SpatialPrivacy: Spatial Sharing for Remote Collaboration in Mixed Reality

Kaiming Cheng
JPMorganChase, UW
kaiming.cheng@jpmorgan.com *

Mengyu Chen
JPMorganChase
mengyu.chen@jpmchase.com

Feiyu Lu
JPMorganChase
feiyu.lu@jpmchase.com

Youngwook Do
JPMorganChase
youngwook.do@jpmchase.com

Blair MacIntyre
JPMorganChase
blair.macintyre@jpmchase.com

ABSTRACT

Mixed Reality (MR) offers opportunities to greatly improve remote collaboration, given its ability to understand and reason about the spatial environment and to output multi-sensory feedback. Passthrough camera access, which is essential for MR applications to reason about the surrounding, is now available on major MR devices such as Microsoft's HoloLens 2 and Apple's Vision Pro. However, how to comfortably share our spatial environment to MR applications becomes an ever-more important question to address, given the wide variety of sensitive visual information it contains. In this paper, we present SpatialPrivacy, a framework that offers spatial control through indoor localization techniques and access control policies configured over digital replicas of the real-world space.

1 INTRODUCTION

As Mixed Reality (MR) technologies continue to advance, their ability to sense, understand, and reconstruct our physical environments has seen remarkable improvements. This increased capability not only enables smooth transitions between the physical and the digital but also opens new possibilities for sharing digitally rendered spaces with others to let remote users "see" what another person is seeing in real-time. Building on these advancements, researchers have, for the past few decades, explored and developed various MR systems and interaction mechanisms that facilitate remote collaboration in various application domains, including health, industry, sports, design, and training [3, 4, 10]. Compared to collaboration through 2D-based video teleconferencing systems, remote collaborations in MR can enhance the sense of co-presence, engagement, and work effectiveness [11, 13]. These MR remote collaboration experiences are possible as MR devices continuously record and sense the user's environment by using sensors (e.g., cameras and depth sensors).

However, this sensing capability that enables the remote collaboration introduces privacy concerns. Particularly, the sensors embedded in MR devices could be used to unwittingly record and infer the potential sensitive data contains in a user's physical surroundings [7]. This concern is not theoretical. Existing industry MR platforms have already recognized the potential concern associated with the outward-facing passthrough cameras and have implemented rigorous privacy controls. For example, passthrough-camera data access is restricted across all Meta Quest products and on visionOS 1.0 for Apple's Vision Pro. This approach, however, limited the potential utility, pushing it to the extreme of the privacy-utility trade-off curve [12]. Recently, visionOS 2.0 has opened the access to passthrough camera only under the enterprise API [1] to enhance the MR experience for enterprise applications. Although these APIs are designed specifically for workplace environments, deploying MR spatial sharing in settings that contain sensitive information still presents privacy challenges. Institutions like hospitals and banks, which must adhere to strict regulatory compliance due to the sensitive information they handle and store, may face heightened risks.

In this statement, we propose SpatialPrivacy, a framework that utilized indoor localization to configure access policy based on the spatial entity's relation to the real space. By regulating the spatial content shared in MR, SpatialPrivacy provides strong privacy guarantees while preserving the necessary utility to help remote collaboration in MR reach its full potential.

2 GOALS AND THREAT MODEL

We start by detailing our threat model. Then, we describe the design goals SpatialPrivacy aim to achieve.

Threat Model We consider the adversary to be either an untrusted application or an untrusted user. We assume attacker-controlled applications have full network access and can share sensitive content to external entities, and the untrusted user can access sensitive areas in physical workspace environments. We classify potential assets in the spatial content to four classes as follows.

- Room space: This includes the semantic information about the room and its indoor location, which could reveal the function of a space and activity happened within.
- Object space: This includes sensitive objects within the space that may contain or display confidential information, such as documents, screens, or proprietary equipment.
- User space: This involves the user's biometric information gathered from the MR environment, including but not limited to eye-tracking, hand-tracking, and input selection.
- Bystander space: This involves the presence and activities of individuals not engaging in the MR environment.

Design Goals To protect these assets when sharing the 3D environment, we made four key design goals as follows.

- Non-image-based solutions: To address privacy concerns, our framework avoids the use of cameras. This approach helps in adhering to compliance rules, as it prevents the processing of sensitive visual data.
- Object customization: We support fine-grained configuration, allowing users to specify which objects in the environment are sensitive and should be excluded from spatial sharing.
- Policy configuration: Our framework includes a policy configuration feature that clearly maps out restricted areas where MR technology should not record or process data. This ensures that sensitive zones within a 3D environment are explicitly protected from any form of data sharing.

*Kaiming is a Ph.D. candidate at the Paul G. Allen School of Computer Science & Engineering, University of Washington and completed this work on an internship at JPMorganChase.

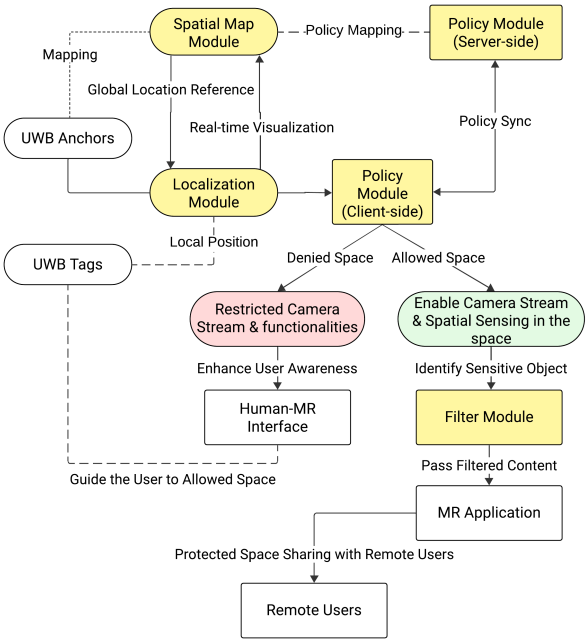

Figure 1: Diagram of the proposed solution. The yellow box represents the four main modules: localization, spatial map, policy, and filter. The green and red boxes represent allowed action and denied action.

- Dynamic tracking: Our framework should handle objects that can move from one location to another, such as users, bystanders, and certain sensitive objects like mobile whiteboards.

3 System Design

We now present the design of SpatialPrivacy, as shown in Figure 1. We envision this framework as a middleware that provides necessary privacy functionalities for MR remote collaboration applications.

3.1 Localization Module

For this project, we use Ultra-wideband (UWB) techniques to perform indoor localization, as prior work has demonstrated that UWB based solution can achieve a few-cm accurate localization [2]. Generally, we can categorize UWB devices into anchors and tags. A UWB anchor continuously transmits signals that are detected by UWB tags, which can be used to determine the relative distance. A UWB tag sends the received data to the server through local network, which will then compute the relative distance between the tag and the anchor and localize them in the global space. We assume that the MR headset will have UWB capabilities and thus will serve as a tag. We can also attach tags to moving objects to support dynamic tracking given its real-time localization accuracy. This setup allows us to determine (1) whether the headset is within a pre-defined area, (2) whether the direction in which the headset is looking overlaps with our pre-defined areas, and (3) the location of the dynamically tracked objects and the associated policy. We assume that the UWB sensors installed in the building are approved by the administrators of the system and are placed statically. We also assume that sensors communicate with the system over a secure channel that provides confidentiality and integrity of sensor data.

3.2 Spatial Map Module

We define physical spaces such as rooms and buildings as *spatial entities*. We assume that these spatial entities can be partitioned into non-overlapping places, termed *spatial units*. Users can define aggregate spatial units to represent one or multiple spatial entities.

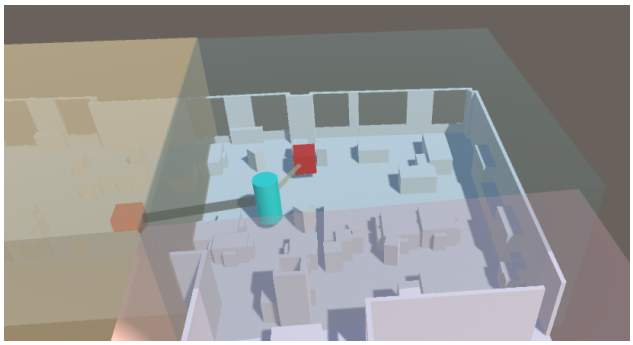

Figure 2: A demo for ultra-wideband (UWB) localization in the global space of a virtual digital twin. The red cube objects represent static anchors that are placed in the physical space, and the blue cylinder object represents the location of a user. Each spatial entity of the digital twin maps to a distinct access control policy defined by admin.

The system generates a 3D spatial map, a digital replica of the real-world environment. Depending on the desired fidelity, the map may be represented as an abstract layout or as a detailed 3D model, with personal information removed to ensure privacy. The localization module utilizes this spatial map to determine the locations of entities within this space and to integrate these findings with the physical world. Figure 2 presents a demonstration of the localization module operating over a spatial map.

3.3 Policy Module

Based on the spatial map module, the administrators can then designate specific policies as allowing or denying certain action over the spatial entities from the server side. The server uses a centralized control engine that manages policy specifications for each spatial entity. This engine can be programmed in a fine-grained manner to support various access control strategies, such as role-based access control (RBAC) [9], spatial hierarchical policy structure, and temporal access control. The control engine can also resolve potential conflicted policy based on the hierarchy. We plan to leverage public key infrastructure inspired by world-driven access control [8] for policy signing and authentication. The policy will then be conveyed to the client side. Depending on the engine's output, this client side policy module can either restrict certain MR capabilities or guide the user to a pre-defined allowed space.

3.4 Filter Module

After identifying and localizing the target objects to blur, we will utilize state-of-the-art 3D object detection algorithms on point cloud data to estimate their 3D poses. Subsequently, we will generate inpainting masks derived from the segmentation bitmaps [5]. During spatial sharing sessions, we plan to employ learning-based video inpainting methods to construct frame-by-frame inpainting visual results for computation efficiency [6]. Balancing the inference latency with the quality of the inpainting result is crucial, as it significantly impacts the overall immersive effect. We plan to conduct extensive experiments to understand the trade-offs between different inpainting methods and to evaluate their perceptual impact on end users.

4 Conclusion

Remote collaboration in MR holds much promise, raising critical privacy risks if deployed inappropriately. These risks need to be addressed while the MR ecosystems and new capabilities are actively being developed. To this end, we build a framework that addresses privacy vulnerabilities introduced by passthrough cameras while preserving the necessary utility. By protecting sensitive content during remote collaboration, we are taking steps towards securing fully-fledged MR remote collaborations in the future.

REFERENCES

[1] Building spatial experiences for business apps with enterprise apis. https://developer.apple.com/documentation/visionOS/building-spatial-experiences-for-business-apps-with-enterprise-apis, 2024.

[2] A. Arun, S. Saruwatari, S. Shah, and D. Bharadia. Xrloc: Accurate uwb localization to realize xr deployments. In *Proceedings of the 21st ACM Conference on Embedded Networked Sensor Systems*, pp. 459–473, 2023.

[3] M. Billinghurst and H. Kato. Collaborative mixed reality. In *Proceedings of the First International Symposium on Mixed Reality*, 1999.

[4] C. G. Fidalgo, Y. Yan, H. Cho, M. Sousa, D. Lindlbauer, and J. Jorge. A survey on remote assistance and training in mixed reality environments. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2291–2303, 2023.

[5] M. Kari, T. Grosse-Puppendahl, L. F. Coelho, A. R. Fender, D. Bethge, R. Schütte, and C. Holz. Transformr: Pose-aware object substitution for composing alternate mixed realities. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 69–79. IEEE, 2021.

[6] M. Liao, F. Lu, D. Zhou, S. Zhang, W. Li, and R. Yang. Dvi: Depth guided video inpainting for autonomous driving. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16*, pp. 1–17. Springer, 2020.

[7] R. McAmis and T. Kohno. The writing on the wall and 3d digital twins: personal information in (not so) private real estate. In *32nd USENIX Security Symposium (USENIX Security 23)*, pp. 2169–2186, 2023.

[8] F. Roesner, D. Molnar, A. Moshchuk, T. Kohno, and H. J. Wang. World-driven access control for continuous sensing. In *ACM Conference on Computer and Communications Security*, 2014.

[9] R. S. Sandhu. Role-based access control. In *Advances in computers*, vol. 46, pp. 237–286. Elsevier, 1998.

[10] A. Schäfer, G. Reis, and D. Stricker. A survey on synchronous augmented, virtual, andmixed reality remote collaboration systems. *ACM Computing Surveys*, 55(6):1–27, 2022.

[11] M. Sereno, X. Wang, L. Besançon, M. J. Mcguffin, and T. Isenberg. Collaborative work in augmented reality: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 28(6):2530–2549, 2020.

[12] D. Smullen, Y. Feng, S. A. Zhang, and N. Sadeh. The best of both worlds: Mitigating trade-offs between accuracy and user burden in capturing mobile app privacy preferences. *Proceedings on Privacy Enhancing Technologies*, 2020.

[13] R. Wolff, D. J. Roberts, A. Steed, and O. Otto. A review of telecollaboration technologies with respect to closely coupled collaboration. *International Journal of Computer Applications in Technology*, 29(1):11–26, 2007.

DISCLAIMER