Text Selection in AR-HMD Using a Smartphone as an Input Device

Rajkumar Darbar* — Inria Bordeaux, France
Joan Odicio-Vilchez† — Inria Bordeaux, France
Thibault Lainé‡ — Asobo Studio, Bordeaux, France
Arnaud Prouzeau§ — Inria Bordeaux, France
Martin Hachet¶ — Inria Bordeaux, France

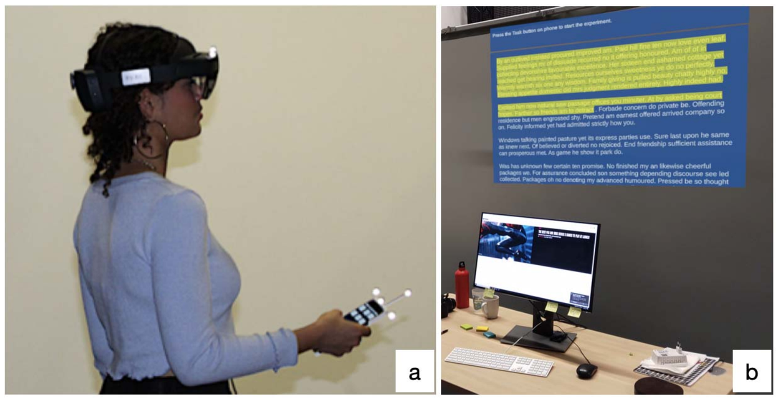

Figure 1: (a) Our implementation consisted of an HoloLens, a smartphone, and an optitrack system. (b) A user is performing text selection in the HoloLens view.

Abstract

Text selection is a common task while reading a PDF file or browsing the web. Efficient text selection techniques exist on desktops and touch devices, but are still under-explored for Augmented Reality Head Mounted Display (AR-HMD). Performing text selection in AR commonly uses hand-tracking, voice commands, and eye/head-gaze, which are cumbersome and lack precision. In this poster paper, we explore the use of a smartphone as an input device to support text selection in AR-HMD because of its availability, familiarity, and social acceptability. As an initial attempt, we propose four eyes-free, uni-manual text selection techniques for AR-HMD, all using a smartphone - continuous touch, discrete touch, spatial movement, and raycasting.

Index Terms: Human-centered computing—Mixed Reality; Human-centered computing—User Interface Design

1 Introduction

The recent focus of the industry on Augmented Reality Head-Mounted Display (AR-HMD), with the development of devices like the Microsoft HoloLens and Magic Leap, made them more and more accessible to us, and their usage is envisioned in our future everyday life. The lack of a physical keyboard and mouse (and in general, the absence of interactive surfaces) with such devices makes text input and text editing difficult and an important challenge in AR research. While text input for AR-HMD has been already well-studied [4], a handful of research focused on editing text that has already been typed by a user. In general, text editing is a complex task, and one of the first steps is to select the text to edit. In this research, we will focus on text selection in AR-HMD.

Performing text selection in AR-HMD commonly uses hand-tracking, eye/head-gaze, voice commands, and handheld controller. However, these techniques have their own limitations. For instance, hand-tracking suffers from achieving character level precision, lacks haptic feedback, and provokes arm fatigue during prolonged interaction. Eye-gaze and head-gaze have the 'Midas Touch' problem [2]. Moreover, frequent head movements in head-gaze interaction increase motion sickness. Voice interaction might not be socially acceptable in public places. In the case of a dedicated handheld controller, users need to carry extra specific hardware with them.

Lately, a number of research efforts have begun to explore the use of smartphones as 6 DoF tracked smartphones in a combination of the AR-HMD because of their availability, familiarity, social acceptance, and tangibility [5]. In this paper, we take a step in this direction and focus on the use of a smartphone to perform text selection in an AR-HMD.

2 Related Work

A few research focused on text editing and text selection in AR. Ghosh et al. presented EYEditor to facilitate on-the-go text-editing on a smart-glass with a combination of voice and a handheld controller [1]. They used voice to modify the text content, while manual input is used for text navigation and selection. The use of a handheld device is inspiring for our work, however, we try not to use a voice command or building our own device. Lee et al. [3] proposed two force-assisted text acquisition techniques where the user exerts a force on a thumb-sized circular button located on an iPhone 7 and selects text which is shown on a laptop emulating the Microsoft Hololens display. They envision that this miniature force-sensitive area (12 mm $\times$ 13 mm) can be fitted into a smart-ring. Although their result is promising, a specific force-sensitive device is required.

We follow the direction of the two papers previously presented and continue to explore the use of a smartphone in combination with

*e-mail: rajkumar.darbar@inria.fr
†e-mail: joan.odicio-vilchez@inria.fr
‡e-mail: tlaine@asobostudio.com
§e-mail: arnaud.prouzeau@inria.fr
¶e-mail: martin.hachet@inria.fr

978-0-7381-1367-8/21/$31.00 ©2021 IEEE
DOI 10.1109/VRW52623.2021.00145

an AR-HMD. While their use for text selection in AR-HMD is still rare, it has been investigated more broadly for other tasks.

3 Proposed Techniques

In the following sections, we describe our smartphone-based eyes-free, uni-manual text selection techniques for AR-HMD - continuous touch, discrete touch, spatial movement, and raycasting. The first two techniques are inspired by mobile touchscreen interaction that is familiar to most users. Whereas, the other two techniques are adapted from typical VR/AR controllers.

To select text successfully using any of our proposed techniques, a user needs to follow the same sequence of steps each time. First, she moves the cursor, located on the text window in an AR display, to the beginning of the text to be selected (i.e., the first character). Then, she performs a double tap on the phone to confirm the selection of that first character. She can see on the headset screen that the first character got highlighted in yellow color. At the same time, she enters into the text selection mode. Next, she continues moving the cursor to the end position of the text using one of the techniques presented below. While the cursor is moving, the text is also getting highlighted simultaneously up to the current position of the cursor. Finally, she ends the text selection with a second double tap.

3.1 Continuous Touch

In continuous touch, the smartphone touchscreen acts as a trackpad. It is an indirect pointing technique where the user moves her thumb on the touchscreen to change the cursor position on the AR display. For the mapping between display and touchscreen, we used a relative touch mode with clutching.

3.2 Discrete Touch

This technique is inspired by the text selection with keyboard shortcuts available in both Mac and Windows OS. In this work, we tried to emulate a few keyboard shortcuts. We particularly considered imitating keyboard shortcuts related to the character, word, and line-level text selection. For example, in Mac OS, hold down the ⇧ and pressing the → or ← extends text selection one character to the right or left. Whereas hold down the ⇧+⌥ and pressing the → or ← allows users to select text one word to the right or left. To perform text selection to the nearest character at the same horizontal location on the line above or below, a user needs to hold down the ⇧ and press the ↑ or ↓ respectively. In discrete touch interaction, we replicated all these shortcuts using directional swipe gestures. Left or right swipe can select text at both levels - word as well as character. By default, it works at the word level. The user performs a long-tap which acts as a toggle button to switch between word and character level selection. On the other hand, up or down swipe selects text at one line above or one line below from the current position. The user can only select one character/word/line at a time with its respective swipe gesture.

Note that, to select text using discrete touch, a user first positions the cursor on top of the starting word (not the starting character) of the text to be selected by touch dragging on the smartphone as described in the continuous touch technique. From a pilot study, we observed that moving the cursor every time to the starting word using discrete touch makes the overall interaction slow. Then, she selects that first word with the double tap and uses discrete touch to select text up to the end position as described before.

3.3 Spatial Movement

This technique emulates the smartphone as an air-mouse for AR-HMD. To control the cursor position on the headset screen, the user holds the phone in front of her torso, places her thumb on the touchscreen, and then she moves the phone in the air with small forearm motions in a plane that is perpendicular to the gaze direction. While moving the phone, its tracked positional data in XY-coordinates get translated into the cursor movement in XY-coordinates inside a 2D window. When a user wants to stop the cursor movement, she simply lifts her thumb from the touchscreen. Thumb touch-down and touch-release events define the start and stop of the cursor movement on the AR display. The user determines the speed of the cursor by simply moving the phone faster and slower accordingly. We applied a CD-gain that has been empirically determined, between the phone movement and the cursor displacement.

3.4 Raycasting

Raycasting is a popular interaction technique in AR/VR. In this work, we developed a smartphone-based raycasting technique for selecting text displayed on a 2D window in AR-HMD. A 6 DoF tracked smartphone was used to define the origin and orientation of the ray. In the headset display, the user can see the ray in a straight line appearing from the top of the phone. By default, the ray is always visible to users in AR-HMD as long as the phone is being tracked properly. In raycasting, the user needs to do small angular wrist movements for pointing on the text content using the ray. Where the ray hits on the text window, the user sees the cursor there. Compared to other proposed methods, raycasting does not require clutching as it allows direct pointing to the target. The user confirms the target selection on the AR display by providing a touch input (i.e., double tap) from the phone.

4 Implementation

To prototype our proposed interaction techniques, we used a Microsoft HoloLens 2 and an OnePlus 5 smartphone (see Figure 1). For spatial movement and raycasting interactions, real-time pose information of the smartphone is needed. An OptiTrack system with three Flex-13 cameras was used for accurate tracking.

In our software framework, the AR application running on HoloLens was implemented using Unity3D (2018.4) and Mixed Reality Toolkit. To render text in HoloLens, we used TextMeshPro. A Windows 10 workstation was used to stream tracking data to HoloLens. All pointing techniques with the phone were also developed using Unity3D. We used UNet library for client-server communications between devices over the WiFi network.

5 Conclusion and Future Work

In this research, we proposed four text selection techniques for AR-HMD using a smartphone as an input device. It is not yet clear which technique will allow users to select text fast and accurately. Although our initial informal test with participants tends to show that continuous touch seems promising for fast and accurate text selection as it is easy to learn and users are already familiar with mobile touch interaction. Of course, this needs to be formally tested in a proper user study that will be the focus of our future work.

References

[1] D. Ghosh, P. S. Foong, S. Zhao, C. Liu, N. Janaka, and V. Erusu. Eyeditor: Towards on-the-go heads-up text editing using voice and manual input. In *CHI '20*. doi: 10.1145/3313831.3376173

[2] M. Kytö, B. Ens, T. Piumsomboon, G. A. Lee, and M. Billinghurst. Pinpointing: Precise head- and eye-based target selection for augmented reality. In *CHI '18*. doi: 10.1145/3173574.3173655

[3] L. Lee, Y. Zhu, Y. Yau, T. Braud, X. Su, and P. Hui. One-thumb text acquisition on force-assisted miniature interfaces for mobile headsets. In *IEEE PerCom '20)*. doi: 10.1109/PerCom45495.2020.9127378

[4] 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 ISMAR*, pp. 279–288, 2019. doi: 10.1109/ISMAR.2019.00026

[5] F. Zhu and T. Grossman. Bishare: Exploring bidirectional interactions between smartphones and head-mounted augmented reality. In *CHI '20*. doi: 10.1145/3313831.3376233