Exploring Visual Techniques for Boundary Awareness During Interaction in Augmented Reality Head-Mounted Displays

Wenge Xu* Hai-Ning Liang†‡ Yuzheng Chen§ Xiang Li¶ Kangyou Yu‖

Xi'an Jiaotong-Liverpool University, Suzhou, China

Abstract

Mid-air hand interaction has long been proposed as a 'natural' input method for Augmented Reality (AR) systems. Current AR Head-Mounted Displays (HMDs) have a limited area for hand-based interactions. Because of this, users may easily move their hand(s) outside this tracked area during interaction, especially in dynamic tasks (e.g., when translating an object). Compared to common mid-air interaction issues, such as gesture recognition, arm/hand fatigue, and unnatural ways of interacting with virtual objects (e.g., selecting a distant object), boundary awareness issues in AR devices have received little attention. In this research, we explore visual techniques for boundary awareness in AR HMDs, focusing on object translation tasks. Through a systematic formative study, we first identify the challenges that users might face when interacting with AR HMDs without any boundary awareness information (i.e., how current systems work). Based on the findings, we then propose four methods (i.e., static surfaces, dynamic surface(s), static coordinated lines, and dynamic coordinate line(s)) and evaluate them against the benchmark (i.e., baseline condition without boundary awareness) to make users aware of the tracked interaction area. Our results show that visual methods for boundary awareness can help with dynamic mid-air hand interactions in AR HMDs, but their effectiveness and application are user-dependent.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed/augmented reality; Human-centered computing—Visualization—Visualization techniques;

1 Introduction

Hand-based interaction is one of the most commonly used interaction methods in Augmented Reality (AR) Head-Mounted Displays (HMDs) [26] (e.g., Meta 2, Magic Leap 1, HoloLens), because it is assumed to be natural, practical, and easy to use. The proliferation of reasonably-priced depth cameras and sensors has warranted the investigation of natural user interfaces that are often based on mid-air hand interactions [18]. Currently, most AR HMDs have enabled mid-air hand interaction, but the supported tracked interaction volume is relatively small and limited. Due to this small tracked area, users often observe that the virtual object may not be responding to their gestures during regular interaction (see Fig. 1a for a typical scenario). Such a situation could lead to unnatural and inaccurate interaction experience in different broad interaction scenarios (e.g., AR remote collaboration [15,34]) and in specific tasks (e.g., hand-based text entry in AR HMDs [35]). One way to avoid or mitigate this issue is by allowing users to see via explicit visual cues the tracked interaction area (see Fig. 1b for an example of such method). By knowing the boundary, it might enhance the performance of hand-based text entry technique in AR HMDs, avoid wasting time in remote collaboration [15], enhance the remote learning experience in training like telemedicine [34].

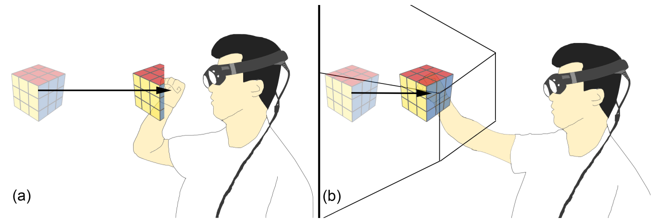

Figure 1: (a; left) (1) The user is trying to drag the object closer to himself. (2) The object partially disappears once the user's hand is outside the AR HMD's interaction boundary. Because the user cannot see the interaction boundary, it can lead to confusion and errors. (b; right) By showing the interaction boundary, the user can interact with the virtual object in AR HMDs and know when to stop his movement to avoid going outside of the tracked interaction area.

This research begins with a formative study to examine the challenges that the user might face when interacting with AR HMDs without boundary information. Then, based on the participants' comments, our observations, and interviews, we propose two preliminary solutions to visualize the interaction boundary of AR HMDs. One is an off-body indicator, which mimics the proprietary guardian boundary system that is used by Oculus Rift HMDs. This visual technique displays a transparent colored surface to remind the user of the boundary. The other one is an on-body indicator, which mimics the hand state notification system that is utilized by Meta 2 AR HMDs. This technique displays a coordinate system on the users' hands to remind the user of the interaction boundary. To understand how to provide such boundary awareness methods and their usefulness, we explore our preliminary solutions to be provided statically—i.e., the system always displays the boundary awareness information, and dynamically—i.e., it displays the information only when it is necessary.

Our investigation of boundary awareness methods in AR HMDs began with one of the most common and essential mid-air interaction tasks—object translation [4]. To understand what the best way is for showing boundary awareness in AR HMDs, we conducted a controlled experiment to assess the accuracy and efficiency of boundary indicators for mid-air hand interactions with AR HMDs. More specifically, we investigated the following two research questions.

RQ1: How accurately and efficiently can users interact with the system in dynamic tasks (i.e., translating virtual objects) when they cannot see the interaction area?

RQ2: How do boundary awareness methods affect the user's subjective feelings of translating virtual objects in AR HMDs?

The contributions of the paper include: (1) the first systematic exploration of visual methods for boundary awareness in AR HMDs,

*e-mail: wenge.xu@xjtlu.edu.cn
†e-mail: haining.liang@xjtlu.edu.cn
‡Corresponding author
§e-mail: yuzheng.chen18@student.xjtlu.edu.cn
¶e-mail: xiang.li18@student.xjtlu.edu.cn
‖e-mail: kangyou.yu18@student.xjtlu.edu.cn

and (2) results of a user study comparing different visual boundary awareness methods for interacting with virtual objects in these systems.

2 Related Work

Our current work builds on prior research on mid-air interaction and its related issues such as hand tracking, gesture recognition, and users' awareness of the boundary of tracked areas.

2.1 Mid-air Interaction

Koutsabasis and Vogiatzidakis [20] indicated that mid-air interaction has the following characteristics: (a) touchless interaction, (b) real-time sensor tracking of (parts of) the user body, (c) body movements, postures and gestures that need to be identified and matched to particular user intentions, goals, and commands. In the following subsections, we describe motion tracking devices/sensors, gesture recognition techniques, and the use of mid-air hand interaction in AR HMDs.

2.1.1 Low-cost Motion Tracking

In 2006, one of the earliest commercial mass production motion tracking products released was the Wiimote controller by Nintendo, which uses an accelerometer and optical sensors to track the user's hand movements. Later, the Sixth Sense [28] presented the first affordable, wearable mid-air gestural interface that enables on-demand augmentation of the physical world with digital information, which can be manipulated via hand gestures. Since then, more and more affordable 3D depth cameras such as Kinect, Leap Motion, and Intel's RealSense have been created to support users' interaction in games or other interactive systems using their bodies to leverage the naturalness of hand and body movements for interaction [37].

2.1.2 Gesture Recognition

Gesture-based interaction alongside other natural methods such as speech improves the efficiency and accuracy of the interactions, and reduces the training time and error rates [5, 17, 24]. Most prior studies on gesture-based interaction are based on the use of one or more RGB cameras [2]. For instance, Dani et al. [11] have proposed a low-cost approach that uses only one monocular RGB camera to enable hand pointing gesture detection and fingertip localization for mobile VR devices. Similarly, Jain et al. [19] presented a low-cost framework that works with just one RGB camera to manipulate objects in mid-air. Kinect, a device that contains an RGB camera and a depth camera, has been widely used for gesture recognition studies. Researchers [30,31] have developed a novel distance metric, the Finger-Earth Mover's Distance (FEMD), to recognize gestures represented from 0 to 9 and using other arithmetic symbols with the data provided by a Kinect. Inspired by FEMD, Wang et al. [6] proposed a novel superpixel earth mover's distance metric for hand gesture recognition. Reyes et al. [25] presented a novel feature weighting approach within the Dynamic Time Warping framework for gesture recognition using depth video data. Combining RGB image and depth image to recognize gestures not only improves the accuracy of the gesture recognition but also allows one hand to overlap with the face or the other hand [1]. In short, with the recent advances of low-cost depth cameras and RGB cameras, many algorithms and techniques (see [7] for a recent review) have been developed to enable gesture recognition for mid-air interaction.

2.1.3 Mid-air Interaction in AR HMDs

There are three main types of interaction approaches for AR HMDs—controller-based, hand-based, hybrid-based (i.e., head pointing and hand gestures) [36]. However, only hand-based input is the most commonly used interaction method for wearable AR HMDs (e.g., HoloLens, Meta 2, Project North Star, and Magic Leap 1) since it is considered intuitive, natural, and cost-effective [5]. In commercial AR HMDs (like Magic Leap 1), users need to perform the following actions to select an object that is close to them. They need first to hover the hand over the virtual object and then perform a grab gesture to select the object [36].

2.2 Boundary Awareness

2.2.1 Issues

According to Bowman et al. [5], current natural interactions (like mid-air hand interaction) provide little additional productivity but make the task more complicated and unnecessarily cumbersome. The main limitations of mid-air interaction in AR HMDs include limited precision with direct input on intangible surfaces [33], arm fatigue [18], and unnatural way of selecting a distant object [5].

In this work, we focus on one limitation of mid-air hand interaction that we refer to as boundary awareness (or lack of it), which is an issue that can occur in motion tracking applications that rely on any type of sensor. For instance, for mid-air interaction, in particular, the user's hand can easily go out of or leave the tracking volume (or area) that the devices' sensor(s) can capture, but the user may not have a conscious awareness that their hands are no longer tracked [27,36] (see Fig. 1a above). This has been observed in early works with motion tracking devices such as Leap Motion [10,27] and Kinect [8] that unavoidably had to have a restricted tracked area due to technical limitations.

For AR systems, lack of boundary awareness could confuse, frustrate and discourage users towards the system because misinterpreted gestures would likely lead to unintentional actions and unresponsiveness for gestures that fall outside of the range and might lead to the users believe that the system recognition is flawed and unusable, thereby leading to an unpleasant experience. For instance, it might affect the text entry accuracy and performance of hand-based text entry techniques (i.e., a text entry technique that involves hand gestures) [35]. It might unnecessarily waste collaboration time due to loss of the hand tracking (due to fewer trackable features in the field of view). [15] reported this in a mock-up Boeing 737 cockpit when using a handheld AR to perform a remote collaboration with tasks like placing annotations, drawing, and live imagery (e.g., of hand gestures). Lack of boundary awareness might also affect other remote collaboration training situations (e.g., remote procedural training of telemedicine [34]).

In short, these above issues become a major problem for interactions where gestures require a wider area of motion [27] and especially for dynamic tasks (e.g., translating virtual objects).

2.2.2 Solutions

Boundary awareness remains a crucial challenge for recent tracking technologies such as Leap Motion and Kinect due to their cameras' limited field-of-view. One solution, as proposed in [22], is to use multiple devices at the same time to increase the tracked area. However, this is not feasible for AR HMDs as the sensors are fixed and mounted on the HMDs. In addition, because AR HMDs are meant to be mobile devices that enable users to move in both indoor and outdoor environments [12,23], setting up multiple depth sensors around the user is not a feasible solution for these AR devices. AR HMDs, unlike standard tracking devices like Leap Motion, is a combination of a tracking and display device, which can not only track users' hands but also provide visual feedback to the users. Therefore, in this paper, we propose and evaluate an alternative solution to allow users to notice the tracking boundary by (1) showing the tracking boundary all the time, or (2) displaying the tracking boundary when their hands are about to leave the device tracking area. To the best of our knowledge, our study represents the first attempt to explore this issue of boundary awareness in AR HMDs.

Table 1: Summary of the advantages and disadvantages of visual methods for boundary awareness that were tested in our study.

Techniques	Advantages	Disadvantages
Static Surfaces	(1) allows users to notice the boundaries easily; and (2) provides this visual information constantly	(1) users have to infer the distance between the hand and the boundary; and (2) may occlude the view
Static Coordinate Lines	(1) provides distance information between the hand and the boundaries; (2) provides this information constantly; and (3) fewer visual objects in the scene when compared to Static Surfaces	(1) visualizes the boundaries in an indirect way
Dynamic Surface(s)	(1) helps visualize the boundaries in a clear way; and (2) provides such information dynamically and as such it does not occlude the interaction space when users' hands are far from the boundary	(1) users have to infer the distance between the hand and the boundary; and (2) there is still some degree of occlusion when users' hands are close to the boundary and the visuals are activated
Dynamic Coordinate Line(s)	(1) provides distance information from the hand to the boundary; and (2) the scene is clearer than (i) Static Coordinate Lines as the lines only appear when users' hands are close to the boundary; and (ii) does not occlude the view	(1) visualizes the boundaries in an indirect way

3 Formative Study

We could not find any prior work that has focused on boundary awareness in AR HMDs. To guide our design, we carried out a formative study to observe and identify challenges faced by users when interacting with AR HMDs with no explicit boundary awareness.

3.1 Formative Study: Method

We recruited six participants (2 females) from a local university, whose ages ranged from 18 to 27. During the one-hour study, we observed participants experiencing a variety of mid-air hand interaction tasks (e.g., manipulating virtual objects, sushi cat, HoloQuarium), while no boundary awareness was provided. After a tutorial, participants interacted with the AR HMD while following a thinking-aloud protocol. They were asked to talk about what they saw, what challenges they had, and possible improvements by having a boundary awareness method to guide their interaction explicitly.

3.2 Formative Study: Findings

Our formative study led to 3 main findings that were extrapolated from participants' comments, our observations during their interaction, and post-experiment interviews.

(1) *Visualizing the boundaries*. During the study, participants had to cope with the system when there was no response to their gestures. In most cases, non-responsiveness was caused by the lack of awareness of the device's tracking area because their hands would stray outside of it. Participants were confused because they were unsure whether it was because of something that they did wrong. This led to 'uncomfortable feelings' and led them to question their ability to work with AR devices in general. This finding led us to hypothesize that if users could be made aware of the tracked area (e.g., via some type of visualization), the cases of non-responsiveness would likely be reduced.

(2) *Distance to the boundary*. We wanted to go deeper into the issue of boundary awareness and asked participants further questions. From the interviews, they indicated that it might be helpful to show how far between their hands were away from the boundary of the tracked interaction area (e.g., *P3*: 'I could be careful of moving hands when I must interact the object near the boundary'). By knowing this, they could prevent their hands from hitting or going outside.

(3) *When to show the boundary*. Although visualizing the boundary seemed necessary, participants also argued that knowing the boundaries may not be that useful when there would not be risks of moving their hands outside the boundary. This was reasonable because the visual FOV of AR HMDs is not large, and having additional visual information would increase the amount of information shown.

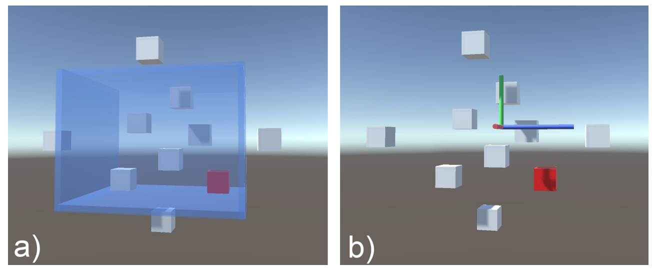

Figure 2: Design of static boundary awareness methods. (a) Static Surfaces (SS) that displays the interaction volume with colored transparent surfaces. (b) Static Coordinate Lines (SCL) that displays the distance to the closest interaction boundary in x-, y-, z-axes via coordinate system.

4 Evaluated Visual Techniques for Boundary Awareness

Findings from the formative study allowed us to propose the following boundary awareness techniques. The testing platforms were all developed and run in Unity3D. We have summarized the advantages and disadvantages of our visual methods for boundary awareness in Table 1.

4.1 Static Surfaces (SS)

This condition provides a visualization of the interaction area in the form of planes or borders (Fig. 2a). The surfaces are shown in blue (i.e., RGB color (0,0,128)) but with 40% opacity to allow users to still see through them. Blue is selected because it works well in indoor environments with white walls [14], which is our experimental environment setting. The area surrounded by the surfaces represents the interaction area. Moving the hand outside the interaction volume leads to tracking issues by the AR headset. The advantages of this method include: (1) allowing users to notice the boundaries easily; and (2) providing such information constantly. On the other hand, the disadvantages of this method include: (1) users have to infer the distance between the hand and the boundary; and (2) because it is visible at all times during interaction, it adds extra visual clutter that may occlude the view of other objects of interest.

4.2 Static Coordinate Lines (SCL)

In this approach, as long as the user's hand is inside the interaction volume, the distance to the volume's surfaces is shown through a 3D coordinate axis. The position of the coordinate center follows the hand position. The length of the line(s) indicates the distance to the boundaries (see Fig. 2b). The advantages of this method include: (1) providing distance information between the hand and the boundaries

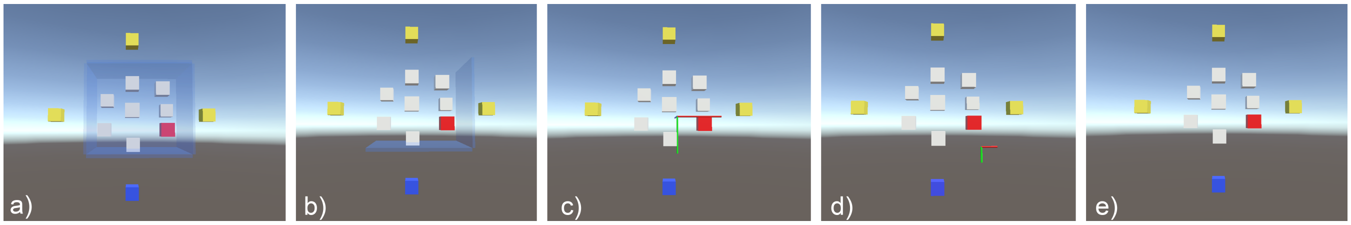

Figure 3: Experiment setting for each boundary awareness technique. a) Static Surfaces, b) Dynamic Surface(s), c) Static Coordinate Lines, d) Dynamic Coordinate Line(s), and e) Benchmark. Note the default Unity3D background was not visible during the experiment.

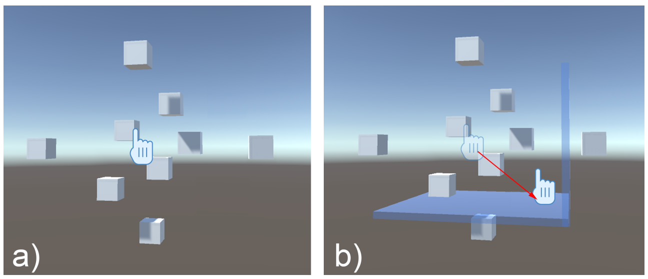

Figure 4: Design of Dynamic Surface(s) (DS) boundary awareness method. (a) The boundaries are not shown if they users are 1.5cm away the interaction boundary. (b) When users are about to move outside the interaction volume, at about 1.5cm to the surface, DS would highlight the corresponding surface(s) to let users be aware of the situation.

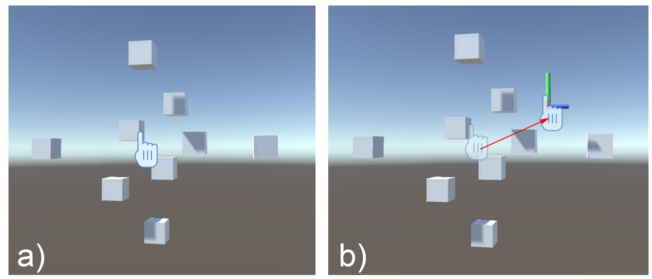

Figure 5: Design of Dynamic Coordinate Line(s) (DCL) boundary awareness method. (a) The coordinate lines are not shown if they users are 1.5cm away from the interaction boundary. (b) When users are about to move outside the interaction volume, at 1.5cm to the boundary, DCL would highlight the corresponding coordinate line(s) to let users be aware that they may possibly be exiting the area.

via simple visuals (in this case lines); (2) providing such information constantly; and (3) there are fewer visual objects in the scene than SS. The disadvantage of this method is that it indirectly visualizes the boundaries.

4.3 Dynamic Surface(s) (DS)

This condition visualizes the surface(s) when the user's hand only gets very close (i.e., 1.5cm) to the corresponding boundary (see Fig. 4). Otherwise, it is analogous to the Benchmark method (no visuals are given). The advantages of this method include: (1) visualizing the boundaries in a clear way; and (2) providing such information dynamically and as such, it does not occlude the interaction space when users' hands are far from the boundary. The disadvantages include: (1) users have to infer the distance between the hand and the boundary; and (2) there will still be some degree of occlusion when users' hands are close to the boundary and the visuals are shown.

4.4 Dynamic Coordinate Line(s) (DCL)

This condition is analogous to the SCL; the only difference is that the system only visualizes the coordinate line(s) when the user's hand gets very close (i.e., 1.5cm) to the corresponding boundary. Like DS, DCL does not show any visual elements for boundary awareness when the users' hands are outside the interaction area (see Fig. 5). The advantages of this method include: (1) providing distance information from the hand to the boundary via simple visual lines; (2) the scene is clearer than (i) SCL as lines only appear when users' hands are very close to the interaction boundary, and (ii) the surfaces approach as line approach does not occlude the view significantly. Its disadvantage is that it is an indirect way to visualize the boundaries.

4.5 Benchmark

This condition does not provide any visual feedback of the tracking boundaries and represents the case of how users currently interact with commercial AR HMDs. This approach acts as the benchmark when there is no boundary information provided to the users. It helps us to understand how users would perform and feel when there are visual cues provided to allow for a comparative analysis with the other four conditions.

4.6 Testbed Environment

The interaction volume is 25cm (width) × 20cm (length) × 16cm (height) and is placed at 42cm in front of the user as Magic Leap 1 can only display virtual items about 40cm away from the user. Users could only perform interaction when their hand is inside the interaction volume. Fig. 3 shows the tested scenes together with the corresponded technique. There are 8 cubes placed inside the interaction volume as target objects and 4 are outside the interaction volume (12 cm away from the surface and is outside the actual visual FOV of the Magic Leap 1) as target translation locations. Visual support are added to help user complete the task in two ways: 1) changing the color of the cube to green when the user's hand is hovering over a cube and 2) displaying an arrow to point out where the target location is when the user selects the cube successfully.

5 EXPERIMENT: OBJECT TRANSLATION

To better understand what the best way is to notify users that they are moving their hands outside the tracking boundary, we looked at user performance and preference for one common and important mid-air interaction—object translation [4]. We conducted a controlled experiment investigating ***RQ1*** and ***RQ2*** (see Introduction section) to explore mid-air translation (dynamic) tasks that would require a more complicated interaction process, from first selecting an object and then moving it to a different location within the AR environment.

5.1 Participants and Apparatus

Twenty participants (7 females, average age: 20.2±2.2 years old, all right-handed with an average arm length of 71.4±4.1 cm) were recruited from a local university campus. They all had normal or corrected-to-normal (using contact lenses) vision. Fourteen of them had prior experience with AR HMDs, but all were not frequent users. None had prior experience with the AR HMD used in the experiment—Magic Leap 1. The experiment was conducted in a university lab.

5.2 Evaluation Metrics

We measured task performance in the form of objective data (speed and accuracy) and collected data describing users' preference to the methods, including subjective feedback (system usability, user experience, workload, arm tiredness, vision tiredness).

5.2.1 Task Performance

The task-completion time was the translation time from the first successful selection of the cube made by the participant to the time when the cube was dragged and dropped at the target location. The error was the number of times the cube hits the boundary as the participant's hand was not rendered by the AR system (i.e., moving outside of the boundary).

5.2.2 User Preference

User Preference was measured by 59 questions compiled from the System Usability Scale (SUS) questionnaire, NASA-TLX workload [16], User Experience Questionnaire (UEQ) [21], Borg CR10 [3], and Computer Vision Syndrome Questionnaire (CVSQ) [32].

5.3 Experiment Design and Procedure

The experiment employed a one-way within-subjects design where the independent variable was Technique (SS, DS, SCL, DCL, and Benchmark). The order of the techniques was counterbalanced.

Before the trials started, participants were asked to complete a pre-experiment questionnaire to gather demographic information and were then given 3 minutes to get familiarized with the Magic Leap 1. Before each condition, they were briefed with the details of the next tested technique. During each condition, a one-minute training session was provided for each participant at the beginning. After each condition, participants were asked to fill in the user preference questionnaires. After the experiment, participants were asked to rank the techniques and give comments on the techniques. The whole experiment lasted about 80 mins.

5.4 Task

During the experiment, one cube would randomly change its color to become the target. Users could use their index finger to target the cube they want to select and select it by using a palm open gesture. The color of the cube would be changed back to the default color and the target location would appear when the selection of the cube was made successfully. To complete the task, the user would need to drag the cube and drop it on the target location (i.e., hitting the center of the cube in the target location). A wrong selection did not cause any effect while an error (i.e., dragging the cube and hitting the boundary) would stop the cube from moving. Participants had to re-select the target if they performed an incorrect selection or made an error. There was a one-second gap for the next target to be highlighted after a successful translation. Each cube would be moved to each target location once, which means that each cube would be moved 4 times. Overall, each participant moved 160 targets (32 cubes × 5 techniques).

5.5 Results

We first applied a Shapiro-Wilk test to evaluate whether the collected data were normally distributed. Then, unless otherwise specified, we employed a one-way repeated ANOVA with Technique as the within-subjects variable. Bonferroni correction was used for pairwise comparisons and Greenhouse-Geisser adjustment was used for degrees of freedom if there were violations to sphericity in the data.

Table 2: Objective measurement and subjective feedback ratings with significant differences between the Boundary Awareness methods. Significance results are highlighted in green.

Method	Task-Completion Time	Error	Mental Workload
SS	2.19±0.67	14.25±12.52	36.00±18.68
DS	1.99±0.33	9.80±7.37	39.00±18.54
SCL	2.87±1.86	22.45±22.13	43.50±18.07
DCL	3.07±1.56	26.55±25.98	49.50±19.12
Benchmark	3.05±1.75	28.55±41.35	47.50±22.91
P	$< .05$	$< .05$	$< .01$

5.5.1 User Performance

The analysis showed that Technique had a significant ($F_{2.733,51.936} = 2.872, p < .05$) effect on the task-completion time. Post-hoc tests confirmed a significantly lower time for the DS compared to DCL. As for errors, a Shapiro-Wilk test indicated that the data were not normally distributed, therefore, we conducted a Friedman's ANOVA where the analysis yielded a significant effect of Technique on errors ($\chi^2(4) = 10.539, p < .05$). Post-hoc analysis with Wilcoxon signed-rank tests were conducted with Bonferroni corrections, resulting in a significance level at $p < .005$. We found that DS had significantly ($p < .001$) smaller number of errors than DCL. Table 2 depicts the mean task-completion time and errors that occurred for all conditions.

5.6 User Preference

NASA-TLX workload. Table 2 depicts the mean mental workload for all conditions. The analysis yielded no significant effect of Technique on overall workload ($F_{4,76} = 1.164, p = .334$). Regarding the NASA-TLX subscales, the analysis yielded a significant influence of Technique on Mental workload ($F_{4,76} = 4.008, p < .01$). Post-hoc tests confirmed that SS caused a significantly (both $p < .05$) lower mental workload than DCL and Benchmark. We did not find any significant effect of Technique on Physical Demand ($p = .301$), Temporal ($p = .582$), Performance ($p = .464$), Effort ($p = .778$), and Frustration ($p = .401$) subscales.

SUS. The analysis revealed that the Technique had no significant ($F_{4,76} = 1.686, p = .162$) effect on the system usability. Benchmark (M = 71.5, SD 13.72) had the highest SUS score while SCL (M = 65.37, SD = 12.23) and DCL (M = 65.37, SD = 13.98) had the lowest.

UEQ. The scores for UEQ was analyzed using the excel tool provided by Laugwitz et al. [21] and had been adjusted between -3 (very bad) to 3 (excellent). The analysis yielded no significant influence of Technique on any of the UEQ subscale: stimulation ($p = .983$), efficiency ($p = .702$), perspicuity ($p = .609$), dependability ($p = .859$), attractiveness ($p = .838$), and novelty ($p = .998$). SCL (M = -0.23, SD = 0.20) had the highest UEQ score while Benchmark (M = -0.92, SD = 0.27) had the lowest.

Borg CR10. The analysis yielded no significant effect of Technique on perceived exertion ($F_{4,76} = .496, p = .739$). DCL (M = 5.33, SD = 2.42) was rated that caused the highest physical fatigue for the participants while SS (M = 4.85, SD 2.25) and SCL (M = 4.85, SD 2.46) were rated the lowest.

CVSQ. A Shapiro-Wilk test indicated that the data were not normally distributed. Therefore, we conducted a Friedman's ANOVA where the analysis yielded no significant effect of Technique on perceived visual fatigue ($\chi^2(4) = 5.272, p = .261$). SCL was rated the worst (M = 2.35, SD = 4.13) while DCL (M = 3.40, SD = 4.12) was the best. The number of participants reported suffering computer vision syndrome in SS, DS, SCL, DCL, benchmark were 4, 3, 2, 4, 3, respectively. A binary logistic regression test showed that each Technique had the same level likelihood to cause computer vision syndrome ($\chi^2(4) = 1.082, p = .897$).

Ranking. The ranking of conditions shows a preference for SS (12 voted SS as the first option) before SCL (15 voted SCL as the first option). Benchmark was selected either the first or the last but mostly placed the last (14 voted it as the last option). Dynamic techniques were equally distrusted in the third and fourth places.

5.7 Qualitative Feedback

In general, most participants stated positive comments to Static and Dynamic Surface(s) boundary indicators: "great/good/wonderful" (*P3*, *P13*, *P20*), "easy to know the position and drag the cube" (*P6*, *P19*). However, we still observe a negative comment, "occluded the view" (*P15*). Regarding the Static Coordinate lines boundary indicator, participants indicated that "[it was] difficult to interact with the cubes" (*P3*, *P10*, *P19*). As for Dynamic coordinate line(s) boundary indicator, they stated that "like it" (*P17*), "easy to interact with cubes" (*P2*, *P5*). For benchmark, they stated, "the view is clear" (*P18*) but "extremely easy to move outside the boundary" (*P10*, *P11*, *P19*).

6 Discussion, Guideline, and Future Work

6.1 Task Performance and User Preference

Task Performance. We found that DS could not only help complete the task faster but also caused fewer errors than DCL. This could be because surface-based boundary awareness is much more apparent, explicit, and obvious than Coordinate Line-based methods. For ***RQ1***, boundary awareness methods, in general, did not help to reduce the errors in translation tasks when compared to Benchmark. However, this was highly user-dependent; for instance, *P2* and *P3* had no issue interact with the Benchmark technique (less than 10 errors) while *P19* and *P20* made more than 100 errors. Moreover, although *P19* and *P20* had many problems interacting with the Benchmark technique, they had no issues interacting with the AR environment with any of the boundary awareness methods, having fewer than 20 errors for all of them.

User Preference. For ***RQ2***, boundary awareness methods could positively affect the user's subjective feelings during the interaction as we found that SS led to a significantly lower mental workload than Benchmark. One possible explanation is that users must be aware that they are moving outside of the tracked boundary in Benchmark condition while they did not have such an issue in SS. Interestingly, although SCL presents the tracking boundary all the time, it was not found to have the same effect as SS.

Based on the ranking data, SS is also preferred as the first option. Coordinate line-based methods are preferred by most users.

All in all, based on our results and user feedback, we suggest that in translation tasks, users should choose a surface-based technique (either SS or DS) over Benchmark as the technique could help users to know the boundary visually to guide their interaction. If users feel that their view is occluded and this interferes with their interaction, they could consider a coordinate line-based technique instead.

6.2 Guidelines for Boundary Awareness

To our knowledge, our paper is the first exploration of boundary awareness for AR HMDs. Based on the results and observations of our study, we formulate the following guidelines and discuss implications for the design of boundary awareness methods in AR HMDs.

6.2.1 User-Dependent

Although there was no significant difference between methods on computer vision syndrome (CVS), we suggest that users should experience all available techniques first and avoid the one(s) which can cause computer vision syndrome to provide a better interaction experience. For example, *P15*, who suffered CVS with SS and Benchmark should not consider using them. In addition, participants, who made 117 errors (*P19*) and 153 errors (*P20*) while using Benchmark, should consider the technique(s) that could help them (e.g., SS for *P19* where only 15 errors occurred and DS for *P20* where only 6 errors occurred). All in all, the boundary awareness method should be tuned to suit the individuals' needs and predispositions.

6.2.2 Providing Boundary Awareness method by Default

During the phase where participants tried the AR device to get to know it, we observed that novice users tended to over-value the FOV of the AR HMD. They would ignore the FOV of the AR device and assume that the interaction would be the same as what they would typically do during actual tasks. Therefore, we suggest that providing a boundary awareness method at the beginning stage to remind the users about the limited size of the tracked area and FOV of the device. It could be disabled when users think they could do without it.

6.3 Limitations and Future Work

The design and results have some limitations, which could frame future research in this area.

Our experiment is limited to the mid-air interaction gestures with one-handed only. Future work can explore whether our findings will also be applicable to two-handed gesture-based interactions where large motions are required. As reported in [27], a gesture that requires a large moving may cause more errors and, therefore, might lead to a different experience.

Several values used in our experiment are pre-defined fixed values due to the lack of related prior work. For instance, we have set the color of the surface(s) blue since it works well in indoor environment with white wall [14]. Future work can 1) implement a dynamic color changing scheme for the surface(s) to suit the background [13, 14]; 2) focus on exploring the most suitable values for opacity of the color and the distance for activating the dynamic visual cues for boundary awareness.

In this research, we have investigated the use of boundary awareness methods in translation tasks [4] with visual methods, which is only the starting point for investigating boundary awareness techniques in AR HMDs. It would be useful to examine the feasibility of boundary awareness methods in other common manipulating tasks in 3D environments (e.g., 3D modeling [9] where the interaction would be more complicated), other AR applications, and even in VR environments (e.g., to compare boundary awareness methods with the one offered by HTC VIVE/Oculus Rift in VR HMDs).

In addition, we have only implemented visual techniques for the boundary awareness problem. Other primary sensory channels [29], such as haptic and auditory, could present feasible and novel solutions but were beyond the scope of the current study. The development of non-visual techniques represents a rich area of future work. For example, audio, haptic, or their combination can be activated when users are about to move their hands outside the tracking boundary. This approach will involve less visual clutter, but more research is needed to understand how well they would work and to determine their optimal parameters.

7 Conclusion

In this paper, we present the first empirical study of visual methods for boundary awareness in Augmented Reality (AR) Head-Mounted Displays (HMDs). We have first conducted a formative study to understand the challenges that users would face when interacting with AR HMDs without boundary information. Then, we have introduced four preliminary candidates for boundary awareness that are then compared to the benchmark, where no boundary information is provided, in the common and important mid-air interaction task of object translation regarding task performance and user preference. Based on the results of our experiment, we suggest the boundary awareness method chosen should be user-dependent. We also list two guidelines for the use of boundary awareness methods in AR HMDs. Because mid-air interaction is an important aspect of current AR and VR devices, issues such as boundary awareness are becoming increasingly critical. Our paper represents a first attempt at exploring and providing low-cost techniques that can improve mid-air interactions for these devices.

ACKNOWLEDGMENTS

We thank all the participants for the time and the reviewers for their comments and suggestions that have helped improve our paper. This research was supported by AI University Research Centre (AI-URC) at Xian Jiaotong-Liverpool University (XJTLU), XJTLU Key Program Special Fund (KSF-A-03 and KSF-02), and XJTLU Research Development Fund.

REFERENCES

[1] M. V. D. Bergh and L. V. Gool. Combining RGB and ToF cameras for real-time 3d hand gesture interaction. *2011 IEEE Workshop on Applications of Computer Vision, WACV 2011*, pp. 66–72, 2011. doi: 10.1109/WACV.2011.5711485

[2] K. K. Biswas and S. K. Basu. Gesture recognition using Microsoft Kinect. In *The 5th International Conference on Automation, Robotics and Applications*, pp. 100–103. IEEE, Wellington, New Zealand, Dec. 2011. doi: 10.1109/ICARA.2011.6144864

[3] G. Borg. *Borg's perceived exertion and pain scales*. Borg's perceived exertion and pain scales. Human Kinetics, Champaign, IL, US, 1998.

[4] D. Bowman, E. Kruijff, and I. Poupyrev Joseph J. LaViola. 3d User Interfaces: Theory and Practice. p. 512. Aug. 2004.

[5] D. A. Bowman, R. P. McMahan, and E. D. Ragan. Questioning naturalism in 3d user interfaces. *Communications of the ACM*, 55(9):78–88, Sept. 2012. doi: 10.1145/2330667.2330687

[6] C. Wang, Z. Liu, and S. Chan. Superpixel-Based Hand Gesture Recognition With Kinect Depth Camera. *IEEE Transactions on Multimedia*, 17(1):29–39, Jan. 2015. doi: 10.1109/TMM.2014.2374357

[7] C. Chen, R. Jafari, and N. Kehtarnavaz. A survey of depth and inertial sensor fusion for human action recognition. *Multimedia Tools and Applications*, 76(3):4405–4425, Feb. 2017. doi: 10.1007/s11042-015-3177-1

[8] H. Cook, Q. V. Nguyen, S. Simoff, T. Trescak, and D. Preston. A Close-Range Gesture Interaction with Kinect. In *2015 Big Data Visual Analytics (BDVA)*, pp. 1–8. IEEE, Hobart, Australia, Sept. 2015. doi: 10.1109/BDVA.2015.7314284

[9] J. Cui, A. Kuijper, D. W. Fellner, and A. Sourin. Understanding People's Mental Models of Mid-Air Interaction for Virtual Assembly and Shape Modeling. In *Proceedings of the 29th International Conference on Computer Animation and Social Agents*, CASA '16, pp. 139–146. ACM, New York, NY, USA, 2016. event-place: Geneva, Switzerland. doi: 10.1145/2915926.2919330

[10] J. Cui and A. Sourin. Feasibility Study on Free Hand Geometric Modelling Using Leap Motion in VRML/X3d. In *2014 International Conference on Cyberworlds*, pp. 389–392. IEEE, Santander, Cantabria, Spain, Oct. 2014. doi: 10.1109/CW.2014.60

[11] M. Dani, G. Garg, R. Perla, and R. Hebbalaguppe. Mid-air fingertip-based user interaction in mixed reality. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 174–178, Oct 2018. doi: 10.1109/ISMAR-Adjunct.2018.00061

[12] N. A. ElSayed, B. H. Thomas, K. Marriott, J. Piantadosi, and R. T. Smith. Situated Analytics: Demonstrating immersive analytical tools with Augmented Reality. *Journal of Visual Languages & Computing*, 36:13–23, Oct. 2016. doi: 10.1016/j.jvlc.2016.07.006

[13] J. L. Gabbard, J. E. Swan, D. Hix, S. Kim, and G. Fitch. Active text drawing styles for outdoor augmented reality: A user-based study and design implications. In *2007 IEEE Virtual Reality Conference*, pp. 35–42, March 2007. doi: 10.1109/VR.2007.352461

[14] J. L. Gabbard, J. E. Swan, D. Hix, R. S. Schulman, J. Lucas, and D. Gupta. An empirical user-based study of text drawing styles and outdoor background textures for augmented reality. In *IEEE Proceedings. VR 2005. Virtual Reality, 2005.*, pp. 11–18, March 2005. doi: 10.1109/VR.2005.1492748

[15] S. Gauglitz, C. Lee, M. Turk, and T. Höllerer. Integrating the physical environment into mobile remote collaboration. In *Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '12, p. 241–250. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2371574.2371610

[16] S. G. Hart. NASA Task load Index (TLX). Volume 1.0; Paper and pencil package. 1986.

[17] G. Heidemann, I. Bax, and H. Bekel. Multimodal interaction in an augmented reality scenario. In *Proceedings of the 6th International Conference on Multimodal Interfaces*. ACM Press, State College, PA, USA, Oct. 2004. doi: DOI: 10.1145/1027933.1027944

[18] J. D. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani. Consumed endurance: a metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, pp. 1063–1072. ACM Press, Toronto, Ontario, Canada, 2014. doi: 10.1145/2556288.2557130

[19] V. Jain, G. Garg, R. Perla, and R. Hebbalaguppe. Gestarlite: An on-device pointing finger based gestural interface for smartphones and video see-through head-mounts. *arXiv preprint arXiv:1904.09843*, 2019.

[20] P. Koutsabasis and P. Vogiatzidakis. Empirical Research in Mid-Air Interaction: A Systematic Review. *International Journal of Human-Computer Interaction*, pp. 1–22, Feb. 2019. doi: 10.1080/10447318.2019.1572352

[21] B. Laugwitz, T. Held, and M. Schrepp. Construction and Evaluation of a User Experience Questionnaire. In A. Holzinger, ed., *HCI and Usability for Education and Work*, vol. 5298, pp. 63–76. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. doi: 10.1007/978-3-540-89350-9_6

[22] S. J. Lee, Y. Motai, and H. Choi. Tracking Human Motion With Multichannel Interacting Multiple Model. *Industrial Informatics, IEEE Transactions on*, 9:1751–1763, Aug. 2013. doi: 10.1109/TII.2013.2257804

[23] A. Liverani, G. Amati, and G. Caligiana. A CAD-augmented Reality Integrated Environment for Assembly Sequence Check and Interactive Validation. *Concurrent Engineering*, 12(1):67–77, Mar. 2004. doi: 10.1177/1063293X04042469

[24] Z. Lv, A. Halawani, S. Feng, H. Li, and S. Réhman. Multimodal Hand and Foot Gesture Interaction for Handheld Devices. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 11:1–19, Oct. 2014. doi: 10.1145/2645860

[25] M. Reyes, G. Domínguez, and S. Escalera. Featureweighting in dynamic timewarping for gesture recognition in depth data. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pp. 1182–1188, Nov. 2011. doi: 10.1109/ICCVW.2011.6130384

[26] S. Malik, C. McDonald, and G. Roth. Hand tracking for interactive pattern-based augmented reality. *International Symposium on Mixed and Augmented Reality, ISMAR 2002*, pp. 117–126, 2002. doi: 10.1109/ISMAR.2002.1115080

[27] B. P. Miranda, N. J. S. Carneiro, T. D. O. de Araújo, C. G. R. dos Santos, A. A. de Freitas, J. M. de Morais, and B. S. Meiguins. Categorizing issues in mid-air infovis interaction. In *2016 20th International Conference Information Visualisation (IV)*, pp. 242–246. IEEE, Lisbon, Portugal, July 2016. doi: 10.1109/IV.2016.70

[28] P. Mistry and P. Maes. SixthSense: a wearable gestural interface. In *International Conference on Computer Graphics and Interactive Techniques, SIGGRAPH ASIA*. ACM, Yokohama, Japan, Dec. 2009. doi: DOI: 10.1145/1667146.1667160

[29] D. H. Mortensen, S. Bech, D. R. Begault, and B. D. Adelstein. The relative importance of visual, auditory, and haptic information for the user's experience of mechanical switches. *Perception*, 38(10):1560–1571, 2009.

[30] Z. Ren, J. Yuan, J. Meng, and Z. Zhang. Robust part-based hand gesture recognition using kinect sensor. *IEEE Transactions on Multimedia*, 15(5):1110–1120, Aug 2013. doi: 10.1109/TMM.2013.2246148

[31] Z. Ren, J. Yuan, and Z. Zhang. Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera. *19th ACM international conference on Multimedia*, (June 2014):1093, 2011.

doi: 10.1145/2072298.2071946

[32] M. d. M. Seguí, J. Cabrero-García, A. Crespo, J. Verdú, and E. Ronda. A reliable and valid questionnaire was developed to measure computer vision syndrome at the workplace. *Journal of Clinical Epidemiology*, 68(6):662–673, June 2015. doi: 10.1016/j.jclinepi.2015.01.015

[33] Z. Szalavári and M. Gervautz. The Personal Interaction Panel - a Two-Handed Interface for Augmented Reality. *Comput. Graph. Forum*, 16:335–346, Sept. 1997. doi: 10.1111/1467-8659.16.3conferenceissue.35

[34] S. Wang, M. Parsons, J. Stone-McLean, P. Rogers, S. Boyd, K. Hoover, O. M. Pastor, M. Gong, and A. Smith. Augmented reality as a telemedicine platform for remote procedural training. In *Sensors*, vol. 17, 2017. doi: 10.3390/s17102294

[35] W. Xu, H. Liang, A. He, and Z. Wang. Pointing and selection methods for text entry in augmented reality head mounted displays. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 279–288, Oct 2019. doi: 10.1109/ISMAR.2019.00026

[36] W. Xu, H.-N. Liang, Y. Zhao, D. Yu, and D. Monteiro. DMove: Directional Motion-based Interaction for Augmented Reality Head-Mounted Displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, pp. 1–14. ACM Press, Glasgow, Scotland Uk, 2019. doi: 10.1145/3290605.3300674

[37] Z. Zhang. Microsoft Kinect Sensor and Its Effect. *IEEE MultiMedia*, 19(2):4–10, Feb. 2012. doi: 10.1109/MMUL.2012.24