Understanding Window Management Interactions in AR Headset + Smartphone Interface

Jie Ren
Tsinghua University
Beijing, China
renj17@mails.tsinghua.edu.cn

Chun Yu
Tsinghua University
Beijing, China
chunyu@tsinghua.edu.cn

Yueting Weng
Tsinghua University
Beijing, China
wengyt19@mails.tsinghua.edu.cn

Yuanchun Shi
Tsinghua University
Beijing, China
shiyc@.tsinghua.edu.cn

Chengchi Zhou
Tsinghua University
Beijing, China
zcc16@mails.tsinghua.edu.cn

CHI '20 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6819-3/20/04.
DOI: https://doi.org/10.1145/3334480.3382812

Abstract
We envision a combinative use of an AR headset and a smartphone in the future that can provide a more extensive display and precise touch input simultaneously. In this way, the input/output interface of these two devices can fuse to redefine how a user can manage application windows seamlessly on the two devices. In this work, we conducted a formative interview with ten people to provide an understanding of how users would prefer to manage multiple windows on the fused interface. Our interview highlighted that the desire to use a smartphone as a window management interface shaped users' interaction practices of window management operations. This paper reports how their desire to use a smartphone as a window manager is manifested.

Author Keywords
Augmented Reality; Smartphone; Window Management.

CCS Concepts
•**Human-centered computing** → **User studies;**

Introduction & Background
Recently, AR headsets have become more and more accessible for personal use. They have a more extensive display compared to smartphones, and we expect a growing trend for people to use AR headsets in daily routines

Figure 1: Our envision of the AR headset + Smartphone interface. Users can use it in multitasking scenarios.

Figure 2: The hardware setup of our interview. A magic leap AR headset and a Huawei P20 were used. They were tracked by OptiTrack.

shortly. However, AR headsets do not perform as well as smartphones in terms of precise operations and tactile feedback. Therefore, we suppose that instead of the replacement of smartphones, a fusion interface combining AR headset and smartphone will emerge for multitasking. Users are supposed to manage multiple windows across the two devices intuitively.

Combining an AR headset and a smartphone could be traced from the previous concept of combining palmtop computers and 3D graphics at the time when smartphone and AR headset were not available for consumers [8]. Recent research on combining an AR headset and a smartphone focused on two issues which are combining the display and input capabilities of two devices to promote target acquisition efficiency within one application [9, 10] and using a smartphone to manipulate virtual objects in AR environment for 3D modeling. Millette et al. [12] explored the bimanual interactions for 3D object manipulation and proposed draw-and-drop and touch-and-draw interactions. Wang et al. [16] used a multi-touch gesture on the tablet to manipulate a 3D object in virtual space. Researchers also explored the feasibility of distributing the input of phone screen touch into multiple devices dynamically to compensate for the input inaccuracy of AR headset [1] and provide detailed 2D sketches for 3D modeling [2]. Besides, prior research proposed a 2D and 3D view combined modeling system [13] and explored how a tablet can be used in the virtual reality environment [15].

For multitasking scenario, previous research in the AR environment has studied what window is to be like [7] and what window layout is appropriate to reduce the application switching time [6]. Specifically, Ethereal Planes [5] introduced seven design dimensions that inspire a window layout design in AR space. Besides, Mark Billinghurst et al. [3] found the body-aligned windows more efficient than the HMD-aligned windows on a search task.

In this paper, we envision a multitasking scenario. It enables users to simultaneously handle multiple tasks by distributing various windows across AR virtual space and a smartphone screen. In this way, they can either explore the wide display of an AR headset or the precise input feature of a smartphone. For instance, a user can open several applications in AR like a video player, a web browser, and a social application, and have another chat window open on the smartphone(See Figure 1). He/She can search for information, leave a comment or reply a message freely by dragging the desired window onto the smartphone and operate it; or can expand the current phone view intuitively by dragging it out into the virtual space. Without complex navigation, users can extend the original hierarchical operations on the phone to a broader space, and enjoy the precise input benefits of the smartphone as well.

However, we have little understanding of how users would manage windows in the combining interface. There lacks a direct and efficient interaction method to achieve this envision. At present, users manage virtual windows through mid-air hand gestures [4, 11] or controller in AR and manage the windows on the smartphone screen through touch-based interaction. This frequent switch consumes time and energy and leads to unnaturalness. Cross-device window management needs further investigation.

To understand the requirements and interaction practices of window management operations in the AR headset + smartphone interface, we conduct a semi-structured interview as well as observe users' behaviors. To reinforce the familiarity with the combining interface, we create a simulated interface for participants to explore the typical window management tasks, which is the advantage of our study.

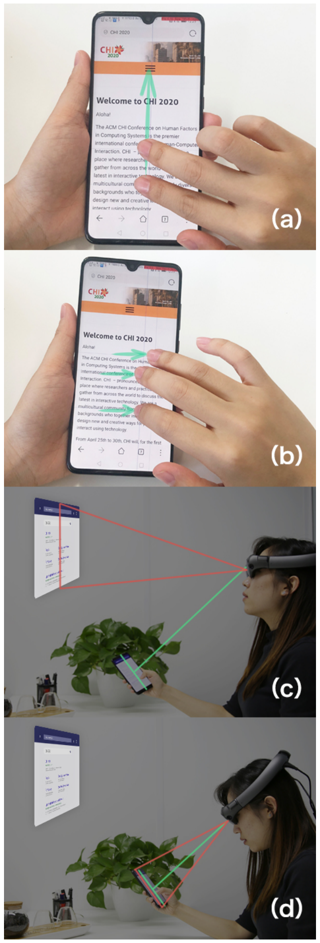

Figure 3: Examples of semantic touch: (a)(b) Perform a 'pull' touch gesture on a smartphone to shift a window into space. (c)(d) Combining head or gaze direction to distinguish touch input.

We found that users tended to treat a smartphone as a window management interface for window management interactions in AR. We present users' interaction preferences in this paper as our key contributions. We design and evaluate our window management operations in future work.

Method

The goal of this formative interview is to better understand users' opinions about window management interactions in the AR headset + smartphone interface. However, users have little experience of using the combining devices. To give them an intuitive impression of the combining interface and inspire them of the interaction methods, we set up a simulated interface that contains a set of windows with an AR headset and a smartphone. Participants can imagine the interaction metaphors and pretend to experience different interaction metaphors in the given tasks through the windows rendered in AR space. Participants can view windows of different alignment modes[5, 9], which are world-aligned windows, head-aligned windows, smartphone-aligned windows, and windows on the smartphone in the combining interface. The world-aligned windows are of different proximity and size as well. Participants need to describe the interaction metaphors with the given tasks under three scenarios and the reasons why they performed such interactions. By providing such assistance(See Figure 2), we hope they can better understand the fusion interface and give more instructive suggestions.

Participant

We recruited and interviewed ten participants(two females, eight males, aged between 20 and 26) from campus. Each interview took about 40 minutes, which was video-recorded under permission. Four of them were quite familiar with AR headset and had AR programming experience. One of them didn't have experience with the AR headset before. The others have once used an AR headset. Moreover, all of them had used a smartphone before and were experts at it.

Apparatus

A participant was given an AR headset (Magic Leap One, tracked by OptiTrack), a smartphone (Huawei P20, tracked by OptiTrack), and a physical keyboard. We rendered six windows (2 sizes * 2 distances user-aligned windows + 1 head-aligned window + 1 smartphone-aligned window) in virtual space to help users understand the combining interface more specifically. A participant controlled the visibility of these windows for different tasks through the physical keyboard freely ('q," w," e," r," t," y' to control each window). These windows were shown to help them imagine how they can do window management tasks. They can call out different windows by pressing the keyboard for different tasks at any time.

Procedure

In the interview, we first informed participants of the multitasking scenarios, namely watching a video, surfing the internet, and chatting and five basic window management tasks, namely open, close, shift, move, and zoom. They were asked to describe and act out the preferable interactions in a think-aloud manner with their imagination. Moreover, they were also encouraged to come up with new windows out of the six kinds of windows and tasks out of the five basic window management tasks as well. Additionally, they needed to explain the reasons why they performed such interactions. During the process, we observed participants' behaviors as well. Our participants were provided with monetary rewards for their participation in our interview.

Data Analysis

We used a grounded theory approach [14] for our analysis. All interviews were conducted in Mandarin, video recorded,

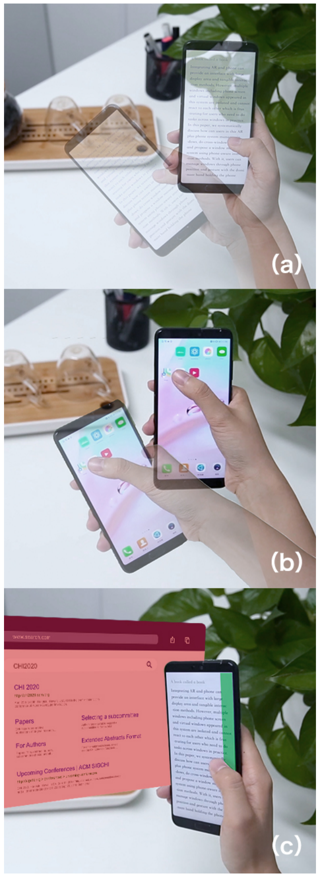

Figure 4: Examples of quick smartphone posture: (a) A shake gesture to pour the window out. (b) Press an icon while performing a shake gesture to open the application in AR. (c) Press the edge and hit the virtual window to close it.

and were later transcribed into Chinese word by word. We used MaxQDA for qualitative analysis along with video playback. Through open coding, 358 codes were produced. Then they were discussed among authors and collaboratively synthesized into higher-level themes through axial coding: semantic touch input, the quick posture of the smartphone, position and movement awareness, and mid-air hand gesture.

Findings

Through the interview, we found that participants would like to use the smartphone interface for window management interactions in AR. They considered it a more efficient and accurate method comparing to the hand gesture because of the accuracy and fatigue level. They employed touch, posture, position, and movement of the smartphone as the inputs of the interface. Participants could select and transform windows in virtual space, and switch windows across devices with the help of its touch, posture, position, and movement. It appeared direct and intuitive to them: *"The smartphone seems a window manager to me. To manage a window in space, I want to use my smartphone at first thought no matter touch on the screen or the smartphone's posture."(P3)* Furthermore, they claimed that with efficient window management interactions, seeking information across applications in AR and smartphone appeared easier, and it would be benefit for the multitasking scenario.

Theme 1: Semantic Touch Input

Semantic touch means to perform touch through a distinct and meaningful hand gesture, or along with the head direction and gaze. Participants paid great attention to touch input for the window management operations because it could afford an explicit tangible feeling and address them to feel more stable. Overall, with the five tasks, all participants mentioned semantic touch for window management interactions. Furthermore, participants implied touch gestures with semantic meanings such as pull and back. They also suggested blending touch input with head movement or gaze, in case it was confusing whether it was the window management or the original function the touch input was employed. They mentioned a lot of semantic-based gestures for window management(See Figure 3). First, to shift the window on the smartphone to a smartphone-aligned window in AR, they aspired to use the pull gesture to pull out the window on the smartphone. Therefore, they used two(or three) fingers or even the palm's side. Swipe them to either side of the screen to pull the window out. To place the window in a user-aligned position, they want to use a throwing gesture to throw the window from the smartphone into space. So, they pressed the screen with multiple fingers and quickly swiped up to shift the window into space. They also claimed that if sliding further on the smartphone, the window would be farther away. Generally, participants expressed great concern on the gestures of pulling in and pushing out for shifting windows. More importantly, when performing a semantic touch, the user wants to avoid conflicts with the original function, so a multi-touch and long-press trigger strategy is given. An example was provided by P4: *"I would like to use two fingers to slide left to pull the window into space. On the one hand, it is intuitive. On the other hand, I think it can be distinguished from the slide operation on the smartphone."*

Besides, we found that touch inputs, along with head and gaze direction, were semantic to participants as well. When the head or eye direction was approximately vertical to the smartphone screen, participants tended to think of touch input as a smartphone operation. Otherwise, participants tended to think of it as a window management operation. *"There is a virtual window. I hope that when I look at the virtual window instead of my smartphone, a click interprets as*

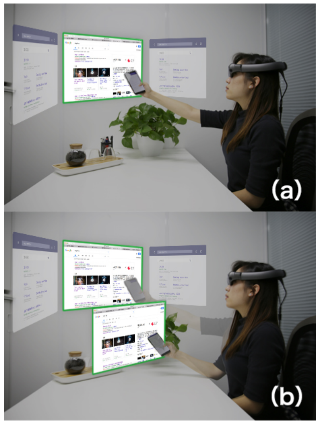

Figure 5: Examples of position and movement aware: (a) Put the smartphone where a virtual window locates to select. (b) Move the smartphone to move the virtual window.

a window management operation rather than a smartphone operation. Otherwise, it will lead to a mistake. For instance, when i look at the virtual window and double click on my smartphone, the window will shut down."(P8)

Theme 2: Quick Posture of the Smartphone

For participants, it was engaging to use the postures of the smartphone to manage windows because they considered it magical. Frequently, participants proposed to perform a discrete operation through the quick smartphone posture, such as a quick tilt, flip over, shake, and knock(See Figure 4). These postures were interpreted as a shift operation or an open-close operation regularly. With a quick posture, a virtual window can open, close, or shift: *"It is nice to open this application in virtual space through a quick shake posture with my smartphone. It appears intuitive."(P3)* Also, P8 told us that it is so natural to perform a flip over posture with a smartphone: *"This flip over posture gives me a hint that the window on screen can be transferred into space."*

Also, the participants revealed the concept of combining the posture and touch input of the smartphone to supplement the window management operations. Since a group of UI elements displayed on the smartphone screen, they suggested to press on a different element and perform a posture of the smartphone to open this specific application window in AR. For example, P10 told us: *"There are icons on the desktop of the smartphone. I can touch this icon while at the same time, shake my smartphone for once. Subsequently, this application will be open in AR."* Similarly, when managing the windows in AR through the smartphone posture, they advised performing it along with touching on a distinct position of the screen, to activate different window management operations. An example was given by p9: *"The screen possesses some unique regions, such as the edges. If I reach the left side of the smartphone and hit a virtual window, it appears on my smartphone. And when I reach the right side and hit a virtual window, the virtual window can be closed."*

Theme 3: Position and Movement Aware

In terms of selecting a window, participants expressed their willingness to place a smartphone where a virtual window approximately located(See Figure 5). They considered it intuitive and simple. After placing the smartphone at a distinguished position, participants can subsequently operate a touch or a posture of the smartphone to manage the virtual window. P1 revealed that he would use a smartphone to approach a virtual window and shut it down: *"For instance, I 'touch' this virtual window through my smartphone and long-press the screen to shut it down."* Furthermore, they preferred to use the non-dominant hand, which held the smartphone to operate a touch or a posture of the smartphone rather than moving the dominant hand because of the convenience and fatigue level. Especially when the combining interface is becoming general: *"I use one hand interacting with my smartphone most of the time. It is simple and convenient. In the combining interface, I want to use my smartphone with one hand as well. So it would be nice if I could still manage windows with only one hand."(P5)*

Also, as virtual windows were in the three-dimensional space, participants were willing to move their smartphones freely to move, rotate, and transform the virtual windows(See Figure 5). One example provided by P4: *"I want to select the window with my smartphone in the real position and then move it. Besides, I hope that there is a three-dimensional window layout thumbnail nearby. With the small range of movement, I can move the window in a real three-dimensional space."* Using the position and move-

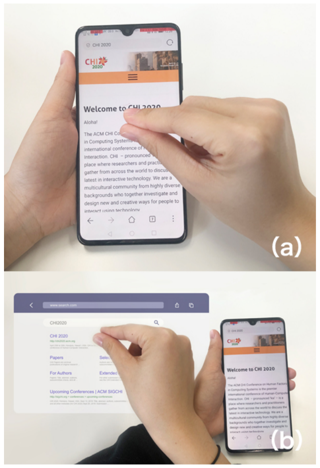

Figure 6: Examples of mid-air hand gesture: (a) Perform a pinch gesture to select the window of a smartphone. (b) Cross hand when performing mid-air hand gesture.

ment of the smartphone to represent the windows in AR is appropriate to the participants.

Theme 4: Mid-air Hand Gesture
Participants also mentioned mid-air hand gestures a lot, such as pinch, grab, and swipe during the interview(See Figure 6). An example provided by P7: *"While using an application on the smartphone, I will perform a pinch gesture above the screen and spread the fingers to place this window in virtual space. "* Moreover, we found an impressive insight that participants were prone to perform distinct gestures on a virtual window to represent different window management operations. For example, P3 told us: *"Now there is a window in front of me. I hope that if I press the window with two fingers, it will be selected and follow my fingers' movement. If I use two fingers to slide down, it will shut down."*

However, they showed great concern about convenience while holding the smartphone simultaneously in some circumstances. Because the non-dominant hand holding the smartphone will block the dominant hand in some ways(See Figure 6) as P1 told us: *"There is a smartphone-aligned window on the left side of my smartphone while I am holding the smartphone with my left hand. It is inconvenient for me to reach this window with my right hand cause the two hands are crossing."* They suggest avoiding such problems when managing windows.

Discussion and Future Work

In the preceding sections, we have illustrated how users would perform window management operations in four aspects. For our participants, they considered the smartphone as a tool and would like to use this tool to manage windows. Also, mid-air hand gestures were suggested. Participants proposed the concept of semantic touch input to distinguish touch input from the original function. Besides, it was highly recommanded to use the posture of the smartphone for discrete window management operations. Also, they encouraged to use the position and movement of the smartphone to interact, which corresponded to continuous operations. Finally, participants would interact through mid-air hand gestures. They proposed to make distinct hand gestures on the position of the window to trigger diverse window management operations.

The combining interface of an AR headset and a smartphone will be widely applied, and it will be helpful in multitasking scenarios. Users will use multiple application windows in the combining interface simultaneously. It is important to manage windows intuitively and efficiently. However, we know little about how to manage application windows in the combining interface. This paper focuses on this area.

In our future work, we are planning to design a set of window management interactions of the combining interface. First, We will extract design dimensions from the observed four findings, and design the interactions based on these dimensions. We will also demonstrate them in a prototype system, which consists of a Magic Leap One AR glasses, a Huawei P20 smartphone, and an OptiTrack tracking system, and evaluate the time, accuracy, fatigue level, and the subjective opinions of users in different window management tasks.

REFERENCES

[1] Mohammed Al-Sada, Fumiko Ishizawa, Junichi Tsurukawa, and Tatsuo Nakajima. 2016. Input forager: A user-driven interaction adaptation approach for head worn displays. *ACM International Conference Proceeding Series* December (2016), 115–122. DOI: http://dx.doi.org/10.1145/3012709.3012719

[2] Rahul Arora, Rubaiat Habib Kazi, Tovi Grossman, George Fitzmaurice, and Karan Singh. 2018. SymbiosisSketch: Combining 2D & 3D sketching for designing detailed 3D objects in situ. *Conference on Human Factors in Computing Systems - Proceedings* 2018-April (2018). DOI: http://dx.doi.org/10.1145/3173574.3173759

[3] M. Billinghurst, J. Bowskill, N. Dyer, and J. Morphett. 1998. An evaluation of wearable information spaces. In *Proceedings. IEEE 1998 Virtual Reality Annual International Symposium (Cat. No.98CB36180)*. 20–27. DOI: http://dx.doi.org/10.1109/VRAIS.1998.658418

[4] Li-Wei Chan, Hui-Shan Kao, Mike Y. Chen, Ming-Sui Lee, Jane Hsu, and Yi-Ping Hung. 2010. Touching the Void: Direct-Touch Interaction for Intangible Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. Association for Computing Machinery, New York, NY, USA, 2625–2634. DOI: http://dx.doi.org/10.1145/1753326.1753725

[5] Barrett Ens, Juan David Hincapié-Ramos, and Pourang Irani. 2014b. Ethereal planes: A design framework for 2D information spaces in 3D mixed reality environments. *SUI 2014 - Proceedings of the 2nd ACM Symposium on Spatial User Interaction* October (2014), 2–12. DOI: http://dx.doi.org/10.1145/2659766.2659769

[6] Barrett M. Ens, Rory Finnegan, and Pourang P. Irani. 2014a. The Personal Cockpit: A Spatial Interface for Effective Task Switching on Head-Worn Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3171–3180. DOI: http://dx.doi.org/10.1145/2556288.2557058

[7] Steven Feiner, Blair MacIntyre, Marcus Haupt, and Eliot Solomon. 1993. Windows on the world: 2 D windows for 3 D augmented reality. In *ACM Symposium on User Interface Software and Technology*. 145–155.

[8] George W Fitzmaurice. 1993. Situated information spaces and spatially aware palmtop computers. *Commun. ACM* 36, 7 (1993), 38–50.

[9] Jens Grubert, Matthias Heinisch, Aaron Quigley, and Dieter Schmalstieg. 2015. MultiFi: Multi-fidelity interaction with displays on and around the body. *Conference on Human Factors in Computing Systems - Proceedings* 2015-April (2015), 3933–3942. DOI: http://dx.doi.org/10.1145/2702123.2702331

[10] S. J. Henderson and S. Feiner. 2009. Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. In *2009 8th IEEE International Symposium on Mixed and Augmented Reality*. 135–144. DOI: http://dx.doi.org/10.1109/ISMAR.2009.5336486

[11] Joon Hyub Lee, Sang-Gyun An, Yongkwan Kim, and Seok-Hyung Bae. 2018. Projective Windows: Bringing Windows in Space to the Fingertip. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, Article Paper 218, 8 pages. DOI: http://dx.doi.org/10.1145/3173574.3173792

[12] A. Millette and M. J. McGuffin. 2016. DualCAD: Integrating Augmented Reality with a Desktop GUI and Smartphone Interaction. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*. 21–26. DOI:http://dx.doi.org/10.1109/ISMAR-Adjunct.2016.0030

[13] Patrick Reipschläger and Raimund Dachselt. 2019. DesignAR: Immersive 3D-Modeling Combining Augmented Reality with Interactive Displays. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces (ISS '19)*. Association for Computing Machinery, New York, NY, USA, 29–41. DOI: http://dx.doi.org/10.1145/3343055.3359718

[14] Anselm Strauss and Juliet Corbin. 1990. *Basics of qualitative research*. Sage publications.

[15] Hemant Bhaskar Surale, Aakar Gupta, Mark Hancock, and Daniel Vogel. 2019. TabletInVR: Exploring the Design Space for Using a Multi-Touch Tablet in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, Article Paper 13, 13 pages. DOI: http://dx.doi.org/10.1145/3290605.3300243

[16] Jia Wang and Robert Lindeman. 2014. Coordinated 3D Interaction in Tablet- and HMD-Based Hybrid Virtual Environments. In *Proceedings of the 2nd ACM Symposium on Spatial User Interaction (SUI '14)*. Association for Computing Machinery, New York, NY, USA, 70–79. DOI: http://dx.doi.org/10.1145/2659766.2659777