On-Surface Gesture Interaction on Google Cardboard

GestOnHMD: Enabling Gesture-based Interaction on Low-cost VR Head-Mounted Display

Yuhe Nie
yn2273
yn2273@nyu.edu

Abstract

In the realm of Virtual Reality (VR) technology, notable advancements have been witnessed, most prominently in the form of cutting-edge devices like Quest3 and Apple Vision Pro. Although these devices have garnered substantial attention, their high costs present a barrier to widespread adoption. In contrast, affordable options such as Google Cardboard, priced at less than $20, have democratized access to immersive VR experiences. However, these budget VR devices suffer from limited interactivity, rendering the VR encounters akin to passive viewing rather than interactive engagement. Addressing this limitation, our project initiative aims to enhance the functionality of Google Cardboard by incorporating on-surface gesture-based interaction, drawing inspiration from the established GestOnHMD research. Leveraging stereo microphones, we plan to discern unique sound patterns generated by finger interactions with Google Cardboard. This auditory data will be pivotal in simulating gesture interactions within the VR environment. Furthermore, we will develop sophisticated real-time interactive software specifically tailored for Google Cardboard. Through meticulous experimentation with this software, we intend to validate its effectiveness and explore diverse avenues for further research and development within the realm of affordable VR technologies.

1. Introduction

In recent years, the immersive VR landscape has experienced significant commercial growth. Alongside the prominence of high-end VR platforms exemplified by devices like Quest3 and Apple Vision Pro, the widespread adoption of smartphone-based Virtual Reality Head-Mounted Displays (HMDs) has been notable, a trend catalyzed by the introduction of Google Cardboard [3] in 2014. Despite the advancements in smartphone capabilities, especially in high-resolution rendering, the interactivity offered by low-cost mobile VR HMDs remains constrained. These HMDs primarily rely on motion sensors to translate the user's hand movements into virtual viewing directions within VR applications. In the initial iteration of Google Cardboard, users could employ an attached magnet, detectable by the in-phone magnetic sensor, to initiate input events. Subsequently, in the second-generation Google Cardboard, a small lever button interfacing with the phone's touchscreen was integrated to enhance button-like inputs. The evolution of input mechanisms in these devices underscores the ongoing pursuit for enhanced user interaction within affordable VR technology.

Numerous approaches have been explored in the realm of research to enhance the interactivity of Google Cardboard Virtual Reality Head-Mounted Displays (VR HMDs). These methods range from voice-based interactions [5] and utilizing the mobile phone's rear camera [1] to eye-tracking techniques [6,8]. While these techniques augment the range of interactions available in VR HMDs, they are beset by challenges such as privacy concerns and the need for additional hardware components.

This research project commences with the replication and refinement of GestOnHMD [4], an innovative gesture-based interaction technique rooted in deep learning methodologies. GestOnHMD utilizes the acoustic signals captured by the built-in stereo microphones in commodity smartphones to recognize gestures on the front, left, and right surfaces of a paper-based mobile VR headset. The method involves training a deep learning model to discern various states of finger interactions with a Google Cardboard surface, thereby achieving gesture recognition effects. Significantly, this approach capitalizes on the inherent nature of Google Cardboard as a cardboard shell, obviating the necessity for extensive object detection models or additional hardware. The achieved results are notable, with a high accuracy rate of 98.2% for both gesture detection and surface recognition, and 97.7% for gesture classification.

Subsequent phases of this project involve the development of a dedicated game application for VR HMDs, providing a practical context to validate the effectiveness of these interactions. Concurrently, our research endeavors continue to explore diverse avenues, delving deeper into the domain to formulate innovative interaction schemes and expand the breadth of research content in this promising field.

2. Related Work

Related work can't be fully shown in 2-page limitation.

One of the traditional interaction methods of mobile VR is using sensors to detect the tilt and rotation of the phone. This can be used to detect head rotation and shaking. In addition, existing studies have also proposed deep learn-based spatial perception, gesture detection, and voice detection. Ahuja et al. developed EyeSpyVR [2], which uses the eye-tracking technique. Ishii et al. proposed FistPointer [7], detecting the gestures of thumb pointing and clicking.

3. Design and Implementation

1. **Definition.** Define a set of referents or effects of actions (Eg. gaming, browsing, text editing).

2. **Data collection.** Collect different acoustic signals as well as label them.

3. **Model Training.** Based on the existing data, training the classification model.

4. **Application Development.** Integrate the model in the project that is based on Unity Engine and implement the function we require. Also based on the interaction, develop a real game application that utilizes these functions.

5. **Experiment.** Invite users to test the application. Collect results and feedback.

6. **Modification.** Generation new ideas during the research and reproduction and adapting them in the project.

4. Demonstration Plan

The project contains several steps:

1. **Research data integration and collection.** Read relevant in recent years and collect training training data and feasible methodologies.

2. **Training Deep Learning Model.** Training a deep learning classification model that takes acoustic signals as input and outputs the gesture type.

3. **Reproducing functionality.** Integrate the model in Unity and provide the functionality of different interaction behaviors.

4. **Developing VR HMDs game application** Implement a real-time application that uses Google Cardboard as verification.

5. **Explore potential research possibilities.** Further, explore potential research area and come up with new ideas.

6. **Final Project Thesis.** Finish the whole project, including the reference, paper, application, and further exploration.

5. Timeline

Oct 25	Research data integration and collection
Nov 05	Training Deep Learning Model
Nov 15	Reproducing functionality
Nov 30	Developing VR HMDs game application
Dec 10	Explore potential research possibilities
Dec 14	Final Project Thesis

Table 1. Project TimeLine Details

References

[1] Karan Ahuja, Chris Harrison, Mayank Goel, and Robert Xiao. Mecap: Whole-body digitization for low-cost vr/ar headsets. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST '19, page 453–462, New York, NY, USA, 2019. Association for Computing Machinery. 1

[2] Karan Ahuja, Rahul Islam, Varun Parashar, Kuntal Dey, Chris Harrison, and Mayank Goel. Eyespyvr: Interactive eye sensing using off-the-shelf, smartphone-based vr headsets. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(2):1–10, 2018. 2

[3] Google Cardboard arvr.google.com/cardboard/, 2014. 1

[4] Taizhou Chen, Lantian Xu, Xianshan Xu, and Kening Zhu. Gestonhmd: Enabling gesture-based interaction on low-cost vr head-mounted display. *IEEE Transactions on Visualization and Computer Graphics*, 27(5):2597–2607, 2021. 1

[5] Jiaxi Gu, Zhiwen Yu, and Kele Shen. Alohomora: Motion-based hotword detection in head-mounted displays. *IEEE Internet of Things Journal*, 7(1):611–620, 2020. 1

[6] Hiroyuki Hakoda, Wataru Yamada, and Hiroyuki Manabe. Eye tracking using built-in camera for smartphone-based hmd. In *UIST '17 Adjunct*, page 15–16, New York, NY, USA, 2017. Association for Computing Machinery. 1

[7] Akira Ishii, Takuya Adachi, Keigo Shima, Shuta Nakamae, Buntarou Shizuki, and Shin Takahashi. Fistpointer: target selection technique using mid-air interaction for mobile vr environment. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 474–474, 2017. 2

[8] Junichi Shimizu and George Chernyshov. Eye movement interactions in google cardboard using a low cost eog setup. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, UbiComp '16, page 1773–1776, New York, NY, USA, 2016. Association for Computing Machinery. 1