I See What You See: A System for Indicating Shared Intentional Engagement while Cruising

Team 2: Auejin Ham, Eunhwa Song, and Seunghwan Kim

Abstract

We propose a new system that provides an effective real-time interaction utilizing a novel attention cueing method regarding distance-based interaction in a possible collocated AR environment. Describing and sharing one's interest upon certain distant objects to the others is one of the most common interactions in a collocated environment. Meanwhile, due to the limited efficiency and information bandwidth from traditional methods (e.g., verbal communication and pointing gestures), many services have been developed to provide location and its contextual information to users, in the form of social networks and online maps (such as the ones provided by Naver or Google). However, the shift of attention from the physical surface to a hand-held digital device inherently generates Information Access Cost (IAC) by its nature, which reduces efficiency in the process of information acquisition. Furthermore, collaborative interactions with personalized hand-held devices, by which users are obliged to notify each one's interest towards real-world objects to the respective individuals in vicinity, is technically limited regarding the gap between the virtual space and the physical world. So far, numerous studies have utilized gaze visualization to improve interaction performance in collaborative environments. In addition, recent studies have started to acknowledge the importance of considering context-awareness not only for augmenting selected target's information, but also for the gaze cue visualization. On the other hand, interaction techniques efficiently benefiting from gaze cues are yet to be deeply studied. Therefore, this project investigated the novelty of spatial information augmented in a HoloLens 2 device based on each user's eye and head position; once two users reach a common region of interest, a notification in the form of square stroke upon the target is projected to the Hololens 2.

Introduction

Needs

In a physical world, people often need to describe distant objects to others to arrive at a common decision. This sort of situation often occurs in collocated environments requiring efficient and timely collaboration, not to mention for a group of people even deciding which restaurant to visit. For example, in the mid-2000s where limited amounts of digital media were presented to obtain and share spatial information, people created huge signboards to promote their businesses, which is deemed commercially conventional. Especially in South Korea, it is not difficult to find unusually large signboards of restaurants, cafes, and other types of businesses on streets. Many sign boards are made large to contain information about even the restaurant's label, menu type, and somehow the owner's photo and name.

Without any assistance from digital modalities, two main methods are mainly utilized to aid distant target selection procedure: verbal communication, and pointing gesture(s). However, verbal communication has its own limitation of dimensionality with the duration while transcribing visual features into verbal modality as its bottleneck. Pointing gestures to locate a distant object has been an ordinary solution in an environment with relatively sparse target density, but eye-hand visibility conflicts still remain to be solved for a precise spatial indication in densely placed targets.

Even with supportive technologies, there is still no means to immediately obtain and share spatial information about a specific spot in real life street scenes. People walk down the street, stop in the middle of the street, look down at their cell phones, and search for information through a search engine. Despite the bothersome procedures, people can hardly acquire information about their destination while taking a look at its digitally projected coordinate. This repetitive transition of grasping physical surroundings and a hand-held digital device is considerably burdensome, inefficient, and inconvenient.

To alleviate the aforementioned situational factors, a new processing medium is needed to instantly obtain and share information in a physical space. Currently available interfaces lack their performance in obtaining and sharing spatial information instantly and easily. It still takes a considerable amount of time to get and share spatial information in currently popular applications (e.g., Map, Mobile SNS, KakaoTalk), and the medium of information also needs to be changed. In short, the user's Information Access Cost (IAC) increases as the user puts a great deal of mental effort to retrieve spatial information through a hand-held device. Information access cost (IAC) refers to the cost (commonly physical or mental effort) associated with retrieving information from a source or area of interest (Wickens and McCarley, 2008). Therefore, we propose providing augmented spatial information via the user's HoloLens2. Through this approach, we expect to significantly reduce the process gap between two regions of respective user's interests.

User Scenario

The user's situation is to walk along the street with a friend and choose a fine restaurant with good reviews. Both users are wearing HoloLens2 and need to get information about the restaurant.

1. The user wearing Hololens2 walks down the street with a collaborative user to choose a restaurant.

2. Information about where the collaborative user is looking is augmented at the user's HoloLens.

3. Information about where the user is looking is augmented on the collaborative user's HoloLens with different colors.

4. When points of interest (POI) overlap each other, the POI is augmented in a square stroke method.

Related Work

Numerous research studies tried to figure out the engagement of a person from their body posture. One of their most famous findings is that people tend to look at objects with which they interact. For example, Maglio et al. (2000) found out that people performing speech-interaction with objects tend to direct their heads towards the target objects before addressing them. Naturally, several studies to utilize the collected human engagement also have been proposed. Reiner et al. (2020) investigated the impact of information presentation in navigation displays. They developed an innovative Augmented Reality(AR) navigation display called Mirror in the Sky (hereafter, MitS), which reflects the topographic formation of the terrain in front of the user. They compared the Mirror in the Sky to an existent electronic map for way confirmation and threat location tasks. The results indicate that the electronic map had benefits over MitS on some measures such as accuracy, duration, and subjective workload, but participants with some VR experience performed route confirmation more precisely with MitS than a map.

Visualizing Gaze Cue in Collaborative Environments

Numerous research studies have shown that sharing gaze cues improves performance of given collaborative tasks in virtual and augmented reality environments. Piumsomboon et al. (2017) visualized gaze information as a view frustum of users and showed context of gazed objects simply as a fixed 2D GUI on a viewing screen. Even without the context not localized near the source object, performance and satisfaction was higher where visualization was present. Gupta et al. (2016) and Yao et al. (2018) utilized one of the most common de-facto methods for visualizing gaze cue as a circular marker color-coded for each participant for a collaborative work and teaching environment. They collected qualitative and quantitative data revealing the increase of performance after visualizing gaze cues. Recent research studies proposed by Bai et al. (2020) and Osers et al. (2020) studied the way to effectively visualize information of gazed objects under a spatial context of a given 3D scene. Their studies proved that considering spatial context and optimizing the location of visualized information of gazed objects can improve the efficiency of utilizing that information.

Visualizing Information of Gazed Target

Quickly revealing information at a glance in the foreground

Visualizing information about the gaze-overlapping targets is one of the most crucial issues in our project. In our project, when the collaborators stare at the same signboard, virtual information that indicates the user is looking at the same area is quickly revealed at a glance at the moment the gaze overlaps. Pfeuffer et al. (2021) explored how the AR systems should unfold information in response to the user's attention. The user evaluation result indicates that when the user has to see a target and interact with the target in conversation, virtual user interface in the background and the real world in the foreground of the field of view is recommended. However, in educational visualization or when the main purpose of the interaction is to provide relevant information, not the conversation with the target, virtual user interface in the foreground, and real world in the background is recommended.

Using visualization technique of an eye trajectory.

Furthermore, they presented a Conversational User Interface which is useful when the user has to navigate an unfamiliar environment and the system should provide context-aware information about the user's surroundings such as relevant restaurants around lunchtime. The way of revealing information can be divided into three, whether no information is offered, it quickly reveals AR information at a glance, or the content is displayed always-on. Although it is desirable to present information by seamlessly supporting these various transitions depending on the context, the authors mention that having the content always-on could end up being distracting to the users. For this reason, in our project, We decided to use eye tracking techniques rather than hand gestures.

In the meantime, in our project multiple users' eye-gaze sharing is visualized by drawing the eye trajectory. In a study exploring visualization techniques when multiple users share their eye gaze, Rahman et al. (2020) used eye tracking for better insights into students' behavior patterns in VR environment displaying students' eye gaze in teacher's VR display. The researchers presented six gaze visualization techniques and conducted a user study to compare them. The eye gaze was presented either in a shape of six different shapes: a ring, a disc, an arrow, a trail based on the user's gaze points over the last three seconds, a trail with arrows in which the line segments with arrows represent gaze point movement, or lastly in a way of heatmaps in which the system changed the color of the object being gazed at. The result suggests that a short particle trail representing eye trajectory is most promising.

Context-aware Visualization of Gaze Cues

The efficiency of collaborative gaze interaction has been repetitively proposed, but few research studies have been investigated about their interactive feedback. Piumsomboon et al. (2017) proposed CoVAR as a system to visualize gaze cues for each user in a collaborative environment. Even though this is one of the first research studies proposing collaborative gaze interaction, this system does not offer feedback when occlusion of gaze frustum. Erickson et al. (2020) acknowledged the necessity of this feedback revealing that the difference in FOV of HMD can affect performance in collaborative gaze interaction. Another paper of Erickson et al. (2020) also proposed the significance of depth information for visualizing gaze cues. This paper is one of the first papers that tries to compensate for head-eye gaze conflict between users in collaborative virtual or augmented environments. They showed that visualizing with ray-casted points can improve localization for selecting targets. Research studies have started to provide context-awareness while visualizing gaze cues, but studies regarding the interactive feedback between the gaze cues are limited.

Method

Gaze Cue Interaction in Collocated Spaces

Being able to share each other's preferences without time consuming debates in the midst of a crowded street is one of the feature functions of our application. It will be plausible for a person to wear an HMD for more efficient communication in upcoming decades. Rather than a group of people sharing multiple types of information via multiple devices to arrive at a unified decision, it will be less cumbersome to glance at integrated information augmented to their real-time physical surroundings. Therefore, we developed a concept diagram describing a sequential operation of our application regarding the aforementioned user needs and scenario.

First of all, a HoloLens 2 device processes a real-time dataset of a user's gaze, head movement and direction. Moreover, since our application aims to aid projection of multiple targets selected by different users, it will more likely carry out the processing of multiple users at a scene. Imagine a couple of users cruising down a street to find a suitable restaurant to help themselves. There are signboards of various types of restaurants hanging on the exterior of building(s) which can serve as physical markers for HoloLens users. Users will take time to skim over the signboards in search of a satisfactory dining. The application intervenes once both users synchronously begin the searching activity within a walking distance of the restaurants available.

On the other hand, the application processes both users' physiological factors without displaying tracks of each user's gaze unless each user expresses his/her interest; this is owing to seamlessly sustaining glanceability of users wearing the HMDs while interacting with the physical environment. Variables of human factors being detected and measured to let the application capture his/her interest upon a specific signboard, are fixed duration of one's gaze and head direction. Once a user's fixed duration of gaze and head direction towards a spatial region of a signboard reaches or exceeds a given threshold for interest determination, a colored square appears multiplied to the signboard to indicate the user's interest. This indication mechanism is equally applicable to both users at the scene, thus simultaneously notifying both users.

In case both users get to have a common interest (e.g., gazing at the same signboard), each user's respectively colored shape will be stroked with a color extracted from another user's shape; the shape and stroke approach is expected to easily deliver an ownership to each user. The conceptualization of the shape and stroke projection was owing to setting each user as an active protagonist sustaining the validity of his/her own decision (the shape), and another person agreeing (the stroke) to the protagonist's choice (vice versa). The underlying assumption of this concept is that everyone's choice counts as a fixed point, not being hindered nor influenced by any types of external pressure. Therefore, this concept was deemed meaningful in that users get to silently understand each other's preference without a price (or possibly a probable cause) of sensitive arguments.

Figure 1 is attached as a concept diagram to describe the above sequence of UI usage. The diagram was illustrated using Adobe Photoshop and Illustrator (ver. CS 6). Overall, its composition and style aimed to achieve intuitiveness and aid the viewer's understanding of possible engagement through our proposed UI design. One noticeable variable that we have not accounted for is the luminosity of physical materials being projected on HoloLens UI. Imagine a weather condition not favorable to a HMD user searching for a physical marker of a place that serves a particular purpose. Our UI will need to seamlessly provide a continuously adapting first person view of the real world to a user. This type of functional concept would allow a user to choose between a thermal to various types of vision formats applicable to the UI, for compensating HSV (Hue Saturation Value) of lacking visual cues. This additional consideration of a concept will be discussed to check its suitability to our UI development.

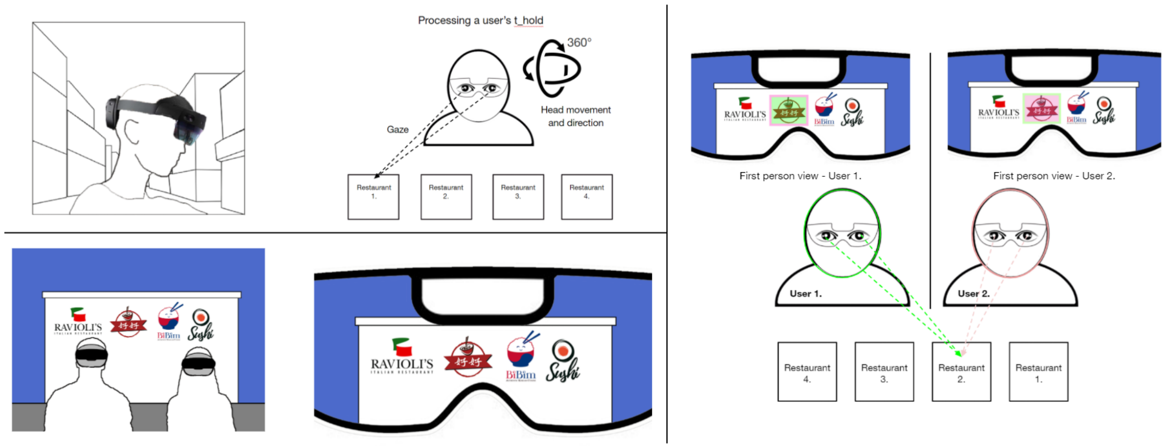

Fig 1. Concept diagram

Based on our concept diagram, we had to think about feasible methods to augment FPV (first-person-view) spatial cues for aiding target identification in shared gaze environments. According to Radu et al., there is a need for concerning features that can support collaborative needs in AR settings where collocated users wear headset devices. The paper regards a practical situation for users who are seemingly similar to our own. Our project's potential users would have to stay together within a close distance so that they may be able to identify each other's interest. Among the categories of collaborative needs drawn from Radu's paper, there are certain categories of needs possibly beneficial to our project: "collaborators need to be aware of others' attention and activities," collaborators need to coordinate attention," "collaborators need to coordinate instructions," and "collaborators need to share the same environment." The aforementioned needs are important due to dynamic, real-time, and seamless awareness cues that must be provided to our users.

For our current research concept, errors for eye-gaze input shared in collocated services may concern accuracy, precision, and latency as a whole. Furthermore, depth information of ROI would certainly have effects on visual target information, so concerning the relationship of depth and angular errors in gaze data, and the shared gaze situations is also another noteworthy approach for our project's scope. The fundamental question of why we have to augment spatial cues for our potential users can be partially answered by the paper by Lee et al. The primary function of spatial AR cues is to help users better understand where to look at while engaged in a particular collaborative situation, therefore, in accordance with argument from Lee et al., the spatial AR cues can be especially effective in guiding users' focus and attention in the real world space to ensure they become aware of where the users' interest (from our scenario) are referring to.

Apparatus

Microsoft Hololens devices offer "Research Mode" where developers can access every raw sensor data of their system. In Hololens 2, following streams are available to be logged; visible light environment tracking cameras, AHAT near-depth sensing camera for hand tracking (45 FPS), long-throw depth sensing camera for spatial mapping (1-5 FPS), two versions of IR-reflectivity stream, accelerometer, gyroscope, and magnetometer. Configuring the exact area of a target could be possible by obtaining logs from a visible light camera and a long-throw depth sensing camera.

Highlighting Targets in Collocated AR Environments

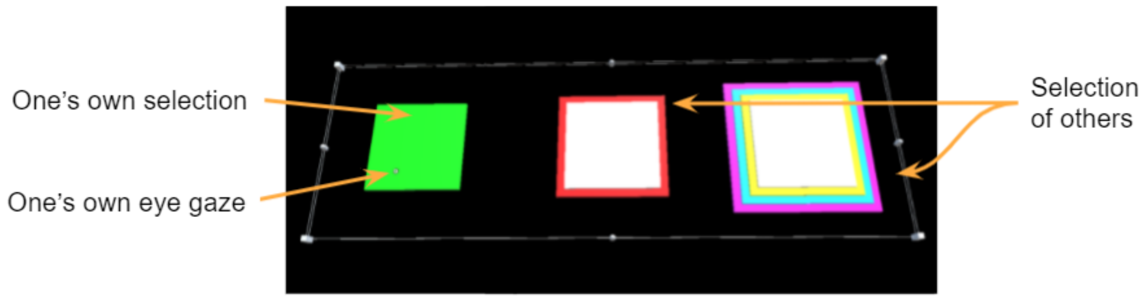

Fig 2. Screenshot of our implementation.

Figure 2 demonstrates the implemented results of our interaction design. First of all, while all the targets are plain white during idle mode, a certain target started with my eye gaze will demonstrate its shape entirely colorized. Moreover, the target would present highlighted strokes layered upon one another in case the target is simultaneously gazed by other people. This interaction schema would feasibly embrace a visual separation between selections of one's own and the others, not to mention selection tendencies visualized as histogram in order to aid context-awareness, and lastly possible real-time interaction in a collocated environment.

Logics for Synchronizing Target Renders

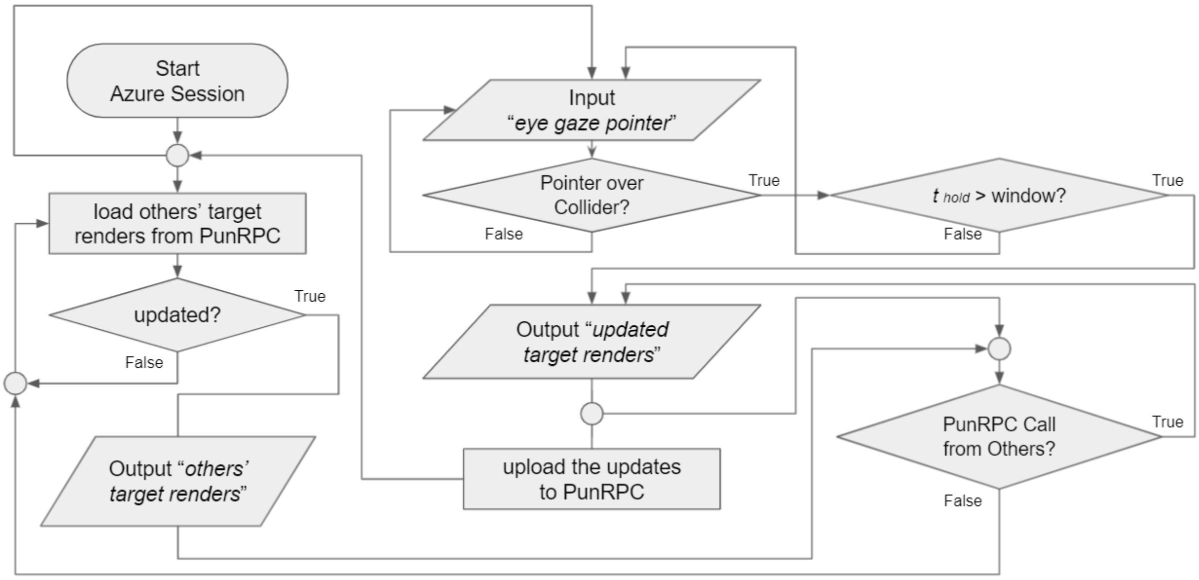

Fig 3. Flowchart for receiving user interaction.

If a user is looking at a target for a specific amount of time, the user's interest area is assumed with the user's head orientation and skeleton data and shared with nearby friends. If interest areas of nearby friends are visible, the user can interact with those and create a newly updated interest area which is also shared with nearby friends.

Figure 3 shows the flowchart of our system. We first briefly introduce the terminology used in the system. The user skeleton is the normal vector of the user's head and eyes in a physical (real) 3D environment. The constant "window" represents a temporal threshold of this system to detect a user's interest in a specific object. If the time of the user's head or eye's normal vector remains the same, hereafter denoted as t_hold, the system recognizes that the user is focused on the specific target, and sets a binding rectangle covering the target user's eye is heading to an Azure anchor.

We preferentially tuned highlights to physical targets based on their positions through Azure Spatial Anchor, then implemented synchronization of target selection among multiple users through Photon Pun RPC. Furthermore, the system regards each user's gaze information and target highlight condition as the user's input. Once a target condition is changed by gazes from more than a single user, a corresponding output call is deployed to synchronously update highlight conditions to all the users involved.

Interest areas are shared and managed as an Azure Spatial Anchor. As Azure is running, the interface collects the user's skeleton and surrounding environments along with other interest areas from the session. When the system detects the user looking at an object for a given time , the system marks the object as a new interest area and uploads it to the session so that near-by users can notice it. If the user's interest area interacts with another's area (e.g., overlapping or overriding), the system applies the corresponding changes to those areas and uploads them to the session.

Real-Time Object Coherence

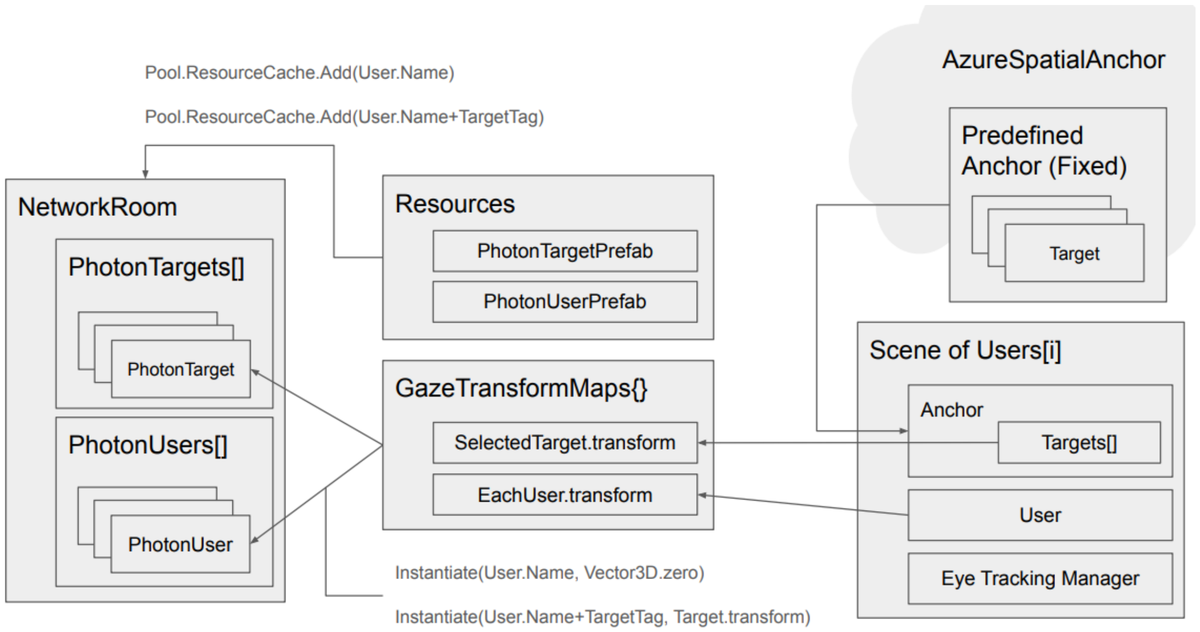

Fig 4. Scheme for implementing real-time object coherence for multiple users.

Azure spatial anchor enables a virtual object to be located at the exact physical space for multiple VR or AR users. However, the process of anchoring and releasing a given object is performed locally and requested to the cloud later, which limits objects from getting modified once they get anchored. This results in inherent limitations for real-time interaction. Therefore, we use Photon PUN RPC which were utilized by numerous VR applications for the real-time colocation of virtual objects between multiple users online. We first set a predefined anchor called a “board” at a fixed location in a physical and virtual world. Then we update displacements of every target (and ROIs) from the board via Photon PUN. This enables multiple objects to be colocated real-time within the minimized dependency of network bandwidth.

The main novelty in interaction this system can offer is locating the region of interest and segmenting the exact 3D object as an anchor. This is different from most AR systems presenting only the control of virtual objects using the human body. If the system can actively harvest the real-world data of users’ interest without the discomfort of information overflow, the possibility of its applicable domain will be unlimited.

Results

Validation

We performed several validation processes for the sake of external validity. First, we ran a simulation study to find the efficacy of finger pointing methodology in collaborative AR environments. Secondly, we validated our apparatus to measure spatial and temporal performance of our system. And finally, we performed user study to compare the performance of visual cue for different attention cueing methods.

Simulation Study

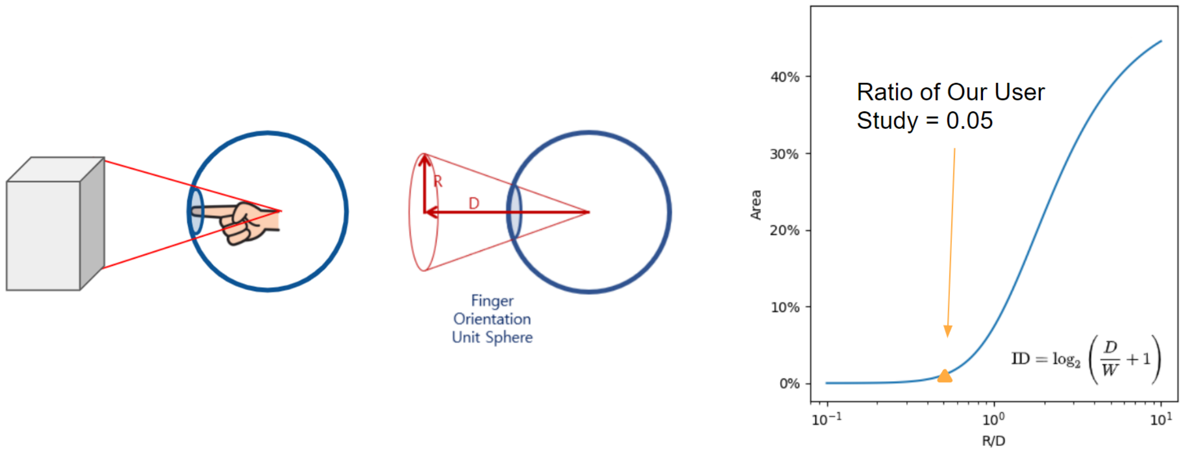

Fig 5. Finger pointing model (left) and the area ratio by pointing difficulty (right).

We first simulated the true positive ratio of finger pointing gestures under an assumption of finger orientation following uniform distribution. The goal of this simulation study is to find out how difficult it is to point at a specific target. We assumed an uniform distribution of finger orientation. Then we calculated a ratio of selection range over FOV (Field of View). We generated a wide range of radius over distance from 0.1 to 10, one of most common indexes representing difficulty in point target selection tasks. The results show that even in the easy circumstance with the ratio of 0.3, the true positive ratio drops to almost zero. With this study, we conclude that the finger pointing gesture is limited for cueing attention towards distant objects.

Spatial and Temporal Accuracy

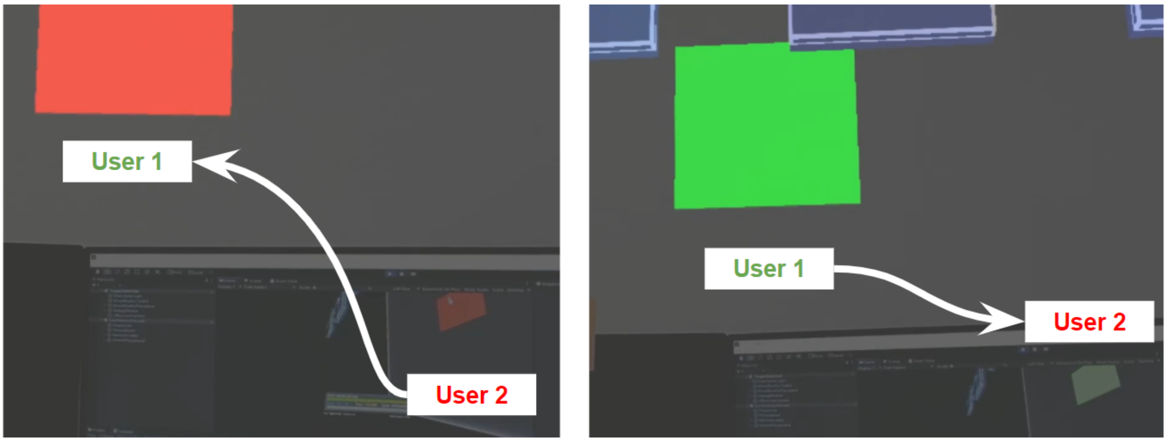

Fig 6. Measuring two different temporal latencies during PunRPC synchronization.

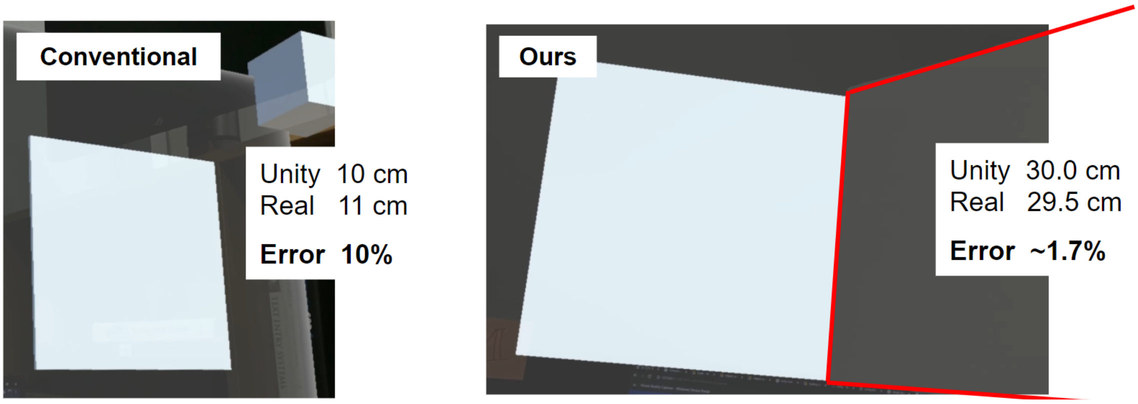

Fig 7. Spatial accuracy for target augmentation.

We measured spatial and temporal accuracy of our solution. We first measured temporal delay of target renderer synchronization implemented with Photon Pun RPC. Considering the rendering pipeline, we measured two directions of PunRPC callback transitions. Both directions of callback transitions resulted in about 0.3 seconds of delay which is fast enough for cueing attention. We also measured the scaling errors of MRTK. Even though unity uses meters as units of object scales, their raw rendering results in about 10% of scaling errors. Therefore, we compensated it by implementing custom scalers for our targets which reduced the errors into 1.7%.

User Study

Fig 8. Observer's view (left) and the demonstrator's view (right) of finger pointing.

Fig 9. Observer's view (left) and the demonstrator's view (right) of our system (ICUC).

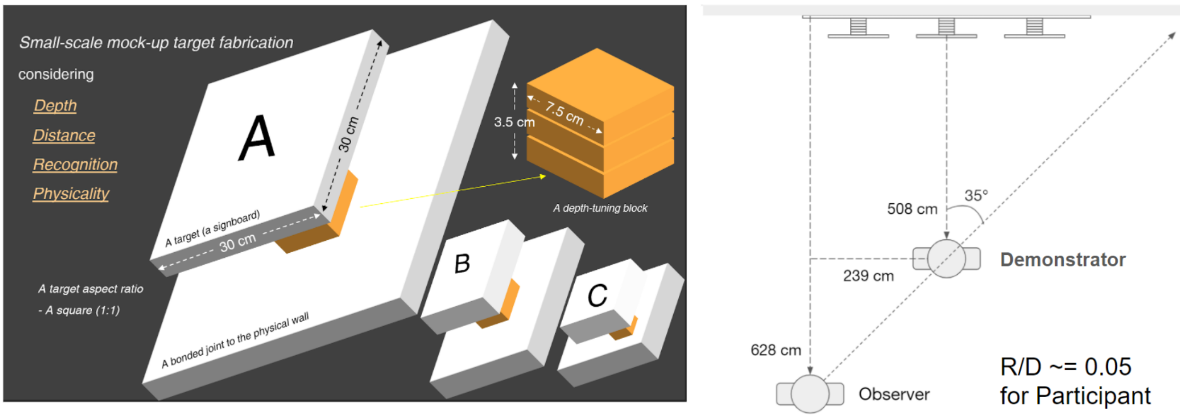

Fig 10. Profile of target (left) and experiment setups (right).

We performed user study where participants observe demonstrations of cue attention methods and try to find out what target the demonstrator is selecting. We adopted the experiment design of Akkil et al. (2016) and reduced the number of targets into three. We first recorded a video of a demonstrator selecting one of three distant targets utilizing one of two attention cueing methods (finger pointing and ICUC). The choices were made 12 times per video and its sequence was randomly generated. The entropy of choice sequence was controlled as the same for all videos. Then, we recruited 12 participants and showed two demonstration videos (one for the attention cueing methods) and asked them to write down what target the demonstrator is selecting. We collected numbers of correctly selected targets for each clip. And a two-way ANOVA was run on a sample of 12 participants to examine the effect of attention cueing methods and order of demonstration videos on the scores.

Considering the learning effect, we set the order of demonstration videos as between subject factors.

Figure 10 illustrates the experimental setups for targets and the demonstration video. Each demonstration clip is recorded by an observer wearing Hololens 2. The radius per distance ratio was set as 0.05. We set the angular difference of the demonstrator and observer as 35 degrees so that azimuth of fingers are not projected into zero and both the targets and fingers are captured in a scene.

We found significant improvement in scores (p=0.001) of participants observing ICUC users compared to those of finger pointing. No significant effects of video orders are found (p=0.341). Our user study shows that participants can understand attention cues more accurate in ICUC than finger pointing.

Discussion & Future Works

In this study, we proposed a new system that provides real-time interaction between visualized gaze cues for collaborative outdoor environments. Information about where the collaborative user is looking is augmented in a HoloLens 2 device based on each user's eye and head position. When the two interests overlap, a notification in the form of square stroke method is augmented to the Hololens 2. The result of the user study indicates that there was significant improvement in scores of participants observing ICUC users compared to those of finger pointing. No significant effects of video orders were found. Our user study shows that participants can understand attention cues more accurate in ICUC than finger pointing.

Our work proposed a new design space of interaction between visualized gaze cues in collocated AR spaces. Our system is heavily based on real-time temporally paired data of the human body and its surrounding environment. Therefore, our system can serve as a fundamental structure of context-aware graphical interface for collaborative and collocated tasks, aiding the limitation of conventional interactions for cueing attentions (e.g., finger pointing, verbal communication).

Due to the enhanced attention cueing performance from our solution, AR contents can be utilized even under highly-dense spatiotemporal domain. Our system can not only serve as a platform for crowdsourced interaction at scale but also a huge opportunity for companies to deploy their targeted ads. Furthermore, our system enables real 3D objects to be bookmarked and shared even with the convenience of instantaneous uploads shown in hugely successful microservices like Instagram stories and Tik-Tok.

However, our study has some limitations although we tried to conduct a reasonable study. First of all, our research did not fully satisfy external validity. In order to meet the internal validity, only three locations of the target were randomly set. Also, the location of participants who looked at these targets were limited to only one location unlike the real environment. In addition, it was conducted in a controlled environment within the lab. As a result, the result did satisfy the internal validity in that it had a controlled experimental environment, but it did not satisfy the external validity sufficiently. More precisely, it did not satisfy ecological validity.

Therefore, in order to properly represent the situation in reality, the target should be placed in more diverse locations, and the location of the person watching the target should be more diverse. Furthermore, to check whether our result can be practically utilized, it will be necessary to conduct the experiment in the field. Therefore, in the future work it would be desirable to see if it is still helpful for users with mobility in real-world environments to engage in gaze-based interactions among myriad targets and targets placed in various locations.

References

Akkil, D., & Isokoski, P. (2016, September). Accuracy of interpreting pointing gestures in egocentric view. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 262-273).

Erickson, A., Norouzi, N., Kim, K. et al. Sharing gaze rays for visual target identification tasks in collaborative augmented reality. J Multimodal User Interfaces 14, 353–371 (2020).

Erickson, A., N. Norouzi, K. Kim, J. J. LaViola, G. Bruder and G. F. Welch, "Effects of Depth Information on Visual Target Identification Task Performance in Shared Gaze Environments," in IEEE Transactions on Visualization and Computer Graphics, vol. 26, no. 5, pp. 1934-1944, May 2020.

Gupta, K., Lee, G. A., & Billinghurst, M. (2016). Do you see what I see? The effect of gaze tracking on task space remote collaboration. IEEE transactions on visualization and computer graphics, 22(11), 2413-2422.

Huidong Bai, Prasanth Sasikumar, Jing Yang, and Mark Billinghurst. 2020. A User Study on Mixed Reality Remote Collaboration with Eye Gaze and Hand Gesture Sharing. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–13.

Iulian Radu, Tugce Joy, Yiran Bowman, Ian Bott, and Bertrand Schneider. 2021. A Survey of Needs and Features for Augmented Reality Collaborations in Collocated Spaces. Proc. ACM Hum.-Comput. Interact. 5, CSCW1, Article 169 (April 2021), 21 pages.

Maglio, P. P., Matlock, T., Campbell, C. S., Zhai, S., & Smith, B. A. (2000, October). Gaze and speech in attentive user interfaces. In *International Conference on Multimodal Interfaces* (pp. 1-7). Springer, Berlin, Heidelberg.

Nancy Yao, Jeff Brewer, Sarah D'Angelo, Mike Horn, and Darren Gergle. 2018. Visualizing Gaze Information from Multiple Students to Support Remote Instruction. In Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18). Association for Computing Machinery, New York, NY, USA, Paper LBW051, 1–6.

Niklas Osmers and Michael Prilla. 2020. Getting out of Out of Sight: Evaluation of AR Mechanisms for Awareness and Orientation Support in Occluded Multi-Room Settings. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–11.

Pfeuffer, K., Abdrabou, Y., Esteves, A., Rivu, R., Abdelrahman, Y., Meitner, S., ... & Alt, F. (2021). ARtention: A design space for gaze-adaptive user interfaces in augmented reality. Computers & Graphics, 95, 1-12.

Piumsomboon, T., Lee, Y., Lee, G. A., Dey, A., & Billinghurst, M. (2017, June). Empathic mixed reality: Sharing what you feel and interacting with what you see. In 2017 International Symposium on Ubiquitous Virtual Reality (ISUVR) (pp. 38-41). IEEE.

Rahman, Y., Asish, S. M., Fisher, N. P., Bruce, E. C., Kulshreshth, A. K., & Borst, C. W. (2020, March). Exploring Eye Gaze Visualization Techniques for Identifying Distracted Students in Educational VR. In 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) (pp. 868-877). IEEE.

Reiner, A. J., Hollands, J. G., Jamieson, G. A., & Boustila, S. (2020). A mirror in the sky: assessment of an augmented reality method for depicting navigational information. *Ergonomics*, *63*(5), 548-562.

Thammathip Piumsomboon, Youngho Lee, Gun Lee, and Mark Billinghurst. 2017. CoVAR: a collaborative virtual and augmented reality system for remote collaboration. In SIGGRAPH Asia 2017 Emerging Technologies (SA '17). Association for Computing Machinery, New York, NY, USA, Article 3, 1–2.

Wickens, C. D., and J. S. McCarley. 2008. Applied Attention Theory. Boca Raton, FL: CRC press.