Annotate All! A Perspective Preserved Asynchronous Annotation System for Collaborative Augmented Reality

Po Yen Tseng[1] Harald Haraldsson[1] Serge Belongie[1,2]

[1]Cornell Tech [2]Cornell University

Abstract

Annotation is a crucial technique for guiding users through a complex sequence of steps such as operating a machine or familiarizing oneself with a new environment. In this work, we developed an annotation system for asynchronous collaboration on an optical see-through head-mounted display. This system offers a set of capabilities for placing, playing back, and organizing virtual annotations to real-world objects. Perspectives of the annotator are recorded to preserve extra context information, allowing the viewer to review the content from any angle. The resulting system lays the foundation for content creation tools in head-mounted augmented reality displays, moving a step closer to an easy-to-use interactive annotation tool that allows general users to create and share meaningful content in mixed reality space.

Figure 1. The Magic Leap One, an Optical See-Through, Head-Mounted Display (OST-HMD). We use the eye tracking and head pose tracking the lightwear provides and the 6DOF control to create visual annotations. To add more context information into asynchronous communication, we record spatial audio.

1. Introduction

In this paper, we describe an asynchronous annotation system, exploring the scenario of creating virtual annotations in the real world on an optical see-through head-mounted display (OST-HMD). The annotator can place annotation marks on physical objects, draw freely in space, and record voice annotations. In addition, to reproduce the perspective of the annotator during annotation, head pose and fixation points of the annotator are tracked and recorded. Viewers can view the annotations from every angle asynchronously.

Annotation is an essential interaction method in daily life. Humans leave physical signs, sticky notes, and illustrations in the real world to share information. Meanwhile, we put text, images, audio in digital formats to communicate asynchronously in the virtual world. Annotating with physical artifacts in the physical world and with digital ones in digital worlds seems straightforward, but is it possible to attach virtual annotation to the physical world? The emergence of augmented reality (AR) technologies blurs the boundary between virtual and real worlds, and thus creates potential opportunities to change the way we share information. Asynchronous annotation is widely adopted to share information across time. For instance, ”Caution, wet floor” signs, comments on code projects, and tags for photos allow viewers to retrieve important information at the moment they come across the annotations. Annotations in mixed reality space augment real-world objects with knowledge that is difficult to attach due to physical constraints. For instance, it is hard to provide comprehensive operation guides to a complicated machine because of the limited space. One can hardly attach long enough user manuals to explain every function a computer can do. For the virtual 2d spaces, such as websites, viewers can navigate through detailed information by manipulating information displayed in a logical sequence. In the world of 3d spatial computing, users can move around in real space to interact with virtual information from all perspectives.

The potential for real-world human interaction has been increasingly studied in the handheld AR space thanks to the

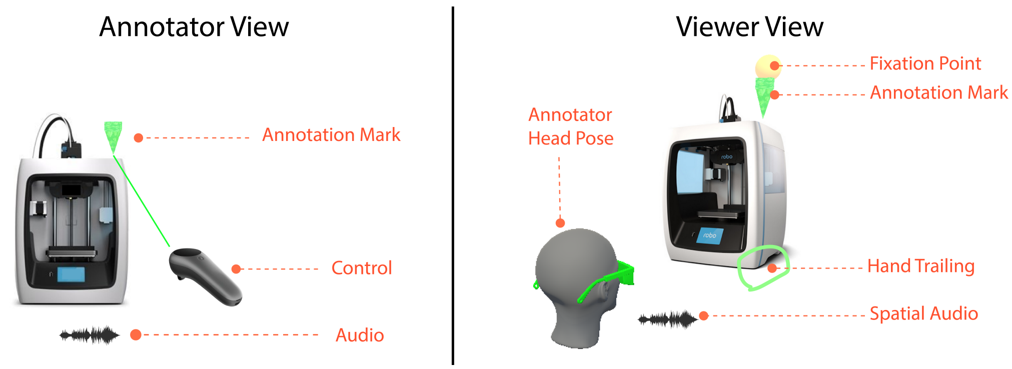

Figure 2. We implement two modes in the system: annotator mode and viewer mode. In the former, the annotator records audio, head pose, fixation point, and hand trails by the control; in the latter, a human head model and fixation point indicates the perspective of the annotator when creating annotations.

growing popularity of easy-to-use AR development frameworks such as Vuforia, ARkit, and ARCore. On the other hand, interaction via optical see-through head-mounted displays (OST-HMDs) is still in its early stage. Especially in the field of utilizing HMD-OST for asynchronous collaboration[6]. These devices offer more immersive experiences and are embedded with multi-sensory detectors offering functionalities such as eye and hand tracking. Combined with 6 degrees of freedom (6DOF) handheld controllers, OST-HMDs create new possibilities for interacting with the physical world. With more ready-to-use OST-HMD devices on the market, we anticipate the need to understand interaction techniques in this field. Thus, we create an annotation system on OST-HMD to explore potential ways of asynchronous collaboration.

2. Related Work

Most current research on annotation systems in the mixed reality space focuses on virtual reality (VR), but attention to the augmented reality (AR) space has been growing, with a focus on annotation for professional education and spatial awareness purposes.

For virtual reality, Poupyrev's Virtual Notepad *et al.* [5] demonstrates handwriting in virtual space and discusses how to create an intuitive VR writing experience. The Harmon *et al.*'s Virtual Annotation System [3] showcases a complete set of tools for making, finding, and reviewing annotations, and shows how voice annotations in a virtual space could be used.

For Augmented Reality, handheld AR for collaboration has been gaining popularity in the medical field. In a system developed by Andersen *et al.* [1], a trainee surgeon looks through a tablet to view the operation field with annotations. Their results show a decrease in placement error when using the AR system, and suggests that using the window metaphor and providing guidance directly onto the field of view of the trainee can improve surgical performance. Langlotz *et al.* [4] let users create textual descriptions and voice annotations on panoramic images on a mobile phone, and then browse the annotations asynchronously. Chen *et al.* [2]introduce an AR system that supports both asynchronous and synchronous collaboration. An interaction model between OST-HMD user and display interface user could potentially help scaling up AR collaboration system by requiring only one party to have the headset. Whereas our project focuses on creating an asynchronous collaboration experience by having the annotator and viewer share the same OST-HMD.

3. Methods

In annotations, perspective information help foster clearer communication. For instance, a technician might be staring at the indicator light above the valve that is being operated. We would naturally follow the point of view of this technician and learn that the light is related to the valve. Smith *et al.* [6] suggest that by capturing the annotators actions, we could provide viewers a broader range of information to understand the context. To visualize the perspective of the annotators, we capture the head pose and fixation points of the annotator in real-time. Additionally, to revisit how annotators interact with real-world space, hand trails and voice are tracked and recorded. These substantial building blocks empower viewers to view the annotation from every angle for an effective asynchronous collaboration experience.

3.1. Hardware

We designed our system on the Magic Leap One OST-HMD for the plentiful interactivity features it provides. Magic Leap One consists of three parts: Lightwear, Light-

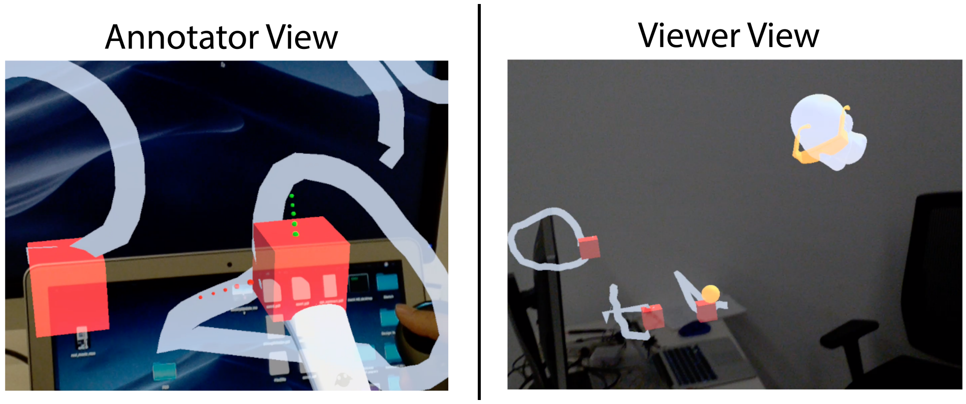

Figure 3. In the annotator view, trails are recorded by control; the viewer is able to inspect the annotation from any angle.

pack, and Control which is shown in Figure 1.

- Lightwear: A headset that collects input including eye tracking, gesture, head pose, and environmental mapping to produce the immersive AR experience.
- Lightpack: An engine drives spatial computing, enabling the machine to see what users see.
- Control: A handheld controller that provides 6 degrees of freedom, bumper, trigger, and touchpad gesture recognition, and raycasting.

3.2. Implementation

The system has two modes: the annotator mode and the viewer mode, which is illustrated in Figure 2. In order to interact with real-world objects, the annotator wears the HMD to scan the surroundings, constructing the meshes that will be overlaid on the field of view. The annotator then casts rays by pointing the controller toward the target location. When the raycast hit is observed, the user leaves an annotation mark on the collided position. Two raycasting method is used for separate purposes. One is to detect collision information from the world mesh generated by the real world, and the other is to interact with the virtual objects such as the added annotation marks and the control menu.

When the annotator leaves an annotation mark, a recording begins and is stopped when the specified button is pressed. Each annotation mark contains a set of recorded objects, including position, head pose, fixation points, voice, and hand trail of the annotator. Shown in Figure 3, multiple annotation marks can be added to the space to create a natural experience of annotation. In the viewer mode, a viewer can raycast on the annotation mark to retrieve the corresponding annotation. A simple UI next to the annotation mark allows for playback control.

4. Conclusion

Our work demonstrates a virtual annotation method for asynchronously sharing information about the real world. Despite the challenges presented by OST HMDs, such as limited field of view and ergonomic factors, the annotation system represents a step forward in allowing non-technical users to create visual and audio annotations without prior training. Viewers can easily review the recorded drawings, head poses, and fixation points by operating the simple playback control of each annotation mark. Furthermore, viewers can move freely in the real-world space to view the recorded annotations in different perspectives. We see our work as a stepping stone toward more complicated content creator tools in the mixed reality space. To evaluate how much could preserving the perspective of the annotator help in collaboration and iterating to a easy-to-use system, user studies is needed in the future.

References

[1] D. Andersen, V. Popescu, E. Cabrera, and A. Shanghavi. Medical telementoring using an augmented reality transparent display. *Surgery*, 159(6):1646–1653.

[2] H. Chen, A. S. Lee, M. Swift, and J. C. Tang. 3D Collaboration Method over HoloLens and Skype End Points. pages 27–30, 2015.

[3] R. Harmon, W. Patterson, W. Ribarsky, and J. Bolter. The virtual annotation system. pages 239–245, 2002.

[4] T. Langlotz, D. Wagner, and A. Mulloni. Online Creation of Panoramic Augmented Reality Annotations on Mobile Phones. *IEEE Pervasive Computing*, 11:56–63, 2012.

[5] I. Poupyrev, N. Tomokazu, and S. Weghorst. Virtual Notepad : Handwriting in Immersive VR HIT Lab , University of Washington. 1998.

[6] R. T. Smith, S. V. Itzstein, and B. H. Thomas. Challenges for Asynchronous Collaboration in Augmented Reality. *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pages 31–35, 2016.