

SWAG Demo: Smart Watch Assisted Gesture Interaction for Mixed Reality Head-mounted Displays

Hyung-il Kim^{1*}, Juyoung Lee¹, Hui-Shyong Yeo², Aaron Quigley², Woontack Woo¹

¹UVR Lab, KAIST, Republic of Korea

²SACHI, University of St Andrews, Scotland, UK

ABSTRACT

In this demonstration, we will show a prototype system with sensor fusion approach to robustly track 6 degrees of freedom of hand movement and support intuitive hand gesture interaction and 3D object manipulation for Mixed Reality head-mounted displays. Robust tracking of hand and finger with egocentric camera remains a challenging problem, especially with self-occlusion – for example, when user tries to grab a virtual object in midair by closing the palm. Our approach leverages the use of a common smart watch worn on the wrist to provide a more reliable palm and wrist orientation data, while fusing the data with camera to achieve robust hand motion and orientation for interaction.

Keywords: Augmented Reality, Wearable Computing, 3D User Interfaces, Hand Interaction, Virtual 3D Object Manipulation.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User interfaces—Input interaction styles

1 INTRODUCTION

Recent development of commodity head-mounted displays (HMD) enabled many users to experience virtual reality (VR) and augmented reality (AR). Popular devices such as Oculus Rift, HTC Vive, HoloLens, Meta and Google glass. While HMD and the mixed reality continuum become more common, interacting with these devices remain difficult.

For gesture interaction and object manipulation in AR and VR environment, the most intuitive approach is to utilize the users' real hand [1,3]. It is more direct and natural compared to using a mouse, a joystick or a pointer.

Common input devices for VR are based on hand-held controllers tracked by multiple external cameras placed around the room (e.g., Oculus Constellation and HTC Light House). This provides accurate and robust 6 DoF tracking for interaction at the cost of portability. Whereas in mobile VR, usually only 3 DOF controller are used (Daydream controller).

There is great amount of research on hand and finger tracking. However, it remains a very challenging problem due to the complex articulations of hand and finger pose, self-occlusions problem and fast motion. Therefore, even the state-of-the-art head-mounted displays like HoloLens and Meta 2 do not support full hand pose tracking. They only support basic hand position tracking with detection hand state (open or close) or finger tapping. Third parties such as the Leap Motion provide a camera add-on kit that can be mounted on the HMD. It supports full hand and finger tracking with rather impressive results. Nonetheless, the remains limitation, especially when the hand is closed where the tracking starts to fail due to the aforementioned self-

occlusions problem. This problem worsens when the hand is closed and rotated, for example when the users grab a virtual object rotate it in midair.

We set our focus to overcome this limitation. Indeed, there are already solutions such as wearing data glove, exoskeleton or using multiple markers with expensive multi-cameras setup. Yet, these are often unpractical due to the high cost and complex setup. On the other hand, smartwatches are getting popular and is rather low cost. We can leverage the fact that users are already wearing smartwatch on their wrist, which contains inertial measurement unit (IMU) and provides fast and high fidelity motion and orientation data for the hand. Taking the data and fusing with the data from camera, we can achieve robust hand motion and orientation tracking.

This demonstration allows the user to experience robust gesture interaction with virtual object using their hand. It is unique because the user can manipulate virtual object with their hand without suffering from intermittent hand tracking failure of common system. The novelty lies in the sensor fusion technique of our system. We believe it will draw a crowd as it is easy to setup and experience (just wear HMD and smartwatch).

2 SYSTEM DESIGN

We assume the user wearing a smartwatch and see-through HMD with near-range depth camera (Figure 1). Near-range depth camera is used for hand localization, and can be also used for generating occlusion mask for improving depth perception. After calibration, wrist worn smartwatch is used for tracking hand orientation. We assume HMD and smartwatch give their orientation in their reference frame using their own internal tracker, and also assumes that HMD and the mounted camera are calibrated.

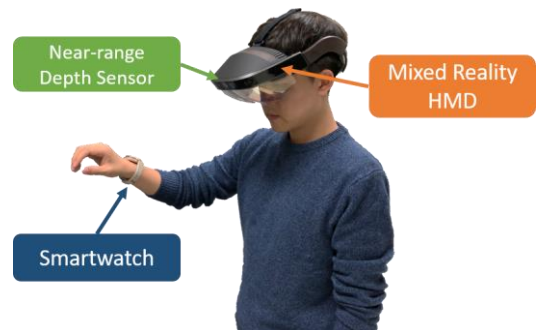


Figure 1: Hardware configuration of proposed system

2.1 Smartwatch and HMD Orientation Calibration

Since smartwatch orientation tracker and HMD orientation tracker have independent reference frame, calibration between smartwatch and HMD is needed (see Figure 2). Let the orientation of watch relative to its reference frame R_{watch} , and the orientation of HMD relative to its reference frame R_{HMD} . For calibration

between smartwatch's reference frame and HMD's reference frame, we use 2D marker tracking. At first stage of system, smartwatch displays 2D marker, and user looks at 2D marker using HMD. Then, marker tracking module detects marker from RGB image, giving relative orientation between marker and HMD, $R_{\text{marker/HMD}}$. At that moment, we can compute rotation offset between smartwatch's reference frame and HMD's reference frame as follows.

$$R_{\text{offset}} = R'_{\text{hmd}} R_{\text{marker/HMD}} R'^{-1}_{\text{watch}} \quad (1)$$

Here R_{HMD} ' and R_{watch} ' denotes HMD and smartwatch rotation at the moment of marker tracking. After 2D marker is tracked, we can compute smartwatch's orientation relative to HMD as follows

$$R_{\text{watch/HMD}} = R^{-1}_{\text{HMD}} R_{\text{offset}} R_{\text{watch}} \quad (2)$$

Here, R_{HMD} and R_{watch} denotes HMD and smartwatch rotation in real time.

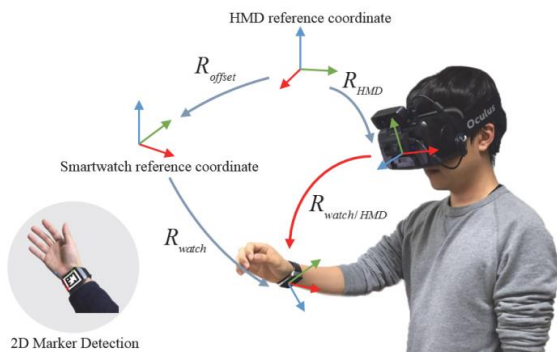


Figure 2: Smartwatch-HMD orientation calibration.

2.2 6-DOF Hand Tracking

Robust 6-DOF hand tracking is very important for 3D interaction in HMD environment, so robust position tracking and orientation tracking are needed. We use both near-range depth camera and smartwatch for position and orientation tracking. We observed that existing vision-based hand tracker [2] gives acceptable position of hand robustly, even if they fail to track hand orientation. Because hand tracker gives 3D position of hand relative to depth camera, one can compute 3D position of hand relative to HMD, assuming that system knows the relationship between near-range depth camera and HMD. For orientation tracking, we assume that watch rotation is same with hand rotation. Proposed system uses orientation of smart-watch relative to HMD, from Section 2.1.

2.3 Natural Interaction for Object Manipulation

We use grasp-to-manipulate metaphor for 3D object manipulation: i) user grasps the object to hold, ii) user moves and rotates the object while grasping, and iii) user releases hand to release the object. For our proposed system, grasp gesture is detected by existing hand tracker, like hand position tracking.

For our object manipulating system, we assume 3D AR space with various objects. When user grasps object with hand, the object in close range is attached to user's hand, and moves and rotates along with hand movement. User releases hands to release holding object.

3 IMPLEMENTATION AND RESULTS

We implemented two versions of our proposed system using commodity devices. The first version uses an Oculus Rift Development Kit 2 with Ovrvision stereoscopic RGB camera for video see-through. For near-range depth camera, we used Creative Senz3D camera. For smartwatch, we used Samsung Gear Live, which uses Android wear as software platform. For the second version, we use a Meta 2 optical see-through HMD.

We implemented our initial prototype in Unity game engine. Our preliminary implementation of object manipulation can be seen at Figure 3. Result showed robust object manipulation with low tracking failure. We also designed simple user test for measuring completion time of five object manipulation tasks, by using existing hand tracker [2] based pinching method and proposed method. Results show that proposed method shows much faster completion time (29.6s in average) than using existing hand tracker [2] (110.1s in average).



Figure 3: Object manipulation using proposed method (version 1).



Figure 4: Object manipulation using proposed method (version 2).

4 CONCLUSION AND FUTURE WORK

We presented robust 6-DOF hand tracking by fusing near-range depth camera image with smart watch sensor information. First, we enabled fast and robust 6-DOF hand tracking by using vision-based hand localization with smartwatch based hand orientation tracking. Also, we introduced method to calibrate orientation between smartwatch and HMD. We also implemented natural object manipulating system using our 6-DOF hand tracker with grasping method.

REFERENCES

- [1] T. Ha, S. Feiner, and W. Woo, "WeARHand: Head-worn, RGB-D camera-based, bare-hand user interface with visually enhanced depth perception," in Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on, pp. 219-228, 2014.
- [2] S. Melax, L. Keselman, and S. Orsten, "Dynamics based 3D skeletal hand tracking," in Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games, pp. 184-184, 2013.
- [3] T. Piumsomboon, D. Altimira, H. Kim, A. Clark, L. Gun, and M. Billinghurst, "Grasp-Shell vs gesture-speech: A comparison of direct and indirect natural interaction techniques in augmented reality," in Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on, pp. 73-82, 2014.