

# Handheld AR Target Selection Method using a Smartphone's Front Camera for Targets Behind the User

Reo Sato  
IPLAB  
University of Tsukuba  
Tsukuba, Ibaraki, Japan  
rsato@iplab.cs.tsukuba.ac.jp

Myungguen Choi  
IPLAB  
University of Tsukuba  
Tsukuba, Ibaraki, Japan  
choi@iplab.cs.tsukuba.ac.jp

Buntarou Shizuki  
IPLAB  
University of Tsukuba  
Tsukuba, Ibaraki, Japan  
shizuki@cs.tsukuba.ac.jp

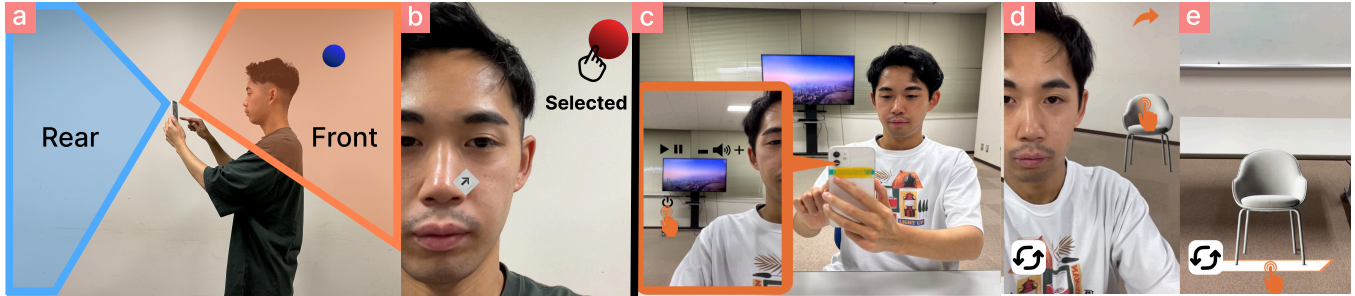


Figure 1: Target selection method using a smartphone's front camera for targets behind the user. a) The blue area represents the rear camera's field of view (FOV) and the orange area represents the front camera's field of view. The blue sphere indicates the target. b) The target is shown in the front camera. When the user touches the target on the touchscreen, the target is selected. c) The user captures a smart TV behind them using the smartphone's front cameras. The TV's menu items are displayed around the TV within the front camera's view. Then the user turns off the display by selecting a menu item. d) The user captures an AR chair behind them using the front camera. e) The user repositions the chair to face forward by switching from the front camera to the rear camera.

## Abstract

In Handheld augmented reality (AR), selecting targets behind the user often requires considerable body or device rotation, increasing selection time and physical effort. In this paper, we present using a smartphone's front camera to enable rear-target selection with minimal body or device movement. This approach supports interaction even when turning around is difficult and can be easily integrated into existing AR applications via a simple camera-switching function. The results of our two user studies show that our method enabled faster and more accurate selection of rear targets, although using both the rear and front cameras introduced confusion due to differences in their operability, leading to increased operation time. Despite this, the fact that most participants chose to use the front camera highlights its potential for supporting 360-degree interaction in Handheld AR.

## CCS Concepts

• Human-centered computing → Mixed / augmented reality.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

SUI '25, Montreal, QC, Canada

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-1259-3/25/11

<https://doi.org/10.1145/3694907.3765930>

## ACM Reference Format:

Reo Sato, Myungguen Choi, and Buntarou Shizuki. 2025. Handheld AR Target Selection Method using a Smartphone's Front Camera for Targets Behind the User. In *ACM Symposium on Spatial User Interaction (SUI '25)*, November 10–11, 2025, Montreal, QC, Canada. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3694907.3765930>

## 1 Introduction

In augmented reality (AR) using a handheld device (Handheld AR), the AR environment is presented by overlaying virtual information onto the real-world image captured by the rear camera of the handheld device, such as a smartphone or a tablet computer. While Handheld AR cannot provide the immersion achieved with video-see-through head-mounted displays (HMDs) or AR glasses, researchers have been exploring Handheld AR applications because it has a notable advantage: It can run on handheld devices many people already own [10, 56], which eliminates the need for specialized hardware. Due to this advantage, Handheld AR is used in various contexts, including spatial design [46, 47], IoT device control [11, 16, 28], 3D sketching [23], virtual try-ons [1], education [36], input methods [2, 5], virtual target selection support [3], and real-world target selection [29].

Many Handheld AR applications require the user to select targets not only in front of them but also within the full 360° space surrounding them [1, 16, 28, 46, 47]. Examples include devices such as fans, lights, TVs, and printers distributed in a room [16], as well as walls or outdoor plants behind or around the user [47]. Selecting

targets behind the user often requires considerable body or handheld device rotation, leading to longer target selection times and increased physical strain. This issue becomes more pronounced when body movement is restricted, such as when the user is seated in a non-rotating chair.

To address this issue, we propose a target selection method that uses the front camera of a handheld device (Figure 1a) to enable interaction with targets behind the user in Handheld AR. While the rear camera is typically used in Handheld AR applications, the front camera is rarely used. By using the front camera, the user can easily select targets behind them via the device's screen, without considerable body or device rotation (Figure 1b). This method is also effective in situations in which turning around is difficult, such as when the user is seated in a non-rotating chair. In such scenarios, the user can utilize Handheld AR to interact with virtual or physical objects behind them, e.g., selecting and turning off a smart TV, without physically rotating their body (Figure 1c).

AR using the front camera can be introduced as an additional feature without disrupting the conventional AR experience based on the rear camera. The user can access this feature simply by switching from the rear to the front camera, which is achieved through a minimal extension of existing AR applications. Furthermore, combining both rear and front cameras realizes new interaction techniques. For example, the user can select a virtual object behind them using the front camera, then move the object to a position in front of them just by switching the camera, without physically repositioning the handheld device (Figure 1d).

This research aims to investigate (1) the performance of selecting a target behind the user in Handheld AR using a front camera, and (2) how the user interacts with the front camera in Handheld AR applications that incorporate both rear and front cameras.

In Study 1, we investigated the performance of selecting targets behind the user in Handheld AR using the front camera. In this study, participants conducted a task for selecting targets behind them with two methods: one that used only the front camera and another that used only the rear camera. The results showed that the front camera enabled significantly faster and more accurate selection of targets behind the user. In Study 2, we investigated when the user chose to use the front camera in situations in which both rear and front cameras were available. In this study, participants performed an object manipulation task involving virtual objects located around them. The results showed that using both the rear and front cameras resulted in longer operation times than using only the rear camera, primarily due to confusion caused by differences in how the two cameras operate. However, many participants frequently used the front camera instead of rotating their bodies backward by more than  $135^\circ$ , which highlights this method's usefulness for facilitating 360-degree interaction in Handheld AR.

Our main contributions are as follows: (a) We show the first Handheld AR target selection method that uses the front camera to select targets behind the user. (b) Through a comparative study, we show that the front-camera-based method enables faster and more accurate selection with reduced physical demand when interacting with targets behind the user. (c) We evaluate the performance of the front camera in a switchable camera setup for 360-degree interaction and analyze practical usage patterns when both rear and front cameras are available.

## 2 Related Work

Since our study explores the performance of selecting AR targets behind the user using a handheld device's front camera, we review target selection methods in Handheld AR, AR systems that utilize the front camera, and interaction methods with occluded and out-of-view targets in the following subsections.

### 2.1 Target Selection and Object Manipulation in Handheld AR

Target selection and object manipulation are fundamental tasks in Handheld AR [10, 39]. Goh et al. categorized interaction techniques for such tasks as touch-based, mid-air gesture-based, or device-based [10].

The touch-based techniques are the most common, and they are often used as a baseline [3, 21, 38, 43, 44]. The user selects or manipulates an object by touching the screen display of the image captured by the rear camera's field of view (FOV). A screen-centered crosshair [3, 21, 33, 43] is often incorporated, and various other enhancements have been proposed, such as freezing the view [4, 43], expanding the target area [33], using a Bubble Cursor [3], and using ray casting [30, 53]. To address more complex scenarios, mid-air gesture-based techniques have utilized pen-based input for mid-air selection (e.g., ARPen [44, 45] or bare-hand gestures for object manipulation [22]), and device-based techniques have been proposed for occlusion-free positioning (e.g., SlidAR [34]).

Prior studies have predominantly focused on selecting targets in front of the user. However, a Handheld AR application inherently requires interaction in a 360-degree space, forcing the user to perform significant body or device rotations to select targets behind them. This could lead to longer selection time and physical strain. Therefore, our research focuses on the challenge of selecting targets behind the user in Handheld AR.

### 2.2 AR Systems That Use the Front Camera

Researchers have developed a limited number of AR systems that utilize the front camera. For example, various companies offer fitting applications [20, 41]. There are also studies on AR systems for trying on sunglasses using the front camera [54, 55]. Zhao et al. [55] proposed a method that utilizes dual cameras to achieve realistic AR rendering. These systems enable the user to examine how an item (e.g., hats, glasses, footwear, and watches) will fit or look while worn by providing an AR rendering. There are also selfie applications with filtering functions [7, 9] and studies that utilize face tracking for makeup [17, 18]. These studies show that the front camera is useful for improving the user's experience and increasing their willingness to make a purchase. In the context of real-world interactions, WorldGaze [29] explored using the front camera to track the user's head direction, allowing mobile voice assistants to respond more precisely to the real-world object or region that the user is looking at.

In contrast, our study focuses on selecting targets behind the user, which differs from these studies regarding the purpose of using the front camera.

### 2.3 Interaction with Occluded and Out-Of-View Targets

Interaction with occluded or out-of-view targets is a known challenge in AR/VR, and various methods have been proposed [25–27, 40, 50]. Lee et al. [25] proposed a selection method using a virtual mirror in VR, although target selection is limited to the area shown on the display. Similarly, vMirror [26] is an interactive widget that uses virtual mirrors to allow the user to observe and select occluded targets in VR. In AR with HMDs, Liliya et al. [27] found that rendering 3D models of targets through obstacles improved task completion times. However, these HMD-based approaches still require the user to turn around to select targets behind them. To avoid pronounced body rotations, Stoakley et al. [40] generated a miniature scene model, called World-In-Miniature, that allows the user to identify and select targets from a bird's-eye view; however, this model can obstruct the main camera view and complicate small target selection. Yan et al. [50] leveraged human spatial memory and proprioception to develop an eyes-free method for selecting targets around the user in VR. However, this approach resulted in lower selection accuracy than conventional methods due to the lack of visual feedback, and its implementation relied on controller-based operations.

These studies developed methods for selecting targets outside the user's FOV by expanding the interaction range in VR environments. In contrast, our method enables the selection of targets behind the user, which are outside the FOV, by using the smartphone's front camera.

## 3 Study 1

We first conducted a user study (Study 1) to investigate the performance of selecting targets behind the user using a front camera.

### 3.1 Participants and Apparatus

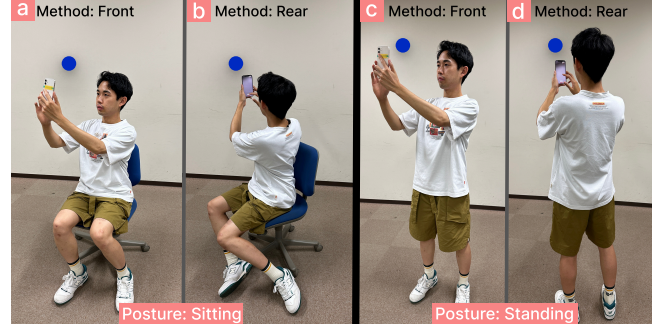
We recruited 12 volunteers (mean age 22.75 years,  $SD = 1.06$ , two females and ten males, ID: P1–P12) from our university. All participants were right-handed. Seven reported using the front camera approximately once per week, four used it once per month, and one used it once per year. Regarding AR application usage, one participant reported using them approximately once per week, one used them once per month, six used them about once per year, and four had never used them.

We used an iPhone 12 (6.9 inch) as a handheld device. The AR application running on the handheld device was developed as a Unity project using Unity ver. 2022.3.30f1 with Apple ARKit XR Plugin [42].

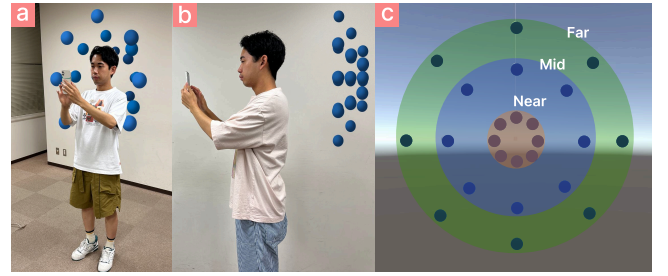
### 3.2 Study Design

We conducted a within-participants study with the following three independent variables: *Method* (*Front*, *Rear*), *Posture* (*Sitting*, *Standing*), and *Target Distance* (*Near* ( $15^\circ$ ), *Middle* ( $30^\circ$ ), and *Far* ( $50^\circ$ )).

*Method* represents the selection method the participants used. In our method (Figure 1), the handheld device displays the front camera's video and overlays virtual information onto the video to present the AR environment. The user can select a target behind the user by rotating the handheld device to include the target in the FOV of the front camera, and then directly touching the position



**Figure 2: Combinations of Method and Posture:** a) *Front*  $\times$  *Sitting*, b) *Rear*  $\times$  *Sitting*, c) *Front*  $\times$  *Standing*, and d) *Rear*  $\times$  *Standing*.



**Figure 3: Target positions.** a) 24 targets are behind the participant. b) All targets are placed 2 m from the front camera. c) Target positions in 2D view.

on the touchscreen. *Rear* is the same as *Front*, except it uses the rear camera instead of the front camera [3, 24, 29–31, 45]. This method serves as the baseline.

*Posture* represents the participant's posture during the task (Figure 2), either sitting and standing. We used these two levels as they are representative postures for performing tasks in Handheld AR. In the *Sitting* condition, participants sit in a chair with a backrest, rotating seat, and casters. In the *Standing* condition, participants stand naturally. To observe natural behavior during target selection behind them, no specific instructions were given regarding body or device movement in either *Posture* condition.

*Target Distance* represents the visual angle between the participant's rearward direction and the target behind them (Figure 3). In the *Near* condition, the target is within the FOV of the front camera, allowing participants to select the target without rotating their body or the device in the *Front* method. In the *Middle* condition, top/bottom targets initially appear at the top or bottom of the front camera's FOV, while horizontal and diagonal targets are outside the FOV when participants hold the device vertically. In the *Far* condition, all the targets are outside the FOV of the front camera. In all *Target Distance* conditions, targets do not initially appear within the FOV of the rear camera.

The order of conditions (combinations of *Method* and *Posture*) was counterbalanced using a Latin square. Participants selected targets in eight directions (top, upper-right, right, lower-right, bottom, lower-left, left, upper-left) for each *Target Distance* (Figure 3c).

Each condition involved three sessions, with 24 trials per session (3 *Target Distances*  $\times$  8 directions). The 24 trials were presented in random order. In total, 3,456 (2 *Methods*  $\times$  2 *Postures*  $\times$  3 sessions  $\times$  24 trials  $\times$  12 participants) valid target-selection data were collected.

### 3.3 Task

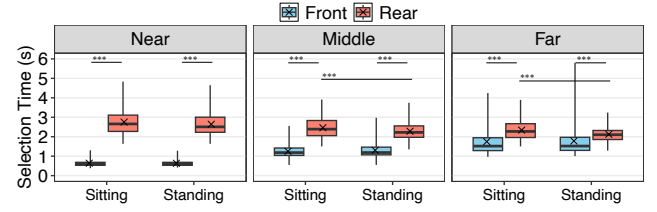
The task was to select blue sphere targets behind participants (Figure 2). During the task, one target was displayed at a time. The task began when participants selected a white square object (i.e., the center of Figure 3c). In the *Front* condition, the square object was displayed 2 m behind the participants, while in the *Rear* condition, it was displayed 2 m in front of them. Participants selected the square by aligning it approximately with the center of the screen and touching it within the handheld device's view. Once selected, the square disappeared, and a target object was presented 2 m behind the handheld device. The target was a blue sphere with an on-screen diameter of approximately 10 mm, which corresponds to the minimum size reliably selectable by direct thumb touch [19, 32]. Additionally, another square object with an arrow indicating the target's direction appeared at the location of the initial square. We provided this directional cue to prevent participants from being hindered by visual search, as the goal of Study 1 was to evaluate the fundamental performance of selecting the target behind the user. Participants were instructed to select the target as quickly and accurately as possible in any comfortable posture (Figure 2). Upon correct selection, the target turned red. After each correct selection, the initial square object reappeared at its previous location. This sequence constituted one trial.

### 3.4 Procedure

Participants were briefed on the user study overview and procedure, after which they completed a pre-study questionnaire querying their gender, age, dominant hand, front camera usage, and AR application usage. Next, as in previous user studies [23, 51, 52, 54, 55], participants held the handheld device vertically (i.e., in portrait orientation) at face height using their non-dominant hand. They were allowed to support the device with their dominant hand if needed, and to use any finger of the dominant hand to perform touch interactions. Participants were located either standing or seated with their backs against a wall.

Participants completed four conditions (*Method*  $\times$  *Posture*), each consisting of four sessions: one practice session and three test sessions. Participants took a break of at least 2 minutes between sessions and conditions, during which they could not use their smartphones.

After completing the task for each *Method*  $\times$  *Posture* condition, participants completed a condition-specific questionnaire asking strategies to select targets, the System Usability Scale (SUS) [6], and the Raw NASA Task Load Index (Raw-TLX) [12, 13]. After completing all tasks, participants were asked via a questionnaire which *Method* they preferred for each *Posture* and *Target Distance*. The user study was video-recorded for later analysis and took approximately 120 minutes per participant.



**Figure 4: Selection time of each independent variable (\*\*\*:  $p < .001$ ).**

### 3.5 Evaluation Metrics

To evaluate the performance of each *Method*, we used selection time, error rate, usability, task load, and user preference as metrics. The selection time is the duration from when participants touched the square object on the touchscreen until the target was selected correctly. The error is a touch outside the designated target area on the touchscreen. The error rate is the proportion of trials in which participants incorrectly touched the touchscreen before correctly selecting the target. Regardless of the number of incorrect touches, each trial with an error was counted as one error trial. Usability was evaluated using the SUS. The task load was evaluated using the Raw-TLX. User preference referred to participants' preferred *Method* for each *Posture* and *Target Direction*.

## 4 Result 1

We excluded outliers from the selection time dataset, which were defined as values more than  $3\sigma$  away from the mean of the dataset for each combination of independent variables. As a result, 44 data were excluded.

We used the non-parametric aligned ranks transformation (ART) method [14, 37, 49] and then a three-way RM-ANOVA on *Method*, *Posture*, and *Target Distance* to analyze selection time and error rate. We used the ART method and then a two-way ANOVA on *Method* and *Posture* to analyze usability and task load. Finally, we used ART-C [8] with Holm correction [15] on these metrics for multiple comparisons.

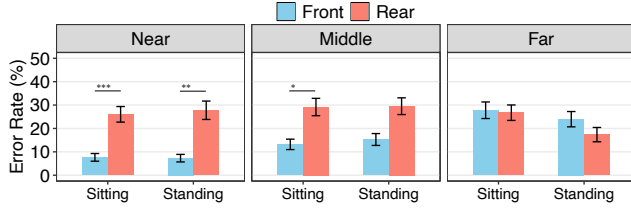
### 4.1 Selection Time

Figure 4 shows the selection times (lower is better) of each combination of *Method*, *Posture*, and *Target Distance*. Those of each *Method* (*Front* and *Rear*) were 1.28 s and 2.46 s. We found that *Method* ( $F_{1,3389.01} = 7794.67$ ,  $p < .001$ ,  $\eta_p^2 = .674$ ), *Posture* ( $F_{1,3389.01} = 39.13$ ,  $p < .001$ ,  $\eta_p^2 = .010$ ), and *Target Distance* ( $F_{2,3389.00} = 160.00$ ,  $p < .001$ ,  $\eta_p^2 = .082$ ) had significant main effects. We also found significant interaction effects between *Method*  $\times$  *Posture* ( $F_{1,3389.01} = 53.03$ ,  $p < .001$ ,  $\eta_p^2 = .013$ ) and *Method*  $\times$  *Target Distance* ( $F_{2,3389.00} = 1349.02$ ,  $p < .001$ ,  $\eta_p^2 = .428$ ). Multiple comparisons showed that the selection time of *Front* was significantly shorter than that of *Rear* in all conditions ( $p < .001$ ).

### 4.2 Error Rate

Figure 5 shows the error rates (lower is better) of each combination of *Method*, *Posture*, and *Target Distance*. Those of each *Method* (*Front*





**Figure 5: Error rate of each independent variable (\*:  $p < .05$ , \*\*:  $p < .01$ , \*\*\*:  $p < .001$ ). The error bars represent the standard errors.**

and *Rear*) were 15.9% and 26.1%. We found that *Posture* ( $F_{1,3433} = 19.03$ ,  $p < .001$ ,  $\eta_p^2 = .006$ ), and *Target Distance* ( $F_{2,3433} = 4.53$ ,  $p < .05$ ,  $\eta_p^2 = .003$ ) had significant main effects. We also found significant interaction effects between *Method*  $\times$  *Posture* ( $F_{1,3433} = 35.10$ ,  $p < .001$ ,  $\eta_p^2 = .010$ ), *Posture*  $\times$  *Target Distance* ( $F_{2,3433} = 13.26$ ,  $p < .001$ ,  $\eta_p^2 = .008$ ), and *Method*  $\times$  *Posture*  $\times$  *Target Distance* ( $F_{2,3433} = 23.40$ ,  $p < .001$ ,  $\eta_p^2 = .013$ ). Multiple comparisons for *Method* showed that the error rate of *Front* at *Near* ( $M = 7.6\%$  ( $SD = 0.28\%$ )) was significantly lower than that of *Rear* ( $M = 26.0\%$  ( $SD = 0.56\%$ )) in both *Sitting* ( $t_{3433} = -4.43$ ,  $p < .001$ ) and *Standing* ( $t_{3433} = -4.01$ ,  $p < .01$ ). Additionally, in *Sitting*, the error rate of *Front* ( $M = 13.2\%$  ( $SD = 0.38\%$ )) was significantly lower than that of *Rear* ( $M = 29.2\%$  ( $SD = 0.63\%$ )) at *Middle* ( $t_{3433} = -3.68$ ,  $p < .05$ ).

### 4.3 Usability and Task Load

The SUS scores of each condition (*Front*  $\times$  *Sitting*, *Rear*  $\times$  *Sitting*, *Front*  $\times$  *Standing*, *Rear*  $\times$  *Standing*) were 76.67, 74.12, 75.63, and 75.63, respectively (higher is better). There was no significant main effect on the SUS score ( $F_{1,33} = 0.88$ ,  $p = .35$ ,  $\eta_p^2 = .0260$ ).

The Raw-TLX mean scores of each condition were 35.07, 38.61, 40.35, and 37.57, respectively (lower is better). We found that *Method* ( $F_{1,33} = 10.34$ ,  $p < .01$ ,  $\eta_p^2 = .239$ ) had significant main effects on the physical demand. In the physical demand, the multiple comparisons showed *Front* ( $M = 38.8$  ( $SD = 27.3$ )) had a significantly lower score than *Rear* ( $M = 60.4$  ( $SD = 22.2$ )) in the *Sitting* posture ( $t_{33} = -3.18$ ,  $p < .05$ ).

### 4.4 Participants' Feedback

**4.4.1 Preference of Method.** Regarding *Posture*, ten out of 12 participants preferred *Front* in *Sitting*; eight preferred it in *Standing*. Regarding *Target Distance*, all participants preferred *Front* for *Near*, nine preferred it for *Middle*, and only four preferred it for *Far*.

**4.4.2 Feedback.** In the participants' responses to the condition-specific questionnaire querying strategies for selecting targets, six reported that when using *Front*, they focused on tilting the smartphone with the wrist rather than moving the body or device. From the video observations, nine mainly used the tilt of the smartphone rather than moving their bodies or chairs. In contrast, when using *Rear* in *Sitting*, three (P2, P5, P12) reported that they focused on locking their arms and rotating or twisting their entire body. There were notable comments: "I felt slightly dizzy at times because I quickly rotated my body using the chair" (P1) and "Spinning in the chair made me feel dizzy during the task" (P3).

In the *Front* condition, participants were confused by the need to rotate their devices and bodies in the opposite direction compared to the *Rear* condition. P4 noted, "It was difficult to capture the target into the touchscreen because the direction of the displayed arrow and the direction to which the smartphone should be moved were opposite." Several participants reported that selecting targets at *Far* using *Front* was difficult due to unnatural wrist movements. For example, participants commented, "It was not possible to find the target just by rotating the wrist" (P12), and "The frustration from the non-intuitive way the smartphone had to be moved to select targets outside the FOV outweighed the benefits of a shorter distance for moving the smartphone" (P2). Additionally, some participants mentioned difficulty in viewing the smartphone screen when *Far* targets were within the frame. P1 and P3 reported that tilting the smartphone largely for selection made it difficult to see the touchscreen, which made selection difficult.

## 5 Study 2

We conducted a user study (Study 2) to compare object manipulation performance between our method, which allows switching between the rear and front cameras, and a baseline method using only the rear camera. The goal of this study was to investigate whether enabling camera switching between the rear and front cameras improves performance in object manipulation tasks, and to examine how the user utilizes this functionality. The study was conducted with the approval of the ethics review committee of our institute (2024R868).

### 5.1 Participants and Apparatus

We recruited 16 participants (mean age 22.69 years,  $SD = 1.21$  years, six females and ten males, ID: P13–P28) from our university. 14 participants were right-handed and two were left-handed. Three reported using the front camera approximately every day, five used it once per week, five used it once per month, and three used it once per year. Regarding AR application usage, one reported using them approximately once per month, four used them once per year, and 11 had never used them. Every participant received approximately 14 USD (2020 JPY). The apparatus used was identical to that in Study 1.

### 5.2 Task

The task was to select a target that appeared in one of seven surrounding directions and reposition it into a goal area located either in front of or behind a participant. The task consisted of three phases: (1) task initiation, (2) target selection, and (3) target repositioning.

Once the study task began, a red sphere, a brown target sphere, a semi-transparent blue goal area, and a diagram indicating the target position and the goal area appeared (see the video figure). The diameters of the red sphere, the target, and the goal area were all 20 mm. The red sphere was located in front of the participants, while the target and the goal area were displayed in specific directions. Both the target and the goal area were located at a vertical angle of  $-10^\circ$  from the device, as this placement allowed them to be easily visible from both the rear and front cameras. All objects were displayed at a distance of 2 m from the device.

First, participants were required to search for the target and the goal area by referring to the diagram or rotating their bodies. They then selected the red sphere with a direct tap (task initiation phase). Once the red sphere was selected, it disappeared, and participants proceeded to select the target using a direct tap with one of the *Methods* (target selection phase).

After the target was selected, it was fixed at the center of the device's screen, at a distance of 2 m from the device. Participants then repositioned the target to the goal area as follows. First, when participants made a touch-down gesture toward the target, the target became attached to their finger. Next, they could drag the target to the desired location and perform a touch-up gesture to place it. This repositioning operation could be repeated multiple times. Once participants were satisfied with the placement, they tapped the task-stop button located in the bottom-right corner of the touchscreen (target repositioning phase). Then, the red sphere and the diagram reappeared, and the positions of the target and goal area were updated to the next position. This sequence constituted one trial. If the task was not completed within 15 seconds after the start of the initiation phase, the current trial ended, and the system moved on to the next task.

### 5.3 Study Design

We conducted a within-participants study with the following three independent variables: *Method* (*Rear&Front*, *Rear*), *Goal Direction* (*Forward* ( $0^\circ$ ), *Backward* ( $180^\circ$ )), and *Target Direction* ( $-45^\circ$ ,  $-90^\circ$ ,  $-135^\circ$ ,  $180^\circ$ ,  $135^\circ$ ,  $90^\circ$ ,  $45^\circ$ ).

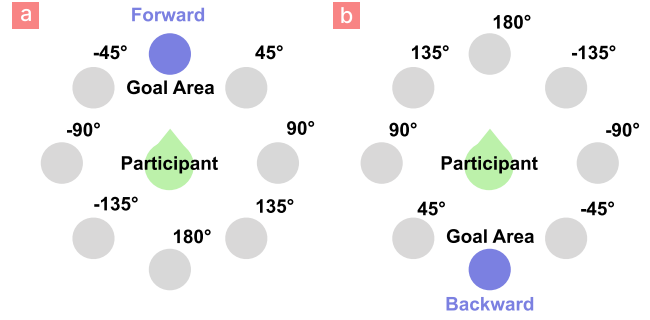
*Method* represents the participant's selection method. *Rear&Front* is our method. In *Rear&Front*, participants can switch between rear and front cameras by tapping a camera switch button (size: 10 mm) located in the bottom-left corner of the screen. Camera switching was optional; participants could decide whether or not to switch between cameras during target selection or placement. *Rear* is a selection method by which the target selection process is the same as that of Study 1. This method served as the baseline.

Figure 6 shows the directions of *Goal Direction* and *Target Direction*. *Goal Direction* represents the direction of the goal area where participants are instructed to place the target. *Forward* corresponds to the participant's front direction at the beginning of the study task, while *Backward* corresponds to the backward direction at the same. *Target Direction* represents the direction of the target relative to the *Goal Direction*. For example, as illustrated in Figure 6b, when *Goal Direction* = *Backward* and *Target Direction* =  $180^\circ$ , the target appears in front of the participant.

The order of *Method* (*Rear&Front* or *Rear*) was counterbalanced. Each condition involved two sessions, with 28 tasks per session ( $2 \text{ Goal Direction} \times 7 \text{ Target Direction} \times 2 \text{ repetitions}$ ). The 28 tasks were presented in random order. In total, 1,792 ( $2 \text{ Methods} \times 2 \text{ sessions} \times 28 \text{ tasks} \times 16 \text{ participants}$ ) valid data were collected.

### 5.4 Procedure

The procedure prior to the study task was identical to that of Study 1. Participants performed the task while standing in a spacious, uncluttered environment to ensure that their movements were not obstructed. They were instructed to select and reposition the



**Figure 6: Directions of Goal Direction and Target Direction. (a) The Goal Direction is Forward. (b) The Goal Direction is Backward. Target Direction consists of seven directions relative to Goal Direction.**

target as quickly and accurately as possible, using any comfortable posture.

Each participant completed two conditions of *Method*, each consisting of one practice session and two test sessions. In the practice session, participants used the assigned *Method* until they became familiar with it. A 30-second break was provided between sessions and conditions, during which participants could not use their smartphones. After completing the task for each condition, participants filled out a condition-specific questionnaire that included questions about the perceived benefits and drawbacks of the *Method*, as well as the SUS and Raw-TLX assessments. After completing all tasks, participants were asked in a final questionnaire which *Method* they preferred. The user study lasted approximately 60 minutes per participant.

### 5.5 Evaluation Metrics

An analysis was conducted on the following metrics:

- **Target Selection Time:** The total time elapsed from the beginning of the target selection phase to its completion.
- **Target Placement Time:** The total time elapsed from the beginning of the target repositioning phase to its completion.
- **Task Completion Time:** The total time elapsed from the beginning of the target selection phase to the end of the target repositioning phase. If the time limit was exceeded, this value was set to 15 s.
- **Target Position Error:** The Euclidean distance between the center of the goal area and the center of the object at the time of task completion.
- **Usability Questionnaire:** Subjective feedback was collected using the SUS and Raw-TLX questionnaires.

## 6 Result 2

We excluded outliers from the dataset of target selection time, target placement time, task completion time, and target position error, which were defined as values more than  $3\sigma$  away from the mean for each combination of independent variables. As a result, 27, 30, 29, and 5 data of each dataset were excluded, respectively. There were 14 data that exceeded the time limit, and these were excluded from the analysis as outliers for task completion time.

Regarding all metrics, we found that they did not follow a normal distribution. Therefore, we used the ART method and then a three-way RM-ANOVA on *Method*, *Goal Direction*, and *Target Direction* to analyze target selection time, target placement time, task completion time, and target position error. Then, we used ART-C with Holm correction on these metrics for multiple comparisons. Finally, we used the Wilcoxon signed rank test with Holm correction to analyze usability and task load.

### 6.1 Target Selection Time

Figure 7a and b show the target selection times (lower is better) of each combination of *Method*, *Goal Direction*, and *Target Direction*. Those of each *Method* (*Rear&Front* and *Rear*) were 1.71 s and 1.37 s, respectively. We found that *Method* ( $F_{1,1722.03} = 420.82$ ,  $p < .001$ ,  $\eta_p^2 = .20$ ), *Goal Direction* ( $F_{1,1722.01} = 104.51$ ,  $p < .001$ ,  $\eta_p^2 = .057$ ), and *Target Direction* ( $F_{6,1722.02} = 41.70$ ,  $p < .001$ ,  $\eta_p^2 = .13$ ) had significant main effects. We also found significant interaction effects between *Method*  $\times$  *Goal Direction* ( $F_{6,1722.02} = 448.36$ ,  $p < .001$ ,  $\eta_p^2 = .061$ ), *Method*  $\times$  *Target Direction* ( $F_{6,1722.01} = 18.62$ ,  $p < .001$ ,  $\eta_p^2 = .061$ ), and *Method*  $\times$  *Goal Direction*  $\times$  *Target Direction* ( $F_{1,1722.02} = 67.54$ ,  $p < .05$ ,  $\eta_p^2 = .19$ ). Multiple comparisons showed that target selection time of *Rear&Front* for *Forward* was significantly longer than that of *Rear* in *Target Direction* =  $-135^\circ$  ( $M = 2.57$  s ( $SD = 0.63$  s) vs.  $M = 1.63$  s ( $SD = 0.45$  s),  $t_{1722.07} = 8.74$ ,  $p < .001$ ),  $180^\circ$  ( $M = 2.30$  s

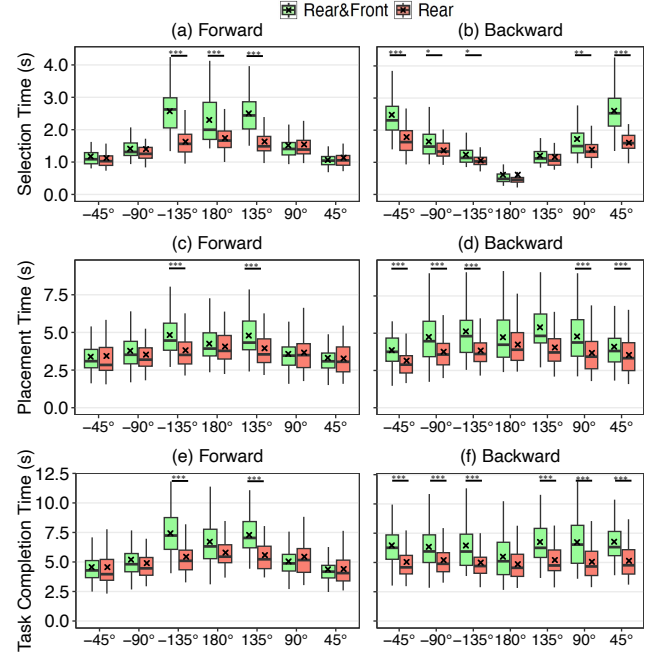
( $SD = 0.78$  s) vs.  $M = 1.74$  s ( $SD = 0.45$  s),  $t_{1722.04} = 4.69$ ,  $p < .001$ ), and  $135^\circ$  ( $M = 2.50$  s ( $SD = 0.66$  s) vs.  $M = 1.63$  s ( $SD = 0.47$  s),  $t_{1722.03} = 8.47$ ,  $p < .001$ ). In addition, target selection time of *Rear&Front* for *Backward* was significantly longer than that of *Rear* in all *Target Directions* ( $p < .05$ ) excluding,  $180^\circ$  and  $135^\circ$  ( $p < .05$ ).

### 6.2 Target Placement Time

Figure 7c and d show the target placement times (lower is better) of each combination of *Method*, *Goal Direction*, and *Target Direction*. Those of each *Method* (*Rear&Front* and *Rear*) were 4.32 s and 3.71 s, respectively. We found that *Method* ( $F_{1,1719.00} = 220.92$ ,  $p < .001$ ,  $\eta_p^2 = .11$ ), *Goal Direction* ( $F_{1,1719.00} = 62.83$ ,  $p < .001$ ,  $\eta_p^2 = .035$ ), and *Target Direction* ( $F_{6,1719.01} = 58.75$ ,  $p < .001$ ,  $\eta_p^2 = .17$ ) had significant main effects. We also found significant interaction effects between *Method*  $\times$  *Goal Direction* ( $F_{1,1719.00} = 38.15$ ,  $p < .001$ ,  $\eta_p^2 = .022$ ) and *Method*  $\times$  *Target Direction* ( $F_{1,1719.00} = 8.57$ ,  $p < .001$ ,  $\eta_p^2 = .029$ ). Multiple comparisons showed that target placement time of *Rear&Front* for *Forward* was significantly longer than that of *Rear* in *Target Direction* =  $-135^\circ$  ( $M = 4.83$  s ( $SD = 1.37$  s) vs.  $M = 3.81$  s ( $SD = 1.28$  s),  $t_{1719} = 6.77$ ,  $p < .001$ ) and  $135^\circ$  ( $M = 4.79$  s ( $SD = 1.39$  s) vs.  $M = 3.95$  s ( $SD = 1.38$  s),  $t_{1719} = 5.36$ ,  $p < .001$ ). In addition, target placement time of *Rear&Front* for *Backward* was significantly longer than that of *Rear* in all *Target Directions*, excluding  $180^\circ$  and  $45^\circ$  ( $p < .001$ ).

### 6.3 Task Completion Time

Figure 7e and f show the task completion times (lower is better) of each combination of *Method*, *Goal Direction*, and *Target Direction*. Those of each *Method* (*Rear&Front* and *Rear*) were 6.10 s and 5.11 s, respectively. We found that *Method* ( $F_{1,1720.01} = 377.92$ ,  $p < .001$ ,  $\eta_p^2 = .20$ ), *Goal Direction* ( $F_{1,1720.01} = 16.68$ ,  $p < .001$ ,  $\eta_p^2 = .0096$ ), and *Target Direction* ( $F_{6,1720.01} = 36.58$ ,  $p < .001$ ,  $\eta_p^2 = .11$ ) had significant main effects. We also found significant interaction effects between *Method*  $\times$  *Goal Direction* ( $F_{1,1720.00} = 32.15$ ,  $p < .001$ ,  $\eta_p^2 = .018$ ), *Method*  $\times$  *Target Direction* ( $F_{1,1720.00} = 10.54$ ,  $p < .001$ ,  $\eta_p^2 = .035$ ), *Goal Direction*  $\times$  *Target Direction* ( $F_{1,1720.01} = 54.73$ ,  $p < .001$ ,  $\eta_p^2 = .16$ ) and *Method*  $\times$  *Goal Direction*  $\times$  *Target Direction* ( $F_{1,1722.01} = 12.98$ ,  $p < .001$ ,  $\eta_p^2 = .043$ ). Multiple comparisons showed that task completion time of *Rear&Front* for *Forward* was significantly longer than that of *Rear* in *Target Direction* =  $-135^\circ$  ( $M = 7.43$  s ( $SD = 1.62$  s) vs.  $M = 5.44$  s ( $SD = 1.48$  s),  $t_{1720.01} = 8.94$ ,  $p < .001$ ) and  $135^\circ$  ( $M = 7.29$  s ( $SD = 1.66$  s) vs.  $M = 5.58$  s ( $SD = 1.61$  s),  $t_{1720.01} = 7.59$ ,  $p < .001$ ). In addition, task completion time of *Rear&Front* for *Backward* was significantly longer than that of *Rear* in all *Target Directions*, excluding  $180^\circ$  ( $p < .001$ ).

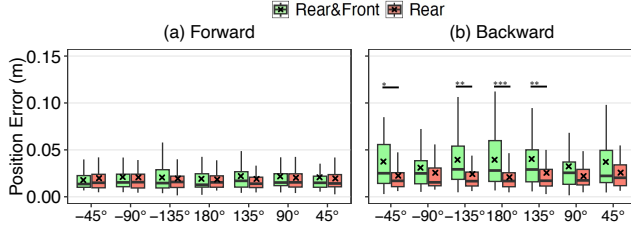


**Figure 7: Time-related metrics of each independent variable\* (\*:  $p < .05$ , \*\*:  $p < .01$ , \*\*\*:  $p < .001$ ). (a, b) Target selection time, (c, d) target placement time, and (e, f) task completion time. (a, c, e) *Goal Direction* = *Forward* and (b, d, f) *Backward*.**

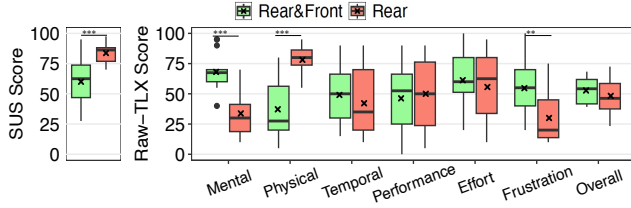
.001,  $\eta_p^2 = .18$ ), *Goal Direction* ( $F_{1,1720.01} = 16.68$ ,  $p < .001$ ,  $\eta_p^2 = .0096$ ), and *Target Direction* ( $F_{6,1720.01} = 36.58$ ,  $p < .001$ ,  $\eta_p^2 = .11$ ) had significant main effects. We also found significant interaction effects between *Method*  $\times$  *Goal Direction* ( $F_{1,1720.00} = 32.15$ ,  $p < .001$ ,  $\eta_p^2 = .018$ ), *Method*  $\times$  *Target Direction* ( $F_{1,1720.00} = 10.54$ ,  $p < .001$ ,  $\eta_p^2 = .035$ ), *Goal Direction*  $\times$  *Target Direction* ( $F_{1,1720.01} = 54.73$ ,  $p < .001$ ,  $\eta_p^2 = .16$ ) and *Method*  $\times$  *Goal Direction*  $\times$  *Target Direction* ( $F_{1,1722.01} = 12.98$ ,  $p < .001$ ,  $\eta_p^2 = .043$ ). Multiple comparisons showed that task completion time of *Rear&Front* for *Forward* was significantly longer than that of *Rear* in *Target Direction* =  $-135^\circ$  ( $M = 7.43$  s ( $SD = 1.62$  s) vs.  $M = 5.44$  s ( $SD = 1.48$  s),  $t_{1720.01} = 8.94$ ,  $p < .001$ ) and  $135^\circ$  ( $M = 7.29$  s ( $SD = 1.66$  s) vs.  $M = 5.58$  s ( $SD = 1.61$  s),  $t_{1720.01} = 7.59$ ,  $p < .001$ ). In addition, task completion time of *Rear&Front* for *Backward* was significantly longer than that of *Rear* in all *Target Directions*, excluding  $180^\circ$  ( $p < .001$ ).

### 6.4 Target Position Error

Figure 8a and b show the target position errors (lower is better) of each combination of *Method*, *Goal Direction*, and *Target Direction*. Those of each *Method* (*Rear&Front* and *Rear*) were 0.029 m and 0.022 m, respectively. We found that *Method* ( $F_{1,1744.00} = 102.02$ ,  $p < .001$ ,  $\eta_p^2 = .055$ ) and *Goal Direction* ( $F_{1,1744.00} = 201.49$ ,  $p < .001$ ,  $\eta_p^2 = .10$ ) had significant main effects. We also found significant interaction effects between *Method*  $\times$  *Goal Direction* ( $F_{1,1744.00} = 86.21$ ,  $p < .01$ ,  $\eta_p^2 = .047$ ) and *Method*  $\times$  *Goal Direction*  $\times$  *Target Direction* ( $F_{1,1744.00} = 2.39$ ,  $p < .01$ ,  $\eta_p^2 = .0081$ ). Multiple comparisons showed



**Figure 8: Target position error of each independent variable** (\*:  $p < .05$ , \*\*:  $p < .01$ , \*\*\*:  $p < .001$ ). (a) *Goal Direction* = Forward and (b) *Backward*.



**Figure 9: The SUS and Raw-TLX Scores of each Method** (\*:  $p < .01$ , \*\*\*:  $p < .001$ ).

that target position error of *Rear&Front* for *Backward* was significantly higher than that of *Rear* in *Target Direction* =  $-45^\circ$  ( $M = 0.038$  m ( $SD = 0.031$  m) vs.  $M = 0.022$  m ( $SD = 0.017$ ),  $t_{1744} = 3.82$ ,  $p < .05$ ),  $-135^\circ$  ( $M = 0.039$  m ( $SD = 0.034$  m) vs.  $M = 0.024$  m ( $SD = 0.022$ ),  $t_{1744} = 4.33$ ,  $p < .01$ ),  $180^\circ$  ( $M = 0.039$  m ( $SD = 0.027$  m) vs.  $M = 0.021$  m ( $SD = 0.014$ ),  $t_{1744} = 5.38$ ,  $p < .001$ ), and  $135^\circ$  ( $M = 0.040$  m ( $SD = 0.034$  m) vs.  $M = 0.026$  m ( $SD = 0.024$ ),  $t_{1744} = 4.26$ ,  $p < .001$ ).

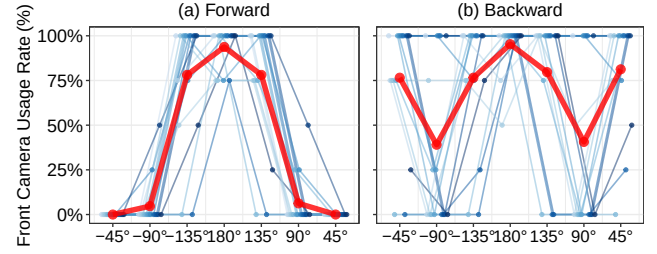
## 6.5 Usability Questionnaire

Figure 9 shows the SUS and Raw-TLX scores. The SUS scores of each *Method* (*Rear&Front*, *Rear*) were 60.00 ( $SD = 19.01$ ) and 83.75 ( $SD = 8.32$ ), respectively (higher is better). We found that *Rear* was significantly higher than *Rear&Front* ( $Z = -3.44$ ,  $p < .001$ ,  $\tau = .860$ ).

The Raw-TLX mean scores of each *Method* (*Rear&Front*, *Rear*) were 52.76 ( $SD = 10.06$ ) and 48.28 ( $SD = 15.15$ ), respectively (lower is better). We found that *Rear* was significantly lower than *Rear&Front* for mental demand ( $M = 33.75$  ( $SD = 17.84$ ) vs.  $M = 68.13$  ( $SD = 14.82$ ),  $Z = 3.38$ ,  $p < .001$ ,  $\tau = .84$ ) and frustration ( $M = 30.00$  ( $SD = 20.66$ ) vs.  $M = 54.69$  ( $SD = 21.87$ ),  $Z = 3.00$ ,  $p < .01$ ,  $\tau = .75$ ). On the other hand, the physical demand of *Rear&Front* was significantly lower than that of *Rear* ( $M = 37.19$  ( $SD = 22.45$ ) vs.  $M = 78.13$  ( $SD = 11.67$ ),  $Z = 3.44$ ,  $p < .001$ ,  $\tau = .86$ ).

## 6.6 Front Camera Usage

We calculated the front camera usage rate, defined as the proportion of trials in which the front camera was used in the *Rear&Front* condition. The rates are summarized in Figure 10. Across all trials, the rate was 53.6%. This result indicates that although participants



**Figure 10: Front camera usage rates for each *Goal Direction* × *Target Direction* condition.** (a) *Goal Direction* = Forward and (b) *Goal Direction* = Backward. The red line indicates the average usage rate, while the blue lines represent the usage rates of individual participants.

were free to switch between the rear and front cameras, more than half of them used the front camera in over half of their trials.

First, in the condition *Goal Direction* = Forward, the front camera usage rate was 37.3%. The front camera was used most frequently when the target appeared at *Target Direction* =  $180^\circ$  (93.8%). The rate was also high for the  $\pm 135^\circ$  directions (78.1%). In contrast, the front camera was rarely used for the  $\pm 45^\circ$  and  $\pm 90^\circ$  directions, with rates of 4.7% and 6.3%, respectively. Next, in the condition *Goal Direction* = Backward, the rate was 69.9%. The front camera was used frequently, with a usage rate of 95.3% when the target appeared at the  $180^\circ$  direction. The rates for the  $45^\circ$ ,  $135^\circ$ ,  $-135^\circ$ ,  $-45^\circ$ ,  $90^\circ$ , and  $-90^\circ$  directions were also higher than those in the *Forward* condition, at 81.3%, 79.7%, 76.6%, 76.6%, 40.6%, and 39.1%, respectively.

Analysis of individual participant rates revealed considerable variation in front camera usage. The rates of each participant are summarized in Table 1. In the condition *Goal Direction* = Backward, six participants (P13, P20–P23, and P27) had a front camera usage rate above 89.3%, and four of them (P13, P22, P23, and P27) used the front camera in 100% of their trials. For these participants, the ability to switch between the rear and front cameras appeared to be consistently valuable. On the other hand, P17 and P24 rarely used the front camera. They only activated the switching function when the *Target Direction* was  $180^\circ$ , and did not use the front camera at all in the *Forward* condition when the target appeared at  $\pm 135^\circ$ . Furthermore, in the *Backward* condition, P17 also avoided using the front camera for the  $\pm 45^\circ$  and  $\pm 135^\circ$  directions.

## 6.7 Preference and Participants' Feedback

We analyzed the responses to the questionnaire asking which *Method* participants preferred, as well as the benefits and drawbacks of each method. The results showed that seven participants preferred *Rear&Front*, while nine preferred *Rear*. Among those who chose *Rear&Front*, five (P13, P14, P19, P21, and P27) cited lower physical demand as the main reason. P24 noted, "I thought a method capable of using two cameras is simply a superior version of one that only has a rear camera." On the other hand, among those who preferred *Rear*, four (P16, P18, P22, and P28) reported that it imposed less cognitive load. Two (P23 and P25) noted that the physical



**Table 1: Front camera usage rates by individual participants for each Goal Direction condition.**

|                 | P13    | P14    | P15    | P16   | P17   | P18   | P19    | P20   |
|-----------------|--------|--------|--------|-------|-------|-------|--------|-------|
| <i>Forward</i>  | 42.9%  | 39.3%  | 28.6%  | 35.7% | 14.3% | 42.9% | 42.9%  | 39.3% |
| <i>Backward</i> | 100.0% | 64.3%  | 71.4%  | 64.3% | 14.3% | 64.3% | 71.4%  | 89.3% |
|                 | P21    | P22    | P23    | P24   | P25   | P26   | P27    | P28   |
| <i>Forward</i>  | 42.9%  | 46.4%  | 42.9%  | 10.7% | 42.9% | 42.9% | 57.1%  | 25.0% |
| <i>Backward</i> | 92.9%  | 100.0% | 100.0% | 17.9% | 46.4% | 71.4% | 100.0% | 50.0% |

effort required to turn their bodies was less demanding than the mental effort involved in learning how to use the front camera.

Regarding the benefits of *Rear&Front*, 11 reported that the method reduced physical demand, as it allowed them to perform the task without rotating their bodies. Three (P16, P22, and P25) specifically noted that they could perform the task with almost no movement, especially when placing a target from front to back or vice versa. In contrast, the drawbacks of *Rear&Front* were also noted. Four (P16, P17, P25, and P26) mentioned that it takes a little bit of time to switch the camera. Additionally, four (P15, P17, P18, and P23) commented that “It is confusing because the direction to move the camera reverses when switching between the rear and front cameras.”

Regarding the benefits of *Rear*, five reported that its operation was intuitive and simple. Additionally, two (P13 and P22) noted that it was easy to understand the target’s position, and three (P14, P18, and P26) commented that the method required low cognitive effort. In terms of drawbacks, five (P14, P18, P20, P24, and P26) stated that the method often required pronounced body movements. Finally, five (P13, P16, P19, P27, and P28) reported experiencing dizziness during its use.

## 7 Discussion

We discuss the target selection performance using a front camera for targets behind the user, the manipulation performance of using a front camera, the purpose of using a front camera, and the limitations of this research.

### 7.1 Selection Performance of Using a Front Camera for Targets Behind the User

The results of Study 1 indicated that participants were able to select targets behind them significantly faster and more accurately using only the front camera (*Front*) than when using only the rear camera (*Rear*). This suggests that the front camera on a handheld device can effectively facilitate target selection behind the user in Handheld AR. In the *Sitting* posture, *Rear* imposed higher physical demands than *Front*. Some participants reported experiencing dizziness due to rapid chair rotations in *Rear*. These findings suggest that, in seated scenarios, *Front* is more suitable than *Rear* for target selection behind the user.

On the other hand, the range in which *Front* outperforms *Rear* in terms of selection time is narrower than the range in which *Rear* is faster. At a target angle of 50° relative to the participant’s rearward direction (*Target Distance* = *Far*), although *Front* yielded faster selection times than *Rear*, as shown in Figure 4, it also resulted

in a higher error rate, as shown in Figure 5, and participants showed a relative preference for *Rear*. These results suggest that *Front* is suitable for *Target Distances* between 0° and ±50°, while *Rear* is suitable outside this range. This discrepancy of range width may be due to differences in the operation methods of *Front* and *Rear*. When using *Front*, participants had to rotate the smartphone in the opposite direction of the target to bring it into FOV, a movement that was unfamiliar and likely contributed to the increased difficulty. Thus, with continued use and increased familiarity, the user may be able to operate *Front* more efficiently across a wider range of angles.

### 7.2 Manipulation Performance of Using Front Camera

Although the switching function for the rear and front cameras significantly reduced physical demand in Study 2, manipulation performance and usability were significantly impaired. In conditions in which the switch function is used frequently (e.g., *Target Direction* = ±135°), *Rear&Front* had significantly longer target selection times, target placement times, and task completion times than *Rear*. Moreover, target position errors, usability scores, and workload scores were worse for *Rear&Front* than for *Rear*.

There are two main reasons why manipulation performance declined when participants used both the rear and front cameras. First, participants became confused because the direction in which the device needed to be moved was reversed between the rear and front camera views. This reversal was unfamiliar to participants and thus likely increased the difficulty of target selection. To mitigate this confusion, introducing a dedicated training session to help the user become accustomed to this control scheme would be useful. Future work should investigate whether this confusion can be overcome with training by evaluating performance after training. Second, the time required to switch between cameras was relatively long, which contributed to increased task completion time. In our application, switching between the rear and front cameras took approximately 0.7 s due to the processing limitations of the smartphone (iPhone 12). This length of time can considerably affect task performance, especially when multiple switches are needed to complete a task. To address this, preparing the switch in advance by predicting when the user is about to tap the camera switch button could reduce perceived latency, task completion time, and task load. One way to achieve this would be leveraging gaze as a predictor [48], which shows that a fixation typically begins 309 ms before the tap at a distance of 35 pixels from the tapped location, and fixation lasts an average of 119 ms.

### 7.3 Purpose for Using Front Camera

The results of Study 2 indicate that participants preferred using the front camera over considerably rotating their bodies, even though the use of the front camera was optional. In the *Goal Direction = Forward* condition, participants used the front camera when the targets were behind them, specifically at the  $180^\circ$  and  $\pm 135^\circ$  directions. In contrast, they did not use it when the targets were located between  $-90^\circ$  and  $90^\circ$ . For targets at  $180^\circ$  and  $\pm 135^\circ$ , participants typically switched to the front camera to select the target, and then switched back to the rear camera to reposition the target in front of them. This allowed them to move rear-located targets forward without substantial body rotations. On the other hand, when targets were located in front or on the side, body rotation was minimal or unnecessary, and thus participants did not use the front camera.

In the *Goal Direction = Backward* condition, participants had a higher front camera usage rate (69.9%) than in the *Goal Direction = Forward* condition (37.3%), and the front camera was frequently used even at *Target Direction =  $\pm 135^\circ$*  (78.1%). This is likely because, with the goal area behind the participants, using the front camera reduced the amount of required body rotation when *Target Distance =  $\pm 135^\circ$* . Notably, even though the total body rotation did not differ, participants also used the front camera for repositioning targets located at  $\pm 90^\circ$  (39.9%). This suggests that the front camera was used not only to reduce body rotation but also to avoid the discomfort or difficulty of rotating backward.

Overall, since many participants used the front camera, the ability to switch between the rear and front cameras could be considered valuable in Handheld AR applications. Admittedly, the high usage rate for targets behind them (e.g., 93.8% at  $180^\circ$  with *Forward*) might have been partly due to the lack of body occlusion in our implementation, which could make the task easier. However, they still frequently used the front camera for diagonally-located targets (e.g., 78.1% at  $\pm 135^\circ$  for *Forward*), where body occlusion is less of a factor. In Study 2, only two (P17 and P24) out of 16 participants rarely used this function; this is most likely simply because the task could be completed without it. Importantly, the presence of the function did not negatively affect the usability of the application for those who chose not to use it. As a means of enabling 360-degree interaction without pronounced body rotations, this function serves as a practical and non-intrusive enhancement to Handheld AR systems.

### 7.4 Limitations and Future Work

This research presents the first attempt to evaluate the performance and usability of using a smartphone's front camera for selecting virtual objects behind the user. As an initial exploration in this area, the findings are preliminary, and further research is needed to fully understand the practical applications and limitations of our front-camera-based AR research.

First, the tasks used in our user studies represent simplified scenarios, i.e., selecting and repositioning isolated spheres in controlled lighting. While this design is suitable for an initial investigation, it may not reflect more realistic situations, such as selecting targets surrounded by distractors in cluttered scenes, selecting targets with unknown locations, or interacting with real-world objects.

Therefore, future work should evaluate our method in such realistic situations.

Second, we exclusively employed a direct-touch selection method in our user studies. While common, this method can suffer from finger occlusion, potentially lowering performance and usability. Therefore, future work should explore how alternative selection methods might perform in this front-camera-based AR (e.g., center-screen targeting [3, 21, 43], ray-casting [11, 53], and a head-based method [29]). These methods could mitigate occlusion issues and potentially improve accuracy for smaller or more distant targets.

Third, in this study, virtual objects remained visible even when they were behind the user, which may have reduced the sense of immersion (e.g., "I did not feel like the target was behind me." (P10)). Ideally, virtual objects behind the user should be occluded by the user's body. During system development, we explored various packages and plugins commonly used in Handheld AR applications; however, none supported occlusion with the front camera, primarily due to the lack of depth-sensing capabilities. To address this limitation, future researchers might utilize a deep learning-based depth estimation from a monocular image [35] to generate the necessary information to estimate the region of the head and body in the image from the front camera. Such information would enable realistic occlusion by not rendering virtual objects that are hidden by the head/body from the front camera's viewpoint. Alternatively, such information could be used to render occluded virtual objects on the head/body with distinct visual styles, such as a single-colored overlay or outlines of the occluded objects, similar to the visual effects used in video games (e.g., *Dark Vision of Dishonored*) to let the player see enemies through walls or obstacles. This approach would keep virtual objects behind the user visible for pointing and thus preserve usability while maintaining the sense of immersion.

Finally, the sample size in our study was relatively small ( $N = 12$  in Study 1 and  $N = 16$  in Study 2), and the participant demographics were skewed in terms of gender and age. Additionally, the large number of conditions may have lowered the statistical power of our analyses. To improve generalizability, future studies should recruit a larger and more demographically balanced participant pool.

## 8 Conclusion

In this paper, we investigated the performance of using the front camera in Handheld AR for selecting targets located behind or around the user. Study 1 demonstrated that the front-camera-only method enabled faster and more accurate selection of targets directly behind the user. Although Study 2 revealed that allowing participants to switch between the rear and front cameras resulted in longer task completion times, participants still used this feature in over half of the trials to reduce body rotation. While several issues remain—such as confusion caused by differences in the operability of the front and rear cameras, and the delay introduced by camera switching—the front camera shows its potential for supporting 360-degree interaction in Handheld AR.

## References

- [1] Shan An, Guangfu Che, Jinghao Guo, Haogang Zhu, Junjie Ye, Fangru Zhou, Zhaoqi Zhu, Dong Wei, Aishan Liu, and Wei Zhang. 2021. ARShoe: Real-Time Augmented Reality Shoe Try-on System on Smartphones. In *Proceedings of the 29th ACM International Conference on Multimedia (Virtual Event, China) (MM*

- '21). Association for Computing Machinery, New York, NY, USA, 1111–1119. doi:10.1145/3474085.3481537
- [2] Tiago Araújo, Carlos Santos, Brunelli Miranda, Nikolas Carneiro, Anderson Marques, Marcelle Mota, Nelson Neto, and Bianchi Meiguins. 2016. Aspects of Voice Interaction on a Mobile Augmented Reality Application. In *Virtual, Augmented and Mixed Reality*, Stephanie Lackey and Randall Shumaker (Eds.). Springer International Publishing, Cham, 199–210.
  - [3] Vinod Asokan, Scott Bateman, and Anthony Tang. 2020. Assistance for Target Selection in Mobile Augmented Reality. In *Proceedings of Graphics Interface 2020* (University of Toronto) (GI 2020). Canadian Human-Computer Communications Society, 56–65. doi:10.20380/GI2020.07
  - [4] Huidong Bai, Gun A. Lee, and Mark Billinghurst. 2012. Freeze View Touch and Finger Gesture Based Interaction Methods for Handheld Augmented Reality Interfaces. In *Proceedings of the 27th Conference on Image and Vision Computing New Zealand* (Dunedin, New Zealand) (IVCNZ '12). Association for Computing Machinery, New York, NY, USA, 126–131. doi:10.1145/2425836.2425864
  - [5] Huidong Bai, Gun A. Lee, Mukundan Ramakrishnan, and Mark Billinghurst. 2014. 3D Gesture Interaction for Handheld Augmented Reality. In *SIGGRAPH Asia 2014 Mobile Graphics and Interactive Applications* (Shenzhen, China) (SA '14). Association for Computing Machinery, New York, NY, USA, Article 7, 7:1–7:6 pages. doi:10.1145/2669062.2669073
  - [6] John Brooke. 1995. SUS: A Quick and Dirty Usability Scale. *Usability Eval. Ind.* 189 (1995), 189–194.
  - [7] Sweet Face Camera. 2017. *Sweet Face Camera: Selfie Edit*. Retrieved August 20, 2025 from <https://apps.apple.com/jp/app/sweet-face-camera-selfie-edit/id1195525697?l=1>
  - [8] 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* (Virtual Event, USA) (UIST '21). Association for Computing Machinery, New York, NY, USA, 754–768. doi:10.1145/3472749.3474784
  - [9] Filter for Selfie. 2017. *Filter for Selfie*. Retrieved August 20, 2025 from <https://play.google.com/store/apps/details?id=covers.maker.filter.selfie>
  - [10] Eg Su Goh, Mohd Shahrizal Sunar, and Ajune Wanis Ismail. 2019. 3D Object Manipulation Techniques in Handheld Mobile Augmented Reality Interface: A Review. *IEEE Access* 7 (2019), 40581–40601. doi:10.1109/ACCESS.2019.2906394
  - [11] Nur Ameerah Binti Abdul Halim and Ajune Wanis Binti Ismail. 2021. ARHome: Object Selection and Manipulation using Raycasting Technique with 3D-model Tracking in Handheld Augmented Reality. In *2021 IEEE 6th International Conference on Computing, Communication and Automation (ICCCA)*. 254–259. doi:10.1109/ICCCA52192.2021.9666416
  - [12] Sandra Hart. 2006. Nasa-task load index (Nasa-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50 (10 2006). doi:10.1177/15419312060500099
  - [13] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139–183. doi:10.1016/S0166-4115(08)62386-9
  - [14] James J. Higgins and Suleiman Tashtoush. 1994. An aligned rank transform test for interaction. *Nonlinear World* 1, 2 (1994), 201–211.
  - [15] Sture Holm. 1979. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* 6, 2 (1979), 65–70.
  - [16] Ke Huo, Yuanzhi Cao, Sang Ho Yoon, Zhuangying Xu, Guiming Chen, and Karthik Ramani. 2018. Scenariot: Spatially Mapping Smart Things Within Augmented Reality Scenes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (CHI '18). Association for Computing Machinery, New York, NY, USA, 219:1–219:13. doi:10.1145/3173574.3173793
  - [17] Ana Javornik, Yvonne Rogers, Delia Gander, and Ana Moutinho. 2017. MagicFace: Stepping into Character through an Augmented Reality Mirror. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (CHI '17). Association for Computing Machinery, New York, NY, USA, 4838–4849. doi:10.1145/3025453.3025722
  - [18] Ana Javornik, Yvonne Rogers, Ana Maria Moutinho, and Russell M. Freeman. 2016. Revealing the Shopper Experience of Using a "Magic Mirror" Augmented Reality Make-Up Application. *Proceedings of the 2016 ACM Conference on Designing Interactive Systems* (2016). <https://api.semanticscholar.org/CorpusID:18704280>
  - [19] Zhao Jin, Tom Plocher, and Liana Kiff. 2007. Touch Screen User Interfaces for Older Adults: Button Size and Spacing. *Lecture Notes in Computer Science* 4554, 933–941. doi:10.1007/978-3-540-73279-2\_104
  - [20] JINS. 2017. *JINS*. Retrieved August 20, 2025 from <https://www.jins.com/jp/>
  - [21] Michal Kapinus, Daniel Bambušek, Zdeněk Materna, Vítězslav Beran, and Pavel Smrž. 2022. Improved Indirect Virtual Objects Selection Methods for Cluttered Augmented Reality Environments on Mobile Devices. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 834–838. doi:10.1109/HRI53351.2022.9889374
  - [22] Minseok Kim and Jae Yeol Lee. 2016. Touch and Hand Gesture-based Interactions for Directly Manipulating 3D Virtual Objects in Mobile Augmented Reality. *Multimedia Tools and Applications* 75, 23 (2016), 16529–16550. doi:10.1007/s11042-016-3355-9
  - [23] Kin Chung Kwan and Hongbo Fu. 2019. Mobi3DSketch: 3D Sketching in Mobile AR. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) (CHI '19). Association for Computing Machinery, New York, NY, USA, 176:1–176:11. doi:10.1145/3290605.3300406
  - [24] Gun A. Lee, Ungyeon Yang, Yongwan Kim, Dongsik Jo, Ki-Hong Kim, Jae Ha Kim, and Jin Sung Choi. 2009. Freeze-Set-Go Interaction Method for Handheld Mobile Augmented Reality Environments. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology* (Kyoto, Japan) (VRST '09). Association for Computing Machinery, New York, NY, USA, 143–146. doi:10.1145/1643928.1643961
  - [25] Joong-Jae Lee and Jung-Min Park. 2020. 3D Mirrored Object Selection for Occluded Objects in Virtual Environments. *IEEE Access* 8 (2020), 200259–200274 pages. doi:10.1109/ACCESS.2020.3035376
  - [26] Nianlong Li, Zhengquan Zhang, Can Liu, Zengyao Yang, Yanan Fu, Feng Tian, Teng Han, and Mingming Fan. 2021. vMirror: Enhancing the Interaction with Occluded or Distant Objects in VR with Virtual Mirrors. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 132, 132:1–132:11 pages. doi:10.1145/3411764.3445537
  - [27] Klemen Liliija, Henning Pohl, Sebastian Boring, and Kasper Hornbæk. 2019. Augmented Reality Views for Occluded Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) (CHI '19). Association for Computing Machinery, New York, NY, USA, 446:1–446:12. doi:10.1145/3290605.3300676
  - [28] Andrea Mattioli and Fabio Paternò. 2023. A Mobile Augmented Reality App for Creating, Controlling, Recommending Automations in Smart Homes. *Proc. ACM Hum.-Comput. Interact.* 7, MHCI, Article 195 (2023), 195:1–195:22 pages. doi:10.1145/3604242
  - [29] Sven Mayer, Gierad Laput, and Chris Harrison. 2020. Enhancing Mobile Voice Assistants with WorldGaze. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–10 pages. doi:10.1145/3313831.3376479
  - [30] Annette Mossel, Benjamin Venditti, and Hannes Kaufmann. 2013. DrillSample: Precise Selection in Dense Handheld Augmented Reality Environments. In *Proceedings of the Virtual Reality International Conference: Laval Virtual* (Laval, France) (VRIC '13). Association for Computing Machinery, New York, NY, USA, Article 10, 10:1–10:10 pages. doi:10.1145/2466816.2466827
  - [31] Takahiro Nagai, Kazuyuki Fujita, Kazuki Takashima, and Yoshifumi Kitamura. 2022. HandyGaze: A Gaze Tracking Technique for Room-Scale Environments using a Single Smartphone. *Proceedings of the ACM on Human-Computer Interaction* 6, ISS (2022), 562:1–562:18. doi:10.1145/3567715
  - [32] Pekka Parhi, Amy K. Karlson, and Benjamin B. Bederson. 2006. Target Size Study for One-handed Thumb Use on Small Touchscreen Devices. In *Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services* (Helsinki, Finland) (MobileHCI '06). Association for Computing Machinery, New York, NY, USA, 203–210. doi:10.1145/1152215.1152260
  - [33] Patrick Perea, Denis Morand, and Laurence Nigay. 2020. Target Expansion in Context: the Case of Menu in Handheld Augmented Reality. In *Proceedings of the 2020 International Conference on Advanced Visual Interfaces* (Salerno, Italy) (AVI '20). Association for Computing Machinery, New York, NY, USA, 32:1–32:9. doi:10.1145/3399715.3399851
  - [34] Jarkko Polvi, Takafumi Taketomi, Goshiro Yamamoto, Arindam Dey, Christian Sandor, and Hirokazu Kato. 2016. SlidAR. *Computers and Graphics* 55, C (2016), 33–43. doi:10.1016/j.cag.2015.10.013
  - [35] Uchitha Rajapaksha, Ferdous Sohel, Hamid Laga, Dean Diepeveen, and Mohammed Bennamoun. 2024. Deep Learning-based Depth Estimation Methods from Monocular Image and Videos: A Comprehensive Survey. *Comput. Surveys* 56, 12, Article 315 (2024), 51 pages. doi:10.1145/3677327
  - [36] Shwetha Rajaram and Michael Nebeling. 2022. Paper Trail: An Immersive Authoring System for Augmented Reality Instructional Experiences. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, 382:1–382:16. doi:10.1145/3491102.3517486
  - [37] K. C. Salter and R. F. Fawcett. 1993. The art test of interaction: a robust and powerful rank test of interaction in factorial models. *Communications in Statistics - Simulation and Computation* 22, 1 (1993), 137–153. doi:10.1080/03610919308813085
  - [38] Leonardo P. Soares, Soraia R. Musse, Márcio S. Pinho, and Jean B. Bousso. 2018. Evaluation of Selection Techniques on a Mobile Augmented Reality Game. In *2018 17th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*. IEEE Computer Society, Los Alamitos, CA, USA, 127–136. doi:10.1109/SBGAMES.2018.00024
  - [39] Becky Spittle, Maite Frutos-Pascual, Chris Creed, and Ian Williams. 2023. A Review of Interaction Techniques for Immersive Environments. *IEEE Transactions on Visualization and Computer Graphics* 29, 9 (2023), 3900–3921. doi:10.1109/TVCG.2022.3174805

- [40] Richard Stokley, 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* (Denver, Colorado, USA) (CHI '95). ACM Press/Addison-Wesley Publishing Co., USA, 265–272. doi:10.1145/223904.223938
- [41] TRYO. 2022. TRYO - Virtual Try On AR App. Retrieved August 20, 2025 from <https://about.tryo.io/>
- [42] Unity. 2022. Apple ARKit XR Plugin. Retrieved August 20, 2025 from <https://docs.unity3d.com/Packages/com.unity.xr.arkit@5.1/manual/index.html>
- [43] Thomas Vincent, Laurence Nigay, and Takeshi Kurata. 2013. Precise Pointing Techniques for Handheld Augmented Reality. In *Human-Computer Interaction – INTERACT 2013*. Springer Berlin Heidelberg, Berlin, Heidelberg, 122–139.
- [44] Philipp Wacker, Oliver Nowak, Simon Voelker, and Jan Borchers. 2019. 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* (Glasgow, Scotland Uk) (CHI '19). Association for Computing Machinery, New York, NY, USA, Article 619, 619:1–619:12 pages. doi:10.1145/3290605.3300849
- [45] Philipp Wacker, Adrian Wagner, Simon Voelker, and Jan Borchers. 2020. Heatmaps, Shadows, Bubbles, Rays: Comparing Mid-Air Pen Position Visualizations in Handheld AR. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–11. doi:10.1145/3313831.3376848
- [46] Zeyu Wang, Cuong Nguyen, Paul Asente, and Julie Dorsey. 2021. DistanciAR: Authoring Site-Specific Augmented Reality Experiences for Remote Environments. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 411, 411:1–411:12 pages. doi:10.1145/3411764.3445552
- [47] Zeyu Wang, Cuong Nguyen, Paul Asente, and Julie Dorsey. 2023. PointShopAR: Supporting Environmental Design Prototyping Using Point Cloud in Augmented Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, 34:1–34:15. doi:10.1145/3544548.3580776
- [48] Pierre Weill-Tessier, Jayson Turner, and Hans Gellersen. 2016. How Do You Look at What You Touch? A Study of Touch Interaction and Gaze Correlation on Tablets. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications* (Charleston, South Carolina) (ETRA '16). Association for Computing Machinery, New York, NY, USA, 329–330. doi:10.1145/2857491.2888592
- [49] 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* (Vancouver, BC, Canada) (CHI '11). Association for Computing Machinery, New York, NY, USA, 143–146. doi:10.1145/1978942.1978963
- [50] Yukang Yan, Chun Yu, Xiaojuan Ma, Shuai Huang, Hasan Iqbal, and Yuanchun Shi. 2018. Eyes-Free Target Acquisition in Interaction Space around the Body for Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (CHI '18). Association for Computing Machinery, New York, NY, USA, 42:1–42:13. doi:10.1145/3173574.3173616
- [51] Hui Ye and Hongbo Fu. 2022. ProGesAR: Mobile AR Prototyping for Proxemic and Gestural Interactions with Real-world IoT Enhanced Spaces. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, 130:1–130:14. doi:10.1145/3491102.3517689
- [52] Hui Ye, Kin Chung Kwan, Wanchao Su, and Hongbo Fu. 2020. ARAnimator: In-situ Character Animation in Mobile AR with User-defined Motion Gestures. *ACM Transactions on Graphics (TOG)* 39, 4 (2020), 83:1–83:12. doi:10.1145/3386569.3392404
- [53] Jibin Yin, Chengyao Fu, Xiangliang Zhang, and Tao Liu. 2019. Precise Target Selection Techniques in Handheld Augmented Reality Interfaces. *IEEE Access* 7 (2019), 17663–17674. doi:10.1109/ACCESS.2019.2895219
- [54] Boping Zhang. 2018. Augmented Reality Virtual Glasses Try-On Technology Based on iOS Platform. *EURASIP Journal on Image and Video Processing* 2018, 1 (2018), 1–19. doi:10.1186/s13640-018-0373-8
- [55] Yiqin Zhao, Sean Fanello, and Tian Guo. 2023. Multi-Camera Lighting Estimation for Photorealistic Front-Facing Mobile Augmented Reality. In *Proceedings of the 24th International Workshop on Mobile Computing Systems and Applications (HotMobile '23)*. ACM, 68–73. doi:10.1145/3572864.3580337
- [56] Fengyuan Zhu and Tovi Grossman. 2020. BISHARE: Exploring Bidirectional Interactions Between Smartphones and Head-Mounted Augmented Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3313831.3376233