The Impact of Occlusion and Collision on Real–Virtual Interaction in Mixed Reality 360° Images

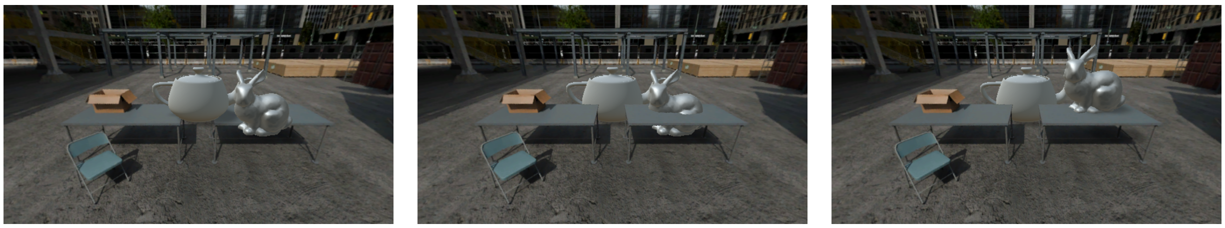
Andrew Chalmers*
Victoria University of Wellington
The University of Melbourne

SeungWon Jeong†
The University of Melbourne

Xuteng Yuan‡
Victoria University of Wellington

Fanglue Zhang§
Victoria University of Wellington

Taehyun Rhee¶
The University of Melbourne



Baseline (MR360) with occlusion with occlusion and collision

Figure 1: Example mixed reality scene in which occlusion and collision effects are added to MR360 [20], enabling interaction with virtual objects (teapot, bunny) within real-world simulated 360° images. The teapot and bunny are occluded behind a table (centre and right), and the bunny is placed on top of the table through collision handling (right).

Abstract

360° images, when paired with virtual reality (VR) headsets, offer an immersive viewing experience of real-world environments. The integration of mixed reality (MR) within 360° videos further enhances users' telepresence by blending virtual assets into the real environment. This paper investigates how high-fidelity visual appearance and physical interaction between virtual and real objects affect user perception in MR 360° videos. We focus on two key features, occlusion and collision, that enable high fidelity interaction with virtual objects within a complex mixed reality 360° scene. To study their impact on visual realism and users' sense of presence, we conduct a user study using simulated 360° videos with pixel-aligned depth data.

Index Terms: mixed reality, 360° image, occlusion, collision

1 Introduction

A 360° camera captures a wide surrounding area, producing high-resolution panoramic imagery. When viewed through Virtual Reality (VR) head-mounted displays, these 360° images offer an immersive experience with high-fidelity telepresence within the captured real environment.

Despite significant advancements in 360° video and interactive MR, previous research has focused primarily on overlaying virtual information onto 2D panoramic videos or empty background regions without considering physical interaction with real-world content [20]. A key limitation is the lack of 3D depth information from real background essential for accurately handling occlusion and collision between virtual objects and the real environment. Utilizing depth or 3D point cloud data to form RGB-D 360° video enables additional MR effects, such as accurate occlusion and collision between virtual and real objects. While capturing and modeling 3D data with 360° video remains an active and challenging area of research in computer graphics and vision, its impact on mixed reality interaction in 360° images has not been thoroughly investigated.

We investigate how occlusion and physics/collision simulation contribute to enhancing virtual object integration (VOI) into the 360° video, presence, and realism. Through user studies, we assess the impact of users' perception of telepresence and visual quality during interactive tasks involving both real and virtual objects. We utilize simulated 360° video with pixel-aligned depth data to begin the exploration of how occlusion and collision between virtual and real objects affect user perception and presence in MR 360° videos.

Our findings suggest that improvements in both visual appearance and physical interaction, especially when both features are combined together, significantly enhance telepresence and perception of visual quality.

2 Related Work

Augmented telepresence aims to enhance remote communication by providing immersive experiences that allow users to feel physically present in a different location [6]. Early works such as video conferencing platforms laid the foundation for modern telepresence [9, 1]. Recent work [14, 21, 28] focusing on immersive telepresence framework allows users to interact within shared virtual spaces as if they were physically present, or to experience the feeling of walking in the environment with an avatar [3]. Some of these previous methods are based on MR360 [20], which developed a method for high visual quality integrating 3D virtual objects into panoramic videos. However, this is only focused on lighting and shadows, with no attention to how the integration of depth information affects the user.

Depth perception is a critical factor in enhancing the quality of user experience and interaction in mixed reality (MR) systems. Teng et al. [24] explored how conflicting sensory cues, such as discrepancies between visual and motion signals, influence perceived object depth in virtual environments. Similarly, Westermeier et al. [25, 26] examined the effects of sensory and cognitive incongru-

*e-mail: andrew.chalmers@vuw.ac.nz
†e-mail: seungwon@student.unimelb.edu.au
‡e-mail: yuanxute@myvuw.ac.nz
§e-mail: fanglue.zhang@vuw.ac.nz
¶e-mail: taehyun.rhee@unimelb.edu.au (corresponding author)

encies on plausibility, spatial presence, and overall user immersion, highlighting significant differences in perception during a simulated maintenance task.

Some immersive systems leverage depth cues to create occlusion effects, thereby enhancing realism. For instance, Gao et al. [7] investigated depth-based perception conflicts through error-related potentials, focusing on the perceptual challenges posed by depth inconsistencies rather than their effects on user performance. Maslych et al. [13] introduced an interactive disocclusion mini-map for VR environments, which aids users in acquiring occluded objects by offering a visual tool for disocclusion, leading to improved task efficiency in complex scenarios. Liao et al. [10] further compared AR interfaces across four tasks involving disocclusion, such as counting dynamic spheres and identifying occluded objects. While these studies provide valuable insights into occlusion effects within VR and AR, their findings cannot be directly applied to 360° content-based MR systems, where interaction dynamics differ significantly, with prior work only considering the effects of resolution in static panoramic environments [15].

3 System Development: Occlusion and Physics Interaction in Mixed Reality 360°

Motivated by challenges in AR/MR [22], we focus on developing two key features, occlusion [12, 15] and physics interaction [18], in our study. Our system builds on MR360 [20] for coherent lighting and shadowing effects, while incorporating 3D depth to enable realistic integration of virtual objects into the 360° MR scene.

Incorporating depth or 3D point cloud data into 360° video allows virtual assets to be positioned accurately relative to real-world objects, whether in front of, behind, or partially hidden [8, 15]. Accurate occlusion ensures that virtual objects are visually blocked by real objects, whether in front of, behind, or partially hidden [8, 15], enhancing depth perception and realism. Achieving occlusion requires acquiring 3D information of the real environment.

Achieving this requires reliable 3D information of the real environment, obtained via manual editing, depth estimation, or sensors [5] and 360° LiDAR [19, 17, 4]. Each method has trade-offs in acquisition time, accuracy, and synchronization. The resulting depth map or point cloud supports occlusion-aware compositing, enabling virtual objects to blend seamlessly with real elements.

However, without collision handling, virtual objects appear as "ghosts," passing through real-world elements with no physical response. To address this limitation, we integrate real-time collision detection and physics simulation between virtual and real objects to eliminate the "ghost" effect. These interactions are enabled by generating collider meshes from the 3D representation of the real environment. For greater stability, we applied plane detection algorithms to fit collider planes to detected surfaces [16, 27, 2, 11]

More advanced scenarios involve object detection and segmentation to recognize specific objects, allowing for tailored colliders or predefined interactions. For example, recognizing human hands enables users to pick up and manipulate virtual objects directly [4], although this is left for future work.

4 User Study

4.1 Study Design and Conditions

We examine how integrating virtual content into 360° videos affects perceived presence and visual quality, hypothesizing that combining occlusion and physics-based interaction with the real environment enhances both.

We designed a within-subjects study, considering the following independent variables: *occlusion* (off/on) and *collision* (off/on). Our design treats collision as a higher-complexity condition that builds upon occlusion rather than operating independently of it. The final set of conditions are C1 (Figure 1-left): lighting on, occlusion off, collision off, C2 (Figure 1-middle): lighting on, occlusion on, collision off, and C3 (Figure 1-right): lighting on, occlusion on, collision on. C1 acts as our baseline, using MR360 [20] to illuminate the virtual objects to match the background video.

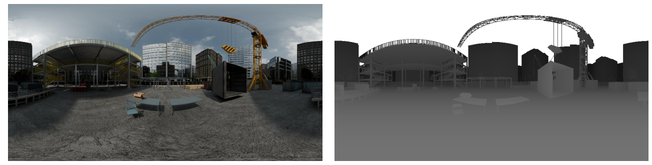


Figure 2: An image frame from our simulated 360° video used in the user study (left: RGB, right: depth).

4.2 Task Design

Across all conditions C1, C2, and C3, participants engaged in a task-oriented game designed to encourage 360° exploration and interaction within the environment. Virtual objects were placed around the environment for the participants to find and move to another location using a VR controller which has a virtual laser pointer to point and click to pick up an object. They could physically move their controller to move the virtual object, and press the down/up buttons to move the object closer/further away. The participant could drop the object by releasing the same button they used to pick up the object. The target location to place the virtual object was visualized using a podium with a virtual icon on top, indicating the target object that participants needed to find and move to the podium. Once the object was placed into a target location, a new target object appeared, and the process repeated. This sequence continued until participants completed four target objects per condition. See Figure 3 for an example of a participant in the study.

4.3 Stimuli

We rendered high quality 360° images to generate corresponding high-fidelity depth maps. These simulated environments act as realistic proxies for real-world 360° footage while providing a controlled setting for our experiments. We made three 360° videos (4k resolution, 25 seconds, continuous loop playback) of a construction site environment, featuring buildings, vehicles, boxes, and other items typically found on a construction site. We conducted the study using a Meta Quest 2 HMD and VR controller. There was no virtual representation of the participant beyond the controller. The system was implemented in Unity (version 2021.3.1f1).

4.4 Measures

We evaluated three key measures of the participant experience across the three conditions: 1) *Presence*, 2) *Integration of Virtual Objects*, and 3) *Task Completion Time*.

Presence was measured using the iGroup Presence Questionnaire (IPQ) [23]. The IPQ uses the term 'virtual environment', but since we are using data that represents real-world footage, we used the term '360 environment' in its place. To evaluate the *Virtual Object Integration (VOI)*, we included an additional question asking how well the virtual objects matched and blended with the background environment: "I felt that the virtual objects were situated in the same environment as the 360 environment.". This measure aimed to gauge the perceived integration of virtual objects within the simulated environment. These questions are answered on a 7-point Likert scale (see Table 1 for the questions).

4.5 Procedure and Participants

Participants first read an information sheet, signed an approved consent form, and completed a demographic questionnaire that collected information on their age group, ethnicity, gender, video game

Table 1: Questions used in 7-point Likert scale questionnaire.

Virtual Object Integration		
VOI	I felt that the virtual objects were situated in the same environment as the 360 environment.	1 = Fully disagree, 7 = Fully agree
Presence		
PRES	In the 360 environment, I had a sense of ”being there”	1 = Not at all, 7 = Very much
Realism		
REAL 1	How much did your experience in the 360 environment seem consistent with your real world experience?	1 = Not consistent, 7 = Very consistent
REAL 2	How real did the 360 environment seem to you?	1 = Completely real, 7 = Not real at all
Involvement		
INV 1	I was not aware of my real environment	1 = Extremely aware, 7 = Not aware at all
INV 2	I was completely captivated by the 360 environment	1 = Fully disagree, 7 = Fully agree
Spatial Presence		
SP 1	Somehow I felt that the 360 environment surrounded me	1 = Fully disagree, 7 = Fully agree
SP 2	I felt present in the 360 environment	1 = Fully disagree, 7 = Fully agree

experience, and whether they had normal or corrected-to-normal vision. After completing each condition, participants removed their HMD and filled in a questionnaire before putting the HMD back on to proceed to the next condition. A total of 16 participants were recruited for this study, with ages ranging from 23 to 36 (M = 26.81, SD = 2.93), with 5 identified as female, and 11 identified as male. Participants were recruited from a university campus, all of whom reported having normal or corrected-to-normal vision.

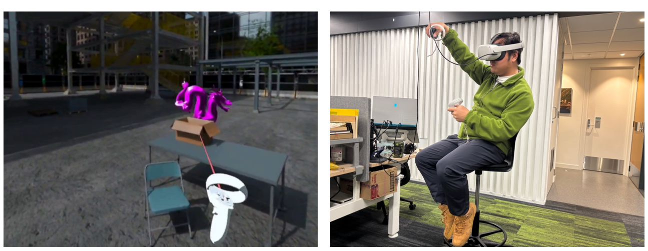


Figure 3: A participant in the user study wearing a Meta Quest 2 HMD using a VR controller.

5 Results

We analyzed participant responses for each of our main measures (IPQ and VOI). The data did not follow a normal distribution, so we used an Aligned Rank Transformation and performed a One-Way Repeated Measures ANOVA. The residuals between pairs followed a normal distribution, so we used a paired t-test. The result for VOI is shown in Figure 4-left, while the IPQ questions for presence, realism, spatial presence, and involvement are shown in Figures 4-right, 5, 6, and 7 respectively. The results are shown in Table 2.

A repeated-measures ANOVA revealed a significant difference in VOI (Figure 4-left), where results show an increasing trend across C1, C2, and C3. No statistical significance was found for the

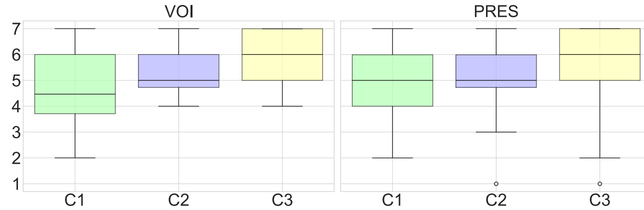


Figure 4: Results for (left) virtual object integration (VOI), and (right) presence (PRES) from IPQ questionnaire.

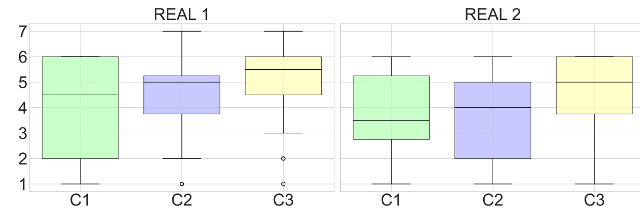


Figure 5: Results for realism (REAL) from IPQ questionnaire.

remaining measures. Presence (Figure 4-right) and realism (Figure 5) showed an increasing trend across C1, C2, and C3. Spatial presence (Figure 6) showed a weaker trend, with all conditions scoring quite high. Involvement with respect to awareness of the real environment (Figure 7-left) showed an improvement (lower is better) across C1, C2, and C3, while involvement with respect to captivation of the 360 environment (Figure 7-right) all scored high. Pairwise comparisons (paired t-tests, Bonferroni-corrected) between C1 and C3 revealed a significant difference for VOI ($p = 0.0143$).

Table 2: Mean values (higher is better) and p-values (*mean value in bold indicates best result between conditions, and p-values in bold indicate statistical significance*).

Question	C1	C2	C3	p-value
VOI	4.56	5.00	**5.81**	**0.0410**
PRES	4.81	4.81	**5.44**	0.4605
REAL 1	4.00	4.31	**4.88**	0.3948
REAL 2	3.69	3.75	**4.38**	0.5008
SP 1	5.19	**5.56**	**5.56**	0.6594
SP 2	4.94	5.06	**5.19**	0.9109
INV 1 (↓)	4.56	4.25	**3.62**	0.2557
INV 2	5.00	5.00	**5.31**	0.7982

6 Findings

The condition that incorporated both occlusion and collision detection (C3) achieved the highest ratings on VOI, presence, and realism metrics. A statistically significant difference was found when comparing both occlusion and collision together (C3) against lighting only (C1). While significance was not found in other comparisons, presence followed an upward trend. The results suggests that occlusion provides a small improvement, and introducing collision has a much more noticeable improvement.

7 Conclusion

This study investigated how occlusion and physical interaction influence user perception in mixed reality 360° videos. Our findings indicate that the combination of occlusion and collision yielded the highest ratings for virtual object integration (VOI), presence, and realism, while occlusion without collision showed a slight improvement. Our results highlight the pressing need for advancements in

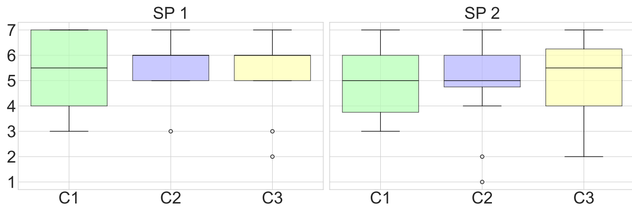


Figure 6: Results for spatial presence (SP) from IPQ questionnaire.

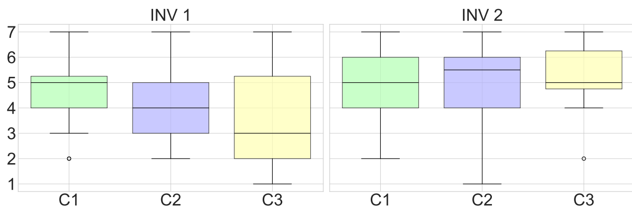


Figure 7: Results for spatial involvement (INV) from IPQ questionnaire. Lower is better for INV 1.

multi-modal sensing, providing a strong motivation for continued development of high-fidelity capture techniques that can support the seamless blending of real and virtual elements.

Acknowledgments

This paper was supported by the Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency grant funded by the Ministry of Culture, Sports and Tourism, South Korea (Project Number: RS-2024-00399136) and the internal research grant (Project Number: 400876) from Victoria University of Wellington.

References

[1] S. Beck, A. Kunert, A. Kulik, and B. Froehlich. Immersive group-to-group telepresence. *IEEE transactions on visualization and computer graphics*, 19(4):616–625, 2013. 1

[2] D. Borrmann, J. Elseberg, K. Lingemann, and A. Nüchter. The 3d hough transform for plane detection in point clouds: A review and a new accumulator design. *3D Research*, 2(2):1–13, 2011. 2

[3] A. Chalmers, F. Zaman, and T. Rhee. Avatar360: Emulating 6-dof perception in 360° panoramas through avatar-assisted navigation. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 630–638. IEEE, 2024. 1

[4] A. Chalmers, F. Zaman, A. Stangnes, S. Finnie, H. S. Nguyen, J. Han, and T. Rhee. Real-time auditorium modeling and visual effects for live performances. In *ACM SIGGRAPH Asia 2023 Real-Time Live!*, pp. 1–1. 2023. 2

[5] R. Du, E. Turner, M. Dzitsiuk, L. Prasso, I. Duarte, J. Dourgarian, J. Afonso, J. Pascoal, J. Gladstone, N. Cruces, et al. Depthlab: Real-time 3d interaction with depth maps for mobile augmented reality. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pp. 829–843, 2020. 2

[6] H. Fuchs, A. State, and J.-C. Bazin. Immersive 3d telepresence. *Computer*, 47(7):46–52, 2014. 1

[7] H. Gao, K. Yue, S. Yang, Y. Liu, M. Guo, and Y. Liu. Exploring depth-based perception conflicts in virtual reality through error-related potentials. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 774–784. IEEE, 2024. 2

[8] A. Holynski and J. Kopf. Fast depth densification for occlusion-aware augmented reality. *ACM Transactions on Graphics (ToG)*, 37(6):1–11, 2018. 2

[9] W. A. IJsselsteijn. History of telepresence. *3D Videocommunication: Algorithms, Concepts and Real-Time Systems in Human Centred Communication*, pp. 5–21, 2005. 1

[10] S. Liao, Y. Zhou, and V. Popescu. Ar interfaces for disocclusion—a comparative study. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 530–540. IEEE, 2023. 2

[11] C. Liu, K. Kim, J. Gu, Y. Furukawa, and J. Kautz. Planercnn: 3d plane detection and reconstruction from a single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4450–4459, 2019. 2

[12] M. C. Macedo and A. L. Apolinario. Occlusion handling in augmented reality: Past, present and future. *IEEE Transactions on Visualization and Computer Graphics*, 29(2):1590–1609, 2021. 2

[13] M. Maslych, Y. Hmaiti, R. Ghamandi, P. Leber, R. K. Kattoju, J. Belga, and J. J. LaViola. Toward intuitive acquisition of occluded vr objects through an interactive disocclusion mini-map. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 460–470. IEEE, 2023. 2

[14] S. Orts-Escolano, C. Rhemann, S. Fanello, W. Chang, A. Kowdle, Y. Degtyarev, D. Kim, P. L. Davidson, S. Khamis, M. Dou, et al. Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th annual symposium on user interface software and technology*, pp. 741–754, 2016. 1

[15] L. Petikam, A. Chalmers, and T. Rhee. Visual perception of real world depth map resolution for mixed reality rendering. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 401–408. IEEE, 2018. 2

[16] J. Poppinga, N. Vaskevicius, A. Birk, and K. Pathak. Fast plane detection and polygonalization in noisy 3d range images. In *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3378–3383. IEEE, 2008. 2

[17] T. Rhee, A. Chalmers, W. K. Hoh, R. Roberts, W. Butcher, S. Finnie, and R. Barrett. Teleport to the augmented real-world with live interactive effects (ifx). In *Proceedings of the SIGGRAPH Asia 2022 Real-Time Live!*, pp. 1–1. 2022. 2

[18] T. Rhee, A. Chalmers, I. Loh, B. Allen, L. Petikam, S. Thompson, and T. Revill. Mixed reality 360 live: live blending of virtual objects into 360° streamed video. In *ACM SIGGRAPH 2018 Real-Time Live!*, SIGGRAPH '18, 2018. 2

[19] T. Rhee, A. Chalmers, F. Zaman, A. Stangnes, and V. Roberts. Real-time stage modelling and visual effects for live performances. In *ACM SIGGRAPH 2023 Real-Time Live!*, pp. 1–2. 2023. 2

[20] T. Rhee, L. Petikam, B. Allen, and A. Chalmers. Mr360: Mixed reality rendering for 360 panoramic videos. *IEEE transactions on visualization and computer graphics*, 23(4):1379–1388, 2017. 1, 2

[21] T. Rhee, S. Thompson, D. Medeiros, R. Dos Anjos, and A. Chalmers. Augmented virtual teleportation for high-fidelity telecollaboration. *IEEE transactions on visualization and computer graphics*, 26(5):1923–1933, 2020. 1

[22] S. Rokhsaritalemi, A. Sadeghi-Niaraki, and S.-M. Choi. A review on mixed reality: Current trends, challenges and prospects. *Applied Sciences*, 10(2):636, 2020. 2

[23] T. Schubert, F. Friedmann, and H. Regenbrecht. The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, 10(3):266–281, 2001. 2

[24] X. Teng, R. S. Allison, and L. M. Wilcox. Manipulation of motion parallax gain distorts perceived distance and object depth in virtual reality. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 398–408. IEEE, 2023. 1

[25] F. Westermeier, L. Brübach, M. E. Latoschik, and C. Wienrich. Exploring plausibility and presence in mixed reality experiences. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2680–2689, 2023. 1

[26] F. Westermeier, L. Brübach, C. Wienrich, and M. E. Latoschik. Assessing depth perception in vr and video see-through ar: A comparison on distance judgment, performance, and preference. *IEEE Transactions on Visualization and Computer Graphics*, 2024. 1

[27] M. Yang and W. Förstner. Plane detection in point cloud data. Workingpaper, University of Bonn, 2010. 2

[28] F. Zaman, C. Anslow, A. Chalmers, and T. Rhee. Mrmac: Mixed reality multi-user asymmetric collaboration. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 591–600, 2023. doi: 10.1109/ISMAR59233.2023.00074 1