An Asynchronous Hybrid Cross Reality Collaborative System

Hyunwoo Cho*
University of South Australia
Bowen Yuan†
University of South Australia
Jonathon Derek Hart‡
University of South Australia
Eunhee Chang§
University of South Australia
Zhuang Chang¶
University of Auckland
Jiashuo Cao‖
University of Auckland
Gun A. Lee**
University of South Australia
Thammathip Piumsomboon††
University of Canterbury
Mark Billinghurst‡‡
University of South Australia

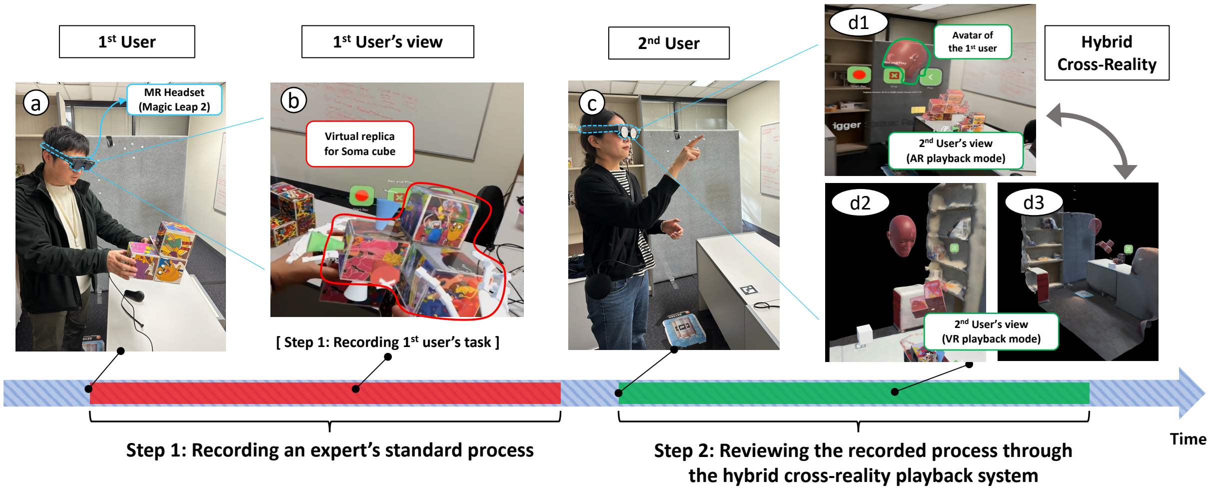

Figure 1: Using the Mixed Realit-based asynchronous hybrid cross reality collaborative system

ABSTRACT

This work presents a Mixed Reality (MR)-based asynchronous hybrid cross reality collaborative system which supports recording and playback of user actions in three-dimensional task space at different periods in time. Using this system, an expert user can record a task process such as virtual object placement or assembly, which can then be viewed by other users in either Augmented Reality (AR) or Virtual Reality (VR) views at later points in time to complete the task. In VR, the pre-scanned 3D workspace can be experienced to enhance the understanding of spatial information. Alternatively, AR can provide real-scale information to help the workers manipulate real world objects, and complete the task assignment. Users can also seamlessly move between AR and VR views as desired. In this way the system can contribute to improving task performance and co-presence during asynchronous collaboration.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Virtual reality; Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality

1 INTRODUCTION

In recent decades, Mixed Reality (MR) collaboration technologies have been applied across many domains, ranging from education and gaming to healthcare and defense [5]. Previous efforts have given rise to numerous advancements, including spatial computing and visual annotations, and the use of gaze and gestures for natural communication [24]. While most MR collaboration research has predominantly focused on synchronous collaboration, recent studies showed an increasing interest in asynchronous collaboration [9, 22]. Asynchronous collaboration entails multiple individuals working on a shared task at different points in time, so recording and playback functions are used to ensure all of the collaborators can communicate and collaborate to complete the task.

With this perspective, we introduce an MR-based system designed to record and playback activities in a temporal and spatial context for asynchronous collaboration. This system aims to simplify the process of recording three-dimensional (3D) task activities and create an immersive environment to review the recorded collaborative tasks and their associated surroundings. When reviewing the recorded activities, the user can freely select Virtual Reality (VR) or Augmented Reality (AR) viewing, or a hybrid mode to enhance task performance. The system not only shares 3D object information via a virtual replica but also visualizes the avatar of the collaborator, thereby enhancing the sense of co-presence.

The main contribution of this work is the introduction of a novel cross reality collaborative system that effectively enhances task efficiency and fosters a higher sense of presence during collaboration on physical tasks. Unlike previous cross reality research, this system supports asynchronous collaboration, which supports collaboration

---

[*] e-mail: hyunwoo.cho@mymail.unisa.edu.au
[†] e-mail: bowen.yuan@mymail.unisa.edu.au
[‡] e-mail: jonathon.hart@mymail.unisa.edu.au
[§] e-mail: eunhee.chang@unisa.edu.au
[¶] e-mail: zcha621@aucklanduni.ac.nz
[‖] e-mail: jcao403@aucklanduni.ac.nz
[**] e-mail: gun.lee@unisa.edu.au
[††] e-mail: tham.piumsomboon@canterbury.ac.nz
[‡‡] e-mail: mark.billinghurst@unisa.edu.au

at different points in time, over longer time scales, and between both AR and VR views.

2 RELATED WORK

2.1 Asynchronous collaboration in MR

Asynchronous collaboration refers to a cooperative scenario where participants interact at different times. Early computer-supported collaborative work (CSCW) literature established conceptual frameworks for time-distributed collaboration and its benefits such as work parallelism and reviewability. Olson and Olson [14] theorized the characteristics and advantages of asynchronous collaboration based on real-world examples of various collaborative environments in space and time.

As an early example, Bradner et al. [2] developed "BABBLE", a chat-based collaboration tool that graphically expressed user activity and supported both synchronous and asynchronous modes. They presented the idea of asynchronous collaboration where users from different time zones could collaborate efficiently using this text-based collaboration tool. Similarly, Tam and Greenberg [20] conducted a study on change awareness to identify the work history when a change occurs in multi-collaborative document creation or graphical workspace. In particular, they presented a work-sharing framework with intuitive visualization to facilitate tracking of various asynchronous collaboration events: such as who changed the artifact, what those changes involved, where changes occurred, when changes were made, how things changed, and why people made the changes.

Previous research on asynchronous collaboration in this CSCW framework has been actively applied in various fields such as video conferencing [21], data visualization [7], web search [13], and film review [15, 17]. There is a growing interest in employing adaptable cross reality that uses the most suitable environment in accordance with the specific requirements of a given collaborative situation. This approach includes the flexibility of freely switching between VR and AR modes based on collaboration needs.

2.2 Shared spatial information in collaboration

In MR collaboration, a 3D reconstruction process is often necessary for sharing a real-world task space with a remote collaborator and for interaction between virtual and real objects. In its broadest sense, 3D reconstruction technology can create a virtual representation of real-world scenes and is a core prerequisite for MR collaboration and telepresence applications. Creating a virtual 3D reconstruction of a physical space will allow the user in that location wearing an AR display to collaborate with the remote worker in the VR copy of that same location.

For example, Lindlbauer and Wilson [11] developed "Remixed Reality" which used multiple depth cameras to reconstruct the user's whole physical environment in real-time. It allows users to see the actual physical world but perform changes quickly in VR. They also demonstrated various types of spatiotemporal manipulation methods, such as erasing the appearance of objects in space, changing textures, and temporal changes such as pausing time or playback of recording.

Thoravi et al. [23] created "LOKI", an MR telepresence system for training on physical tasks. They configured the symmetric learning environment in the space of the learner and the tutor in a remote location, and the learning space was reconstructed in 3D using multiple RGB-D cameras. The reconstructed spatial data is shared with both learners and tutors, and a 3D widget called "Hologlyph" was used to support visual annotations, scale changes, and space rotations. In their study, various real-world tasks such as "teaching guitar", "sculpting", "Coaching Baseball", and "Workshop Learning through Peers" using spatial interfaces were implemented, and its effectiveness was confirmed through user studies.

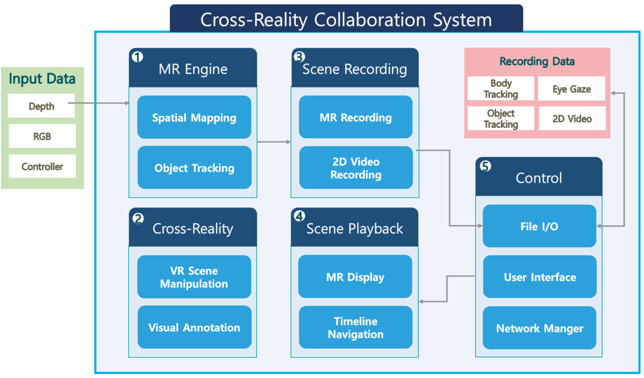

Figure 2: The framework of the Cross Reality Asynchronous Collaborative System, showing the different modules implemented

Finally, Wang et al. [26] conducted a study to enhance the user experience for AR-based interior design in a remote environment through a novel 3D reconstruction and authoring interface. They created 3D scene textures using projecting pixel and pyramid blending techniques to show more realistic reconstruction results. They created a 3D interface for navigating remote spaces and arranging objects by switching between third and first-person views. They conducted a user study where participants showed better understanding and navigation performance in the remote 3D scene, suggesting that the authoring interface could improve remote collaboration.

2.3 Cross Reality Collaboration

Cross reality collaboration is a collaboration between users using systems with multiple realities [10] located on the Reality-Virtuality Continuum (RVC) [12], such as between AR, VR, and MR. Depending on the collaborative tasks, different combinations of these cross reality interfaces could be used to enhance communication and understanding between collaborators. For example, Jing et al. [8] used AR to share the eye gaze cues of two local collaborators in real-world physical tasks so that both users could notice each other's attention during the collaboration. Similarly, Dey et al. [4] explored the effects of sharing the user's real-time heart rate feedback with the collaborators while performing collaborative tasks such as exploring a virtual environment, escaping from a virtual room, and arranging virtual furniture.

In addition to symmetric collaboration using AR-AR and VR-VR interfaces, asymmetric collaboration between AR and VR users has been widely researched. For example, Piumsomboon et al. [18] presented "CoVAR", a remote collaboration system combining AR and VR, where the AR user captured and shared the local task environment with a remote VR user. They also demonstrated scaling and binding the VR user's body to the AR view. Pazhayedath et al. [16] explored bi-directional pinpointing techniques in a cross reality collaboration where an external user in the real world collaborated with a VR user via either a PC or a tablet. Moreover, other researchers also used projectors to allow the real world user to collaborate with the VR user [6, 25].

Overall, current cross reality collaboration systems assigned the user at fixed position on the RVC [19]. However, less research has been done in cross reality collaboration using a transitional interface [1, 3] where the user could freely transition along the RVC. Our work presents an asynchronous cross reality collaboration system where the user can move along the RVC while undertaking a collaborative task.

3 SYSTEM DESIGN

We have developed a Cross Reality Asynchronous Collaborative System that allows people to collaborate using AR or VR views and move smoothly between them. The system uses a see-through

MR headset (Magic Leap2), in a room-scale indoor area measuring 10 m x 10 m. The system's operation encompassed two distinct phases: recording and playback. During the recording phase, the spatial task activities of a user, including tasks such as assembling, moving, and placing objects, are promptly captured and stored in a database. Subsequently, in the playback phase, a different user has the opportunity to observe the pre-recorded sequences depicting the previous user's task execution. These playback scenes are presented using virtual cues within the real 3D space, visible through the MR headset. Furthermore, to enrich the collaborative experience, users have an option to view the playback scene remotely in a VR environment, facilitated by a seamless conversion and supporting diverse visual annotations.

Figure 2 shows an outline of our cross reality asynchronous collaboration system, comprising five distinct modules: (1) MR Engine, (2) Scene Recording, (3) Playback, (4) Cross Reality, and (5) System Control. The MR Engine module is responsible for processing input data, conducting spatial mapping, and tracking objects to facilitate the visualization of MR scenes on the user's device. Subsequently, the Scene Recording module captures the user's task performance in two formats: 3D data and 2D video. The Scene Playback module retrieves the stored recording data and renders them in MR, while also supporting timeline navigation features such as adjusting playback speed and rewinding. The Cross Reality module is responsible for converting the collaborative environment into a VR space and supports VR-based visual annotations on recorded 3D scenes. Lastly, the System Control module handles functions related to input/output operations, encompassing tasks such as file storage and loading, network between users, as well as managing the user interface.

3.1 MR Engine

Our system is designed to gather diverse information concerning the user's activities and the task environment, with the purpose of facilitating immersive and efficient asynchronous collaboration. To achieve this, we used the Vuforia AR tracking library, which enabled the system to accurately track and record the movements of real-world objects. For realistic MR playback of the recorded data, we used pre-scanned 3D object models to visualize them as virtual replicas [24]. Moreover, the MR headset effectively tracked the 3D poses of the user's head and hands, allowing for the animation of an upper body virtual avatar to represent the user in the MR environment. We also integrated an eye tracking module to evaluate usability and record the visual information that held prominence for users during collaborative interactions.

3.2 Recording and Playback

Within the Scene Recording module, a total of four types of information was systematically recorded while the user engaged in task performance; (1) the trajectories of the user's avatar (comprising head and hand movements), (2) the trajectories of task objects being manipulated, (3) the gaze point trajectories, and (4) the first-person view video capturing the user's perspective. All recorded data was promptly stored in files in real-time during the user's task execution.

Subsequently, in the Playback module, users can re-experience the recorded data of the preceding user. This is accomplished by selecting the appropriate recording and activating the play button. As demonstrated in Fig. 1(d), the recorded scene is rendered with a translucent effect in the MR space, alongside the user's avatar and the manipulated object, enabling a comprehensive visualization. Through a dedicated timeline interface, users can effectively review the task performance in 3D, which also provides them with valuable functionalities such as rewinding, adjusting playback speed, and pausing the playback process.

In addition, both the recorded data and user interface (UI) event information are shared in real-time among two or more local users through the system's network module. As a result, multiple users within the task space can simultaneously view synchronized playback images.

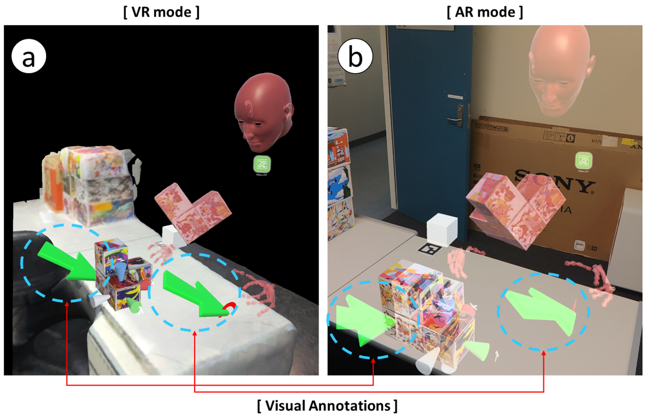

Figure 3: (a) Visual annotations (arrows) in the VR mode, (b) the same visual annotations shown when switching to the AR mode

3.3 Cross Reality Collaboration

The prototype system facilitates cross reality collaboration through two core functionalities: VR scene manipulation and VR-based visual annotation. During the playback phase, the user has the ability to switch from the AR mode to the VR mode, resulting in the natural disappearance of the camera input. When going into VR mode, the recorded scene aligns and is played back on a virtual model that accurately reconstructs the task space in three dimensions. This can be especially useful when the user needs to review the playback remotely from the task environment. Within this context, users are empowered to manipulate, scale, translate, and rotate the VR scene, thereby gaining the advantage of observing the scene from multiple perspectives. Moreover, users can use the VR-based Visual Annotation feature to add various visual annotations to the playback scene while operating within the VR mode. These recorded annotations can be saved along with the existing recorded scene, facilitating asynchronous communication with other users.

4 USE CASE SCENARIO

The prototype system facilitates two distinct types of asynchronous collaborative tasks: object placement and assembly. To test this, Soma cubes were employed as the real-world objects to be tracked, enabling users to engage in activities such as precisely positioning cubes within the designated task space or assembling them to form specific shapes.

The operational procedure of the demonstration unfolds as follows: During the recording phase, the user, wearing the MR headset (Fig. 1(a) and (b)), undertakes the task of assembling or placing Soma cubes while their activities are captured and recorded. In the subsequent playback phase, other users can review the recorded scene, presented within the MR space through the aid of a timeline interface, and can also opt to continue the task initiated by the first user (Fig. 1(c) and (d)). This review process provides access to comprehensive 3D information, including the user's trajectories and the movements of the Soma cubes, from multiple viewing angles.

This step also incorporates network synchronization of playback video for multiple users, facilitating concurrent participation in collaborative tasks by two or more users. By employing spatial mapping-based localization, the user devices share a common coordinate system, ensuring the alignment of the playback scene to the same position in the real-world space. Additionally, synchronization of events (e.g., play, rewind, and pause) for recorded scenes allows for real-time sharing of the same information among users. Moreover, users also have the freedom to playback the recordings in VR mode, enabling them to edit the playback scene through view manipulation and visual annotation within the VR environment, as depicted in Fig. 3 (a). This could be especially useful when the user wants to review the recordings remotely from the physical task space. Upon returning to the physical task space in AR mode, the edited video naturally aligns with the real-world space, as shown in Fig. 3 (b).

Consequently, users can collaboratively tackle intricate spatial tasks with ease and efficacy through asynchronous collaboration. Notably, owing to the file-based storage structure of the recording data, the review process can be simultaneously experienced by two or more users.

5 CONCLUSION

Growing interest has been shown in asynchronous collaboration using MR technology. Our prototype system employs time-space recording and playback in a hybrid cross reality interface and aims to provide improved task efficiency and a higher feeling of presence. Notably, this system allows the worker to freely transition between VR and AR views to enhance the collaboration quality. Various types of visual annotations such as text notes, avatars, and virtual replicas also help to support immersive experiences during asynchronous collaboration.

In future work, we intend to expand the range of tasks investigated and perform a user study with large groups of participants. We are particularly interested in real-world collaborative scenarios such as interior design and training body motion. We believe these steps will allow us to further advance the field of MR-based asynchronous collaboration, providing meaningful contributions to improve both the theoretical understanding and practical application of this technology.

REFERENCES

[1] M. Billinghurst, H. Kato, and I. Poupyrev. The magicbook: a transitional ar interface. *Computers & Graphics*, 25(5):745–753, 2001.

[2] E. Bradner, W. A. Kellogg, and T. Erickson. The adoption and use of 'babble': A field study of chat in the workplace. In *ECSCW'99*, pp. 139–158. Springer, 1999.

[3] K. Cho, H. Kim, and Y. Lee. Augmented reality coloring book with transitional user interface. *Indian Journal of Science and Technology*, 9(20):1–5, 2016.

[4] A. Dey, H. Chen, C. Zhuang, M. Billinghurst, and R. W. Lindeman. Effects of sharing real-time multi-sensory heart rate feedback in different immersive collaborative virtual environments. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 165–173. IEEE, 2018.

[5] B. Ens, J. Lanir, A. Tang, S. Bateman, G. Lee, T. Piumsomboon, and M. Billinghurst. Revisiting collaboration through mixed reality: The evolution of groupware. *International Journal of Human-Computer Studies*, 131:81–98, 2019.

[6] J. Gugenheimer, E. Stemasov, J. Frommel, and E. Rukzio. Sharevr: Enabling co-located experiences for virtual reality between hmd and non-hmd users. pp. 4021–4033, 05 2017. doi: 10.1145/3025453.3025683

[7] J. Heer, F. B. Viégas, and M. Wattenberg. Voyagers and voyeurs: Supporting asynchronous collaborative visualization. *Communications of the ACM*, 52(1):87–97, 2009.

[8] A. Jing, K. May, G. Lee, and M. Billinghurst. Eye see what you see: Exploring how bi-directional augmented reality gaze visualisation influences co-located symmetric collaboration. *Frontiers in Virtual Reality*, 2:697367, 2021.

[9] G. Lee, S. Ahn, W. Hoff, and M. Billinghurst. Ar tips: Augmented first-person view task instruction videos. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 34–36, 2019.

[10] H.-N. Liang, L. Yu, and F. Liarokapis. Workshop: Mixing realities: Cross-reality visualization, interaction, and collaboration. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 298–300. IEEE, 2023.

[11] D. Lindlbauer and A. D. Wilson. Remixed reality: Manipulating space and time in augmented reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2018.

[12] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE Trans. Information Systems*, vol. E77-D, no. 12:1321–1329, 12 1994.

[13] M. R. Morris and E. Horvitz. Searchtogether: an interface for collaborative web search. In *Proceedings of the 20th annual ACM symposium on User interface software and technology*, pp. 3–12, 2007.

[14] G. M. Olson and J. S. Olson. Distance matters. *Human–computer interaction*, 15(2-3):139–178, 2000.

[15] A. Pavel, D. B. Goldman, B. Hartmann, and M. Agrawala. Vidcrit: video-based asynchronous video review. In *Proceedings of the 29th annual symposium on user interface software and technology*, pp. 517–528, 2016.

[16] P. Pazhayedath, P. Belchior, R. Prates, F. Silveira, D. S. Lopes, R. Cools, A. Esteves, and A. L. Simeone. Exploring bi-directional pinpointing techniques for cross-reality collaboration. In *2021 IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW)*, pp. 264–270. IEEE, 2021.

[17] J. Phalip, E. A. Edmonds, and D. Jean. Supporting remote creative collaboration in film scoring. In *Proceedings of the seventh ACM conference on Creativity and cognition*, pp. 211–220, 2009.

[18] T. Piumsomboon, Y. Lee, G. Lee, and M. Billinghurst. Covar: a collaborative virtual and augmented reality system for remote collaboration. In *SIGGRAPH Asia 2017 Emerging Technologies*, pp. 1–2. 2017.

[19] J.-H. Schröder, D. Schacht, N. Peper, A. M. Hamurculu, and H.-C. Jetter. Collaborating across realities: Analytical lenses for understanding dyadic collaboration in transitional interfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–16, 2023.

[20] J. Tam and S. Greenberg. A framework for asynchronous change awareness in collaborative documents and workspaces. *International Journal of Human-Computer Studies*, 64(7):583–598, 2006.

[21] J. Tang, J. Marlow, A. Hoff, A. Roseway, K. Inkpen, C. Zhao, and X. Cao. Time travel proxy: using lightweight video recordings to create asynchronous, interactive meetings. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3111–3120, 2012.

[22] S. Thanyadit, P. Punpongsanon, T. Piumsomboon, and T.-C. Pong. Xr-live: Enhancing asynchronous shared-space demonstrations with spatial-temporal assistive toolsets for effective learning in immersive virtual laboratories. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1):136, 2022.

[23] B. Thoravi Kumaravel, F. Anderson, G. Fitzmaurice, B. Hartmann, and T. Grossman. Loki: Facilitating remote instruction of physical tasks using bi-directional mixed-reality telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pp. 161–174, 2019.

[24] H. Tian, G. A. Lee, H. Bai, and M. Billinghurst. Using virtual replicas to improve mixed reality remote collaboration. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2785–2795, 2023.

[25] Y. Wang, P. Wang, Z. Luo, and Y. Yan. A novel ar remote collaborative platform for sharing 2.5 d gestures and gaze. *The International Journal of Advanced Manufacturing Technology*, 119(9-10):6413–6421, 2022.

[26] Z. Wang, C. Nguyen, P. Asente, and J. Dorsey. Distanciar: Authoring site-specific augmented reality experiences for remote environments. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2021.