Multimodal Interaction with Gaze and Pressure Ring in Mixed Reality

Zhimin Wang[1] Jingyi Sun[1] Mingwei Hu[2] Maohang Rao[1] Yangshi Ge[1] Weitao Song[2]
Feng Lu[1] *

[1] State Key Laboratory of VR Technology and Systems, School of CSE, Beihang University
[2] School of Optics and Photonics, Beijing Institute of Technology

Abstract

Controller-free augmented reality devices currently rely on gesture interaction, which can cause arm fatigue and limit hand mobility. This paper proposes a multimodal interaction technique that combines gaze and pressure ring worn on the index finger. The proposed technique eliminates the need for direct hand capturing using the camera, allowing for more flexible hand movements and reducing fatigue. The experiment conducted in this study demonstrates the effectiveness of the pressure ring in enhancing interaction efficiency.

Index Terms: Human-centered computing—Human computer interaction (HCI)—HCI design and evaluation methods—User studies; Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed/augmented reality

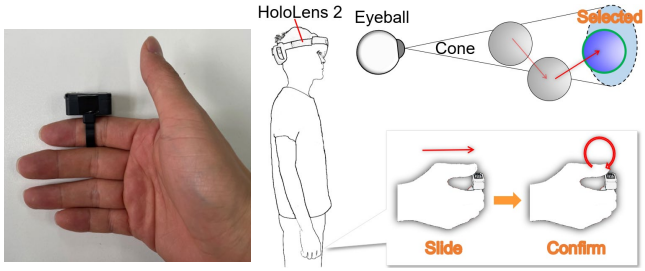
1 Introduction

Virtual Reality (VR) and Augmented Reality (AR) have blurred the boundaries between the digital and physical worlds, resulting in heightened user immersion compared to traditional display technologies [5]. The interaction with virtual objects is crucial for enhancing the user experience. However, existing devices often focus on enhancing the interaction through a single modality. Free-hand interaction aligns more closely with everyday human interaction patterns, but prolonged use can lead to arm fatigue [2]. While eye gaze-based interaction requires less physical effort than gesture-based interaction, it faces challenges due to limited gaze accuracy and the Midas Touch problem [2, 4]. Consequently, there is a need for further investigation into maximizing the advantages of these interaction methods.

Multimodal interaction aims to improve usability by leveraging the strengths of each modality [3]. One intuitive idea is to combine gestures and eye movements to achieve hand-eye coordination. By capitalizing on the fast speed of eye movement and the precision of hand gestures, a "gaze select, hand manipulates" paradigm offers a natural and efficient interaction experience [6]. Apple's Vision Pro has taken hand-eye coordination to new heights [1]. Regarding UI manipulation, it significantly enhances interaction efficiency by integrating eye movements and gestures to carry out actions like clicking, scrolling, and zooming within the interface. What's more, Bao *et al.* propose eye gaze as guidance in conjunction with gestures to select objects that are heavily occluded and perform object translation in depth, thereby expanding the possibilities of hand-eye coordination [2].

Despite the complementary natures in hand-eye coordination, this combination still faces certain limitations. Gesture-based interaction requires the camera on the AR headset to capture hand movements. Apple's Vision Pro alleviates this limitation by installing a camera beneath the headset to capture hand motion, allowing for a broader range of hand movements. However, it is still necessary to capture the hand in the camera's view. Particularly, when users interact while standing, raising their arms to a certain height can still be fatiguing.

To address these concerns, we propose a multimodal interaction technique that integrates gaze and gesture. Our approach encompasses the following steps: 1) Developing a wearable interaction module equipped with a compact pressure ring. This ring is worn on the index finger and controlled by the thumb. Consequently, users can interact with their hand without needing to position hand in front of the camera of an AR headset, even keeping them hidden from view. 2) we propose the control mode of the pressure ring and design the *Gaze-Gesture* interaction, allowing this combination to perform complex operations comparable to those achieved by hand-eye coordination. Our experiment demonstrates that the pressure ring eliminates the need for direct hand capturing using the camera, enabling more flexible hand movements and reducing fatigue.



(a) Pressure Ring (b) Multimodal Interaction

Figure 1: We propose a *Gaze-Gesture* interaction using a pressure ring that enables users to manipulate objects without the need for hand gesture captured by AR HMD. The ring is worn on the index finger and operated with the thumb.

2 System Design

Pressure Ring Design: In this study, we utilize the Distributed Flexible Pressure Sensor (FPS) as our sensor of choice. The FPS, which is a compact pressure-sensitive touch interface resembling a coin, is employed to facilitate ease of use. To accommodate this sensor, we have designed a specially crafted ring, as shown in Fig. 1. The pressure ring has overall sizes of 31mm x 21mm x 25mm, with the flexible pressure sensor securely attached to its top. The ring is fastened to the index finger, allowing the user to operate the device by placing their thumb on the sensor positioned at the top of the finger clip.

***Gaze-Gesture* interaction (*GG*)**: During the selection phase, the user gazes at the target object and confirms the selection by long-pressing the sensor. This action triggers the appearance of a viewing area on the interface, which is a cone with a 6° angle surrounding the

*Corresponding Author: Feng Lu (lufeng@buaa.edu.cn).

line of sight. Objects within the cone become selectable. The user can adjust the position of the viewing area by sliding their fingertip in eight directions on the sensor. Once the adjustment is made, the user long-presses to enter the Depth Selection step. In this step, the user can change the target object among the selectable objects by sliding their fingertip up and down. The user confirms the selection by long-pressing.

Before confirmation, unselected objects that are closer than the current target object become transparent, allowing the user to observe obstructed objects. Upon confirmation, the selection phase ends, and the translation phase begins. Similarly, the user gazes at the destination for the translation and confirms by long-pressing the sensor. The user then fine-tunes the direction of the destination by sliding in various directions on the sensor and long-presses to enter the Depth Translation step. Finally, the user adjusts the depth of the placement location by sliding up and down to reach the target position.

3 Task Design

Four spheres are placed in the center of a 1.5m×1.2m area. They are at distances ranging from 2m to 5m away from the participant and are completely covered by 3 to 4 interfering cubes each. Additionally, four target cubes are positioned at the corners of a 3m x 3m rectangular area located 3m away from the participant. To complete the task, the participant must deal with the challenges posed by the interfering objects, select the target sphere that is completely occluded, and place it in the corresponding target position.

4 Experimental results

We recruited 16 subjects on campus. We conducted our experiments using Microsoft HoloLens 2, which is equipped with gaze estimation and speech detection capabilities. We employed three gaze-related interaction techniques to investigate the merits and demerits of various interaction techniques in our experiment. Apart from the *Gaze-Gesture* technique that utilizes the pressure ring, we design two additional techniques: Gaze-only Interaction and Gaze-Speech Interaction.

Gaze-Only Interaction (*Gaze*): The user selects an object by gazing at it, with the closest object chosen if multiple ones are in focus. During the translation phase, the user directs the object to its destination using a gaze-activated menu bar with forward, backward, pause, and release buttons. The interaction concludes by gazing at the release button once the object reaches its intended location.

Gaze-Speech Interaction (*GS*): The user utilizes a combination of gaze and speech for object selection and manipulation, employing commands like confirm, release, stop, and directional instructions. The selection phase involves gazing at the target and using speech commands to adjust the viewing area and select objects, followed by confirmation to initiate the translation phase. During translation, users direct the object's movement towards a desired destination using speech commands, concluding the process with a confirmation or release command.

Evaluation Metric: We selected the average finish time as our metric, which is calculated as the total finish time divided by the number of completed targets. In cases where no targets are completed, both the total finish time and average finish time are set to 3 minutes. This time encompasses the duration spent on both the selection phase and translation phase.

Result: The result is shown in Fig.2. In the statistical analysis, repeated-measures ANOVAs ($\alpha = 0.05$) were conducted to examine the differences among the three interaction techniques in terms of Average Finish Time. The results indicated significant differences ($F(2, 30) = 40.416$, $p < 0.001$).

Further pairwise comparisons were performed to assess the specific performance of each interaction technique. The performance of *GG* was found to be significantly better than *Gaze* (p = 0.017),

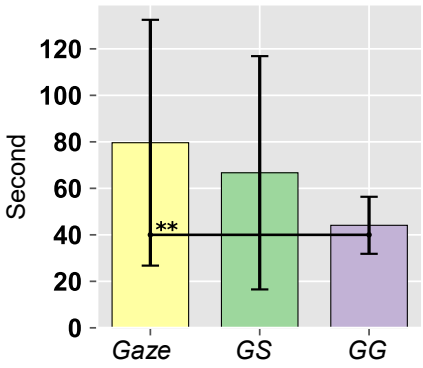


Figure 2: Bar charts of the interaction techniques' performance under Average Finish Time. Error bars indicate the standard error. The statistical significances are labeled with **($p < 0.05$)

demonstrating its superior effectiveness. Additionally, *GG* exhibited a slight advantage over *GS*, although this difference did not reach statistical significance (p = 0.086).

These findings highlight the potential of *GG* as a promising interaction technique in improving Average Finish Time compared to both *Gaze* and *GS*.

5 Conclusion

In conclusion, the proposed *Gaze-Gesture* interaction technique offers a promising solution to the limitations of current controller-free AR devices that rely on gesture interaction. By combining gaze and a small, pressure ring worn on the index finger, this technique allows for more flexible hand movements and reduces arm fatigue. The experiment conducted in this study shows that the pressure ring enhances interaction efficiency, demonstrating the potential of this technique to improve the user experience in AR applications. The *Gaze-Gesture* interaction technique opens up new possibilities for more natural and intuitive interaction in AR, paving the way for further advancements in this field.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant 62372019 and the Academic Excellence Foundation of BUAA for PhD Students.

References

[1] Apple Vision Pro. [Online]. https://www.apple.com/apple-vision-pro/, Accessed January 31, 2024.

[2] Y. Bao, J. Wang, Z. Wang, and F. Lu. Exploring 3d interaction with gaze guidance in augmented reality. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 22–32, 2023. doi: 10.1109/VR55154.2023.00018

[3] M. Lee, M. Billinghurst, W. Baek, R. D. Green, and W. Woo. A usability study of multimodal input in an augmented reality environment. *Virtual Real.*, 17(4):293–305, 2013.

[4] Z. Wang, H. Wang, H. Yu, and F. Lu. Interaction with gaze, gesture, and speech in a flexibly configurable augmented reality system. *IEEE Transactions on Human-Machine Systems*, 51(5):524–534, 2021.

[5] Z. Wang, Y. Zhao, and F. Lu. Gaze-vergence-controlled see-through vision in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3843–3853, 2022. doi: 10.1109/TVCG.2022.3203110

[6] D. Yu, X. Lu, R. Shi, H. Liang, T. Dingler, E. Velloso, and J. Gonçalves. Gaze-supported 3d object manipulation in virtual reality. In *CHI '21: CHI Conference on Human Factors in Computing Systems*, pp. 734:1–734:13. ACM, 2021. doi: 10.1145/3411764.3445343