MRUnion: Asymmetric Task-Aware 3D Mutual Scene Generation of Dissimilar Spaces for Mixed Reality Telepresence

Michael Pabst[1,2,3], Linda Rudolph[1], Nikolas Brasch[1], Verena Biener[2], Chloe Eghtebas[1], Ulrich Eck[1], Dieter Schmalstieg[2], and Gudrun Klinker[1]

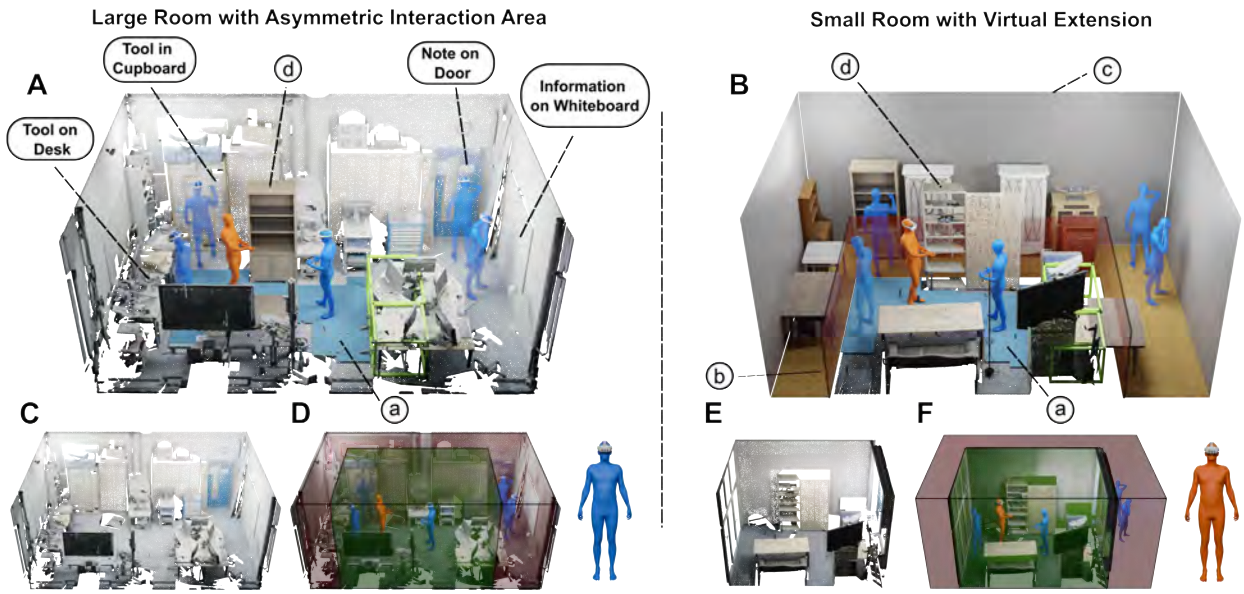

Fig. 1: The blue user in the large room (A/C) is able to access the entire space in the *Union* room layout during collaboration and can use different kinds of physical tools and information artifacts (desk, cupboard, door, whiteboard) without disappearing from the view of the orange user, who is in the small room (B/E). The smaller room is virtually extended to allow collaboration not only in the intersection area (shaded green in D, F), but also in the extended area (shaded red in D, F). To communicate spatial limitations in the generated scenes (A, B), we (a) highlight the mutual space, (b) show semi-transparent walls, (c) synthetic floors, ceilings, and walls, and (d) virtual furniture.

Abstract—In mixed reality (MR) telepresence applications, the differences between participants' physical environments can interfere with effective collaboration. For asymmetric tasks, users might need to access different resources (information, objects, tools) distributed throughout their room. Existing intersection methods do not support such interactions, because a large portion of the telepresence participants' rooms become inaccessible, along with the relevant task resources. We propose MRUnion, a Mixed Reality Telepresence pipeline for asymmetric task-aware 3D mutual scene generation. The key concept of our approach is to enable a user in an asymmetric telecollaboration scenario to access the entire room, while still being able to communicate with remote users in a shared space. For this purpose, we introduce a novel mutual room layout called *Union*. We evaluated 882 space combinations quantitatively involving two, three, and four combined remote spaces and compared it to a conventional *Intersect* room layout. The results show that our method outperforms existing intersection methods and enables a significant increase in space and accessibility to resources within the shared space. In an exploratory user study (N=24), we investigated the applicability of the synthetic mutual scene in both MR and VR setups, where users collaborated on an asymmetric remote assembly task. The study results showed that our method achieved comparable results to the intersect method but requires further investigation in terms of social presence, safety and support of collaboration. From this study, we derived design implications for synthetic mutual spaces.

Index Terms—mixed reality, mutual space, telepresence.

1 Introduction

Traditional 2D screen-based conferencing systems restrict interactions to a flat representation, limiting the immersive experience. As augmented reality (AR), virtual reality (VR), and mixed reality (MR) technologies advance, we observe an increasing demand to have shared virtual spaces where remote users can interact as naturally as if they were copresent [36]. These MR telepresence systems improve human and social interactions by enabling a sense of co-presence among remote users [48, 60].

However, in most telepresence scenarios, we face dissimilar spaces that differ in size and interior arrangements. Therefore, when insufficient spatial correspondences between local surroundings are established, only a fraction of a participant's room is accessible to the telepresence application, and collaborative interaction with physical objects in the non-shared areas is unavailable. In asymmetric collaboration

- [1]*Technical University of Munich,* [2]*University of Stuttgart and* [3]*International Max Planck Research School for Intelligent Systems.*
- *E-mail: michael.pabst@visus.uni-stuttgart.de*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

tasks, where users require different information artifacts and tools, limited sharing of space may adversely affect the collaborative task.

Furthermore, the lack of spatial correspondences between remotely shared spaces can lead to implausible behavior, potentially resulting in remote avatars colliding with physical objects, facing line-of-sight problems, or disappearing [1]. These issues can disrupt users' sense of co-presence [56]. Recent work on mutual space generation for remote users considers each user's physical surroundings and constraints [24, 25, 28, 31]. However, all of these works focus on functional subareas for collaboration and do not allow access to the full room during collaboration.

We present MRUnion (Fig. 1), a novel method for asymmetric task-aware mutual scene generation. MRUnion enables users to collaborate within a synthetic mutual space, taking into account users' spatial constraints, while also visually communicating these constraints to the other users. Our method allows users to access all physical space for collaboration.

Our quantitative evaluation showed that our method succeeded in providing more space for collaboration and increasing access to resources distributed in the room for the collaborative task. In a user study, we evaluated our approach in an asymmetric collaboration task, both in MR and VR. The study demonstrates that our method represents an effective alternative to conventional methods. However, under the current study design, users favored the Intersect layout specifically in the metrics of social presence, safety and support of collaboration, suggesting the need for further investigation of these aspects in future studies. In addition, we introduce a novel visualization to better communicate the spatial limitations of remote users. Our results indicate that these concepts are helpful for collaboration. The contributions of this paper are as follows:

1) A method for 3D mutual scene generation in mixed reality telepresence, including layout generation, virtual furniture generation, and visualization of spatial limitations
2) A taxonomy of mutual space visualization in dissimilar spaces for telepresence scenarios
3) A quantitative evaluation on 882 scene combinations highlighting the effectiveness of our method in providing more space and access to resources in highly dissimilar spaces
4) A user study of MRUnion across two distinct mutual scene layouts for users of MR and VR, confirming that our method supports collaborative telepresence scenarios

2 Related work

2.1 Asymmetric telecollaboration in MR

Various studies have explored symmetric and asymmetric telecollaboration systems [8, 44]. A key difference here is that users of asymmetric telecollaboration systems take on different roles in different locations, characterized by distinct responsibilities, levels of authority, and access to information. Asymmetric telecollaboration has been utilized in various fields, including multi-user asymmetric collaboration [65, 66], remote assistance [10, 12, 15, 35, 38, 52] and teleconsultation in medicine [14, 43, 49, 63, 64]. Previous work has investigated the use of mobile devices for asymmetric telecollaboration [35], in which visual instructions can be annotated on a target object. Other work uses live 360° video for remote collaboration which allows the remote user to observe the task space of the local user [33, 42, 50, 65]. However, these systems restrict the remote user's movement to the 360° camera's view and limit interaction with physical objects, relying on the local user for engagement in the real environment. Additionally, the remote user operates in VR, disconnecting them from their physical space, which may limit access to tools or information needed to assist the local user effectively. Johnson et al. [20] integrated a live 3D view of the local user's task space with a static 3D reconstruction of the remote expert's user space. However, in room-scale tasks involving dissimilar spaces, experts may lose track of users when the local user's task space exceeds the bounds of the remote user's space. This issue becomes even more relevant when incorporating 3D avatar representations within the collaboration space, as demonstrated in recent work on remote telecollaboration to enhance co-presence and workspace awareness [9, 65, 66]. However, in such work, it is assumed that the required tools and information are close at hand and that the remote user does not need to access resources beyond the collaboration space in their physical environment. This assumption often arises from simplifying the system, neglecting the possibility of the avatar disappearing, which can break immersion. [1, 56]. Therefore, none of these works have investigated how dissimilar spaces can affect asymmetric remote collaboration tasks and how a shared space should be created to enable asymmetric telecollaboration in dissimilar spaces without disrupting the immersive experience. Addressing this challenge is the focus of our work.

2.2 Mutual space generation in telepresence

Prior work on mutual spatial alignment is based on finding the intersection area of dissimilar spaces [24, 25, 34]. However, in MR, this method limits the total interaction area to the intersection space of the input spaces. Other methods rely on redirected walking (RDW) for mutual space generation [26, 27, 29, 30]. Kim et al. [29] employ this method using relative translation gains to stretch the space along the x-axis and y-axis to obtain a larger mutual walkable space of the input rooms. However, their approach is limited to only walkable spaces, and thus fails to enable interaction with physical objects. Keshavarzi et al. [25] introduce a method for the generation of a synthetic mutual scene, including walkable, sittable, and workable spaces. This system is designed specifically for VR experiences. Although virtual objects are aligned with physical objects, the generated scenes cannot be used directly in MR environments, because only a fully virtual scene is created. Kim et al. [31] generated a shared virtual space using affordance graphs to consider the relationships between objects and cluster them, allowing interaction with multiple objects in the shared space. Other work [16, 17, 19] has used object-centric alignment (e.g., with tables or whiteboards) for spatial alignment of rooms. More recent work [28] has focused on collaborative, interactable subspaces to enable interaction within a shared target area. The resulting shared space allows interaction beyond the target object, but is still restricted to the intersection area of the aligned spaces.

All of the methods mentioned above allow for the creation of mutual spaces, but a key limitation arises when aligning a large space with a small one. In this scenario, a significant portion of the large space becomes inaccessible for collaboration due to the restrictions of the smaller space, leading to narrow collaborative areas and limited access to tools and information distributed within the physical room. In our work, we address this limitation by generating a shared mutual space that allows interaction even in the inaccessible parts of the room.

2.3 Overcoming space restrictions in telepresence

Besides mutual space generation, there is also a need to access areas beyond the collaborative subspace in order to utilize more space from the user's own physical environment. Previous work has investigated avatar retargeting [6, 58, 61, 62] and motion adaptation [53] to grant users full access to the entire space during a telepresence scenario. Yoon et al. [62] proposed a learning-based approach to avatar retargeting, enabling collaboration across the entire space in a telepresence scenario. This method places the avatar in a new position that best preserves the semantics, including interaction, pose, and functional aspects. Even when users can access the entire space, the delay in transition when the avatar is repositioned remains a primary drawback, since it can be disruptive in remote collaboration. Wang et al. [53] addressed this limitation by reducing the transition delay and allowing a smooth transition between consecutive motions of the remote user. However, their work focused on scenarios that involve lengthy transitions between two user positions, which interfere with the goal of dynamic collaboration. Other works used RDW to overcome space restrictions [26, 30, 41, 54]. However, a key limitation of RDW methods is the lack of interaction with physical objects in the environment as a result of the distortion of the virtual and physical space.

To transcend the limitations of physical spaces in telepresence experiences, it is feasible to employ virtual extensions in MR. Recent studies, such as those of Kim et al. [32], have explored the impact of various room distortion effects on user perspectives, including the use of elongation distortion to virtually extend spaces. This approach not only enhances the sense of spatial freedom, but also opens up new possibilities for remote collaboration and interaction, allowing users to interact within a seemingly larger virtual environment while remaining within their actual physical confines. We expand on this idea by sharing physically inaccessible portions of the remote space. Although these portions cannot be actively roamed, they can at least be visually monitored.

3 Methodology

In this section, we propose a pipeline for 3D mutual scene generation for asymmetric tasks that require full-room access and present a taxonomy of mutual space visualization in dissimilar spaces. Fig. 2 provides an overview of the MRUnion pipeline. Given the collection of input rooms of the telepresence participants, we generate an asymmetry-aware mutual scene. To achieve this, we 1) extract semantic information from the scenes, 2) apply context-aware spatial matching to initialize the collaborative area, 3) generate an asymmetry-aware room layout that allows participants to fully utilize their space within the collaborative task, and 4) generate asymmetry-aware virtual furniture to communicate spatial limitations among participants.

3.1 Semantic extraction

The rooms of each telepresence participant are subsequently provided to the system as a collection of semantically labeled bounding boxes along with their room layout (floor, walls, and ceiling) as highlighted in Fig. 2. These are obtained through surface reconstruction [21, 22], semantic segmentation [5, 39, 40, 51], object bounding box estimation [2] and room layout estimation [3]. In this work, we focus on rooms with four walls (E1 - E4, Fig. 2). The bounding boxes are categorized into functional categories such as *non-contact space* (red in Fig. 2a), and *asymmetric task space* (yellow and green in Fig. 2a). The detected floor is classified as *walkable space*, consisting of a collision-free space, in which the user can walk freely. Non-contact, and asymmetric task spaces are connected to 3D objects in a scene. Hence, all objects in the scene that can be potential sources of information and tools are considered *asymmetric task space*. All other objects that do not belong to these categories are classified as *non-contact space*. We further divide *asymmetric task space* into two subgroups: *asymmetric workable space* and *asymmetric resource space*. Asymmetric *workable space* includes desks, tables, and similar items, while *asymmetric resource space* describes all other objects, information, and tools that are potentially required for the asymmetric task.

3.2 Context-aware spatial matching

In this work, we focus on two collaboration categories, namely floor- and table-centric collaboration. We calculate the corresponding maximum mutual function by considering both the *mutual walkable space*, and *mutual workable space*. For spatial matching, we adapt a previously utilized method to align spaces by calculating optimal mutual functional areas, aiming to find a transformation that maximizes the intersection of spaces [23, 25, 34]. As we work with an axis-aligned mesh and use axis-aligned bounding boxes, we can use a simplified optimization approach that focuses on translation within the floor plane, while allowing only right-angle rotations. We select the variant that yields the largest intersection space between the input rooms. We use a weighted sum of *mutual walkable space* and *mutual workable space* for our objective. Fig. 2, b) illustrates the spatial match between the input rooms and highlights the resulting spatial alignment and mutual space.

3.3 Asymmetry-aware room layout generation

Based on the spatial alignment in the previous step, the room layout is generated in this step. The intersection of the two scenes would restrict the user in the larger space to the brownish-shaded area (Fig. 2c), as they would disappear behind the walls for the user in the smaller space when moving or interacting in the red-shaded area (Fig. 2c). Therefore, non-intersecting areas of the rooms are virtually extended for users in smaller rooms by removing walls inside the *Union* space and keeping those that are part of the union boundary. This allows free movement within non-overlapping areas for the user in the larger room while staying within their physical confines. We refer to this as the asymmetric interaction area, highlighted in light blue in Fig. 2c. Additionally, this approach enables each user to visually observe the movements of the other participant, even in areas that cannot be physically accessed. Users can fully utilize their own space during collaboration, except in *non-contact space* areas. We refer to this as the *Union* layout, which combines all scenes into a single space. This concept is illustrated in Fig. 2, c).

3.4 Asymmetry-aware virtual furniture generation

Based on the generated room layout, virtual furniture is asymmetrically placed in the scenes to communicate the users' spatial limitations. As highlighted in Fig. 2, d) we display only the virtual furniture from the remote space because a physical object is already present in the local room, making overlaying it with virtual furniture unnecessary. Fig. 2, d) shows objects from the large space (L1-11) that are virtually displayed in the small space, referred to as V1-11, and vice versa. Our method removes virtual furniture from the scene when it overlaps with physical objects beyond a certain threshold. Physical contact objects that collide with non-contact objects are categorized as non-contact objects. Virtual furniture highlighted in red (Fig. 2d) corresponds to objects that could cause virtual collisions, while the virtual objects highlighted in orange (Fig. 2d) are inaccessible to the user in the small space and serve only to indicate the spatial limitations of the user in the large space. Once all bounding boxes are generated and categorized, we retrieve the most similar mesh model from 3D-FUTURE [11] based on the dimensions of the bounding box and the class category label, utilizing Euclidean distance as a measurement.

3.5 Visualization of spatial limitations

Mutual space In the application, we have visually distinguished the mutual space to enhance user awareness and understanding of the shared environment. The *mutual walkable space* (Fig. 3a) is visually represented in blue as a flat floor surface, providing a clear indication of the areas where both users can freely move and collaborate. Additionally, for areas designated as *mutual workable space*, we employ bounding boxes which are colored in light green to highlight these regions.

Semi-transparent walls We highlight inaccessible areas, using semi-transparent walls to enhance spatial awareness (Fig. 3b). The virtual walls are intended to signal to the user the physical boundaries in the respective physical environment. This approach is useful in situations where the placement of virtual furniture might limit another user's movement within their own space. By using virtual walls that appear only for users who are facing physical constraints, we provide a flexible solution that maintains the open navigability of the environment for remote users.

Synthetic floors, ceilings, and walls We added a synthetic floor, ceiling and walls to the shared scene to expand the space for the orange user in the smaller room (Fig. 3c). This decoration indicates to the user in the smaller room the spatial limitations of the combined space.

Virtual furniture We show virtual furniture that serves as an intuitive visual cue, indicating to users that the space is occupied by a physical object in one of the remote spaces and should be avoided, thus preventing any potential virtual collision (Fig. 3d). Additionally, we add virtual furniture, which is outside the physical boundaries for the user of the small space, but provides an intuition of the spatial limitations of the user in the large space. We have chosen virtual furniture as a visual cue because it allows seamless integration into a user's physical environment along the physical furniture in the room.

4 Experimental prototype

The scenes generated by the MRUnion pipeline were integrated into an experimental prototype, enabling their application in MR telepresence scenarios.

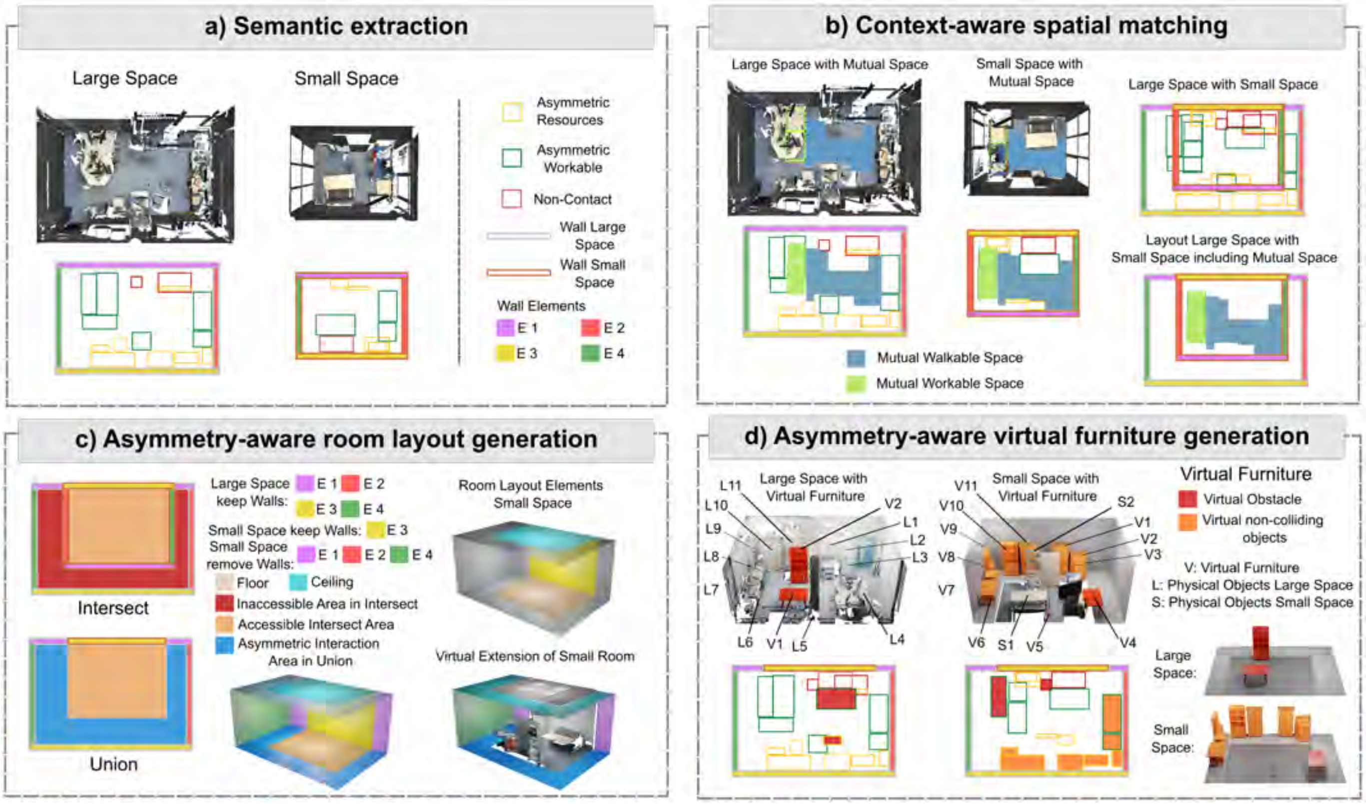

Fig. 2: Scene generation for a floor- and table-centric collaboration (*mutual walkable space* + *mutual workable space*): a) Objects and layout are semantically extracted from the input rooms, b) context-aware spatial matching to obtain initial alignment and mutual space, c) room layout is generated to virtually extend space in the smaller room, allowing asymmetric interaction in the blue-shaded area for the user in the large space, d) virtual furniture is generated to communicate spatial limitations between users (In orange and red is the virtual furniture highlighted in the corresponding room: the left shows the large space with L1-11 selected physical objects and V1-2 virtual objects from the small space, while the middle shows the small space with S1-2 selected physical objects and V1-11 virtual objects from the large space).

4.1 System design

The synthetic mutual scene that is calculated by the MRUnion pipeline serves as input for the experimental prototype. We implemented our prototype using the Unity game engine[1]. We used Photon[2] Fusion for multiplayer capabilities and Photon Voice to provide real-time audio communication between players. For development, we used the Meta XR Core Software Development Kit (SDK), Meta XR Interaction SDK, and the Meta Avatars SDK to visualize 3D avatars[3]. In our experiments, a Meta Quest 3 head-mounted display (HMD) was used, and we deployed and ran the application directly on the device without any connection to an external computer. For the experiments with MR, we used the pass-through mode of the Meta Quest HMD. Upon first use, each user is required to generate a scene model of their room using the HMD, followed by a calibration step to align the precomputed mutual scene with the user's actual environment.

4.2 Scene representation

The method for obtaining a scene representation of the rooms of our labs follows the procedure presented by Yeshwanth et al. [59]. The scans from our rooms are generated using a Faro Focus Premium laser scanner. Poisson surface reconstruction is applied on measured point clouds [21, 22] to obtain a mesh surface representation, further simplified using quadric edge collapse [13]. Finally, the segments are annotated in a 3D web interface to get a semantic scene representation. The resulting simplified mesh of the large room, along with its semantics, is shown in Fig. 4.

[1]https://unity.com/
[2]https://www.photonengine.com/
[3]https://developer.oculus.com/

4.3 Calibration

To integrate the synthetic mutual space, calibration and anchor alignment between the involved scenes are required.

Mutual space calibration The calculation of the mutual space is conducted offline, based on 3D scans of the rooms involved. Integrating this mutual space into the real environment requires calibration within each participant's physical room. This calibration is achieved through the placement of spatial anchors, which serve as reference points to align the virtual mutual space with the real-world environment. In a telepresence scenario, each participant must manually calibrate their room using the spatial anchor, because we do not have access to the Meta Quest 3 camera stream. We used a visualization of the rooms' corresponding bounding boxes for all objects in the scene as a reference for accurate placement of the spatial anchor. The calibration process with the highlighted bounding boxes can be seen in Fig. 5, a).

Anchor alignment After placing and storing the spatial anchors of each player within the scene model, the rooms need to be aligned via the spatial anchor. Additionally, the pose of the Avatar needs to be considered and transformed relative to the spatial anchor. The largest room is taken as the reference anchor, and all other rooms are aligned with the reference anchor (Fig. 5b).

5 Evaluation of Mutual Scene Generation

To perform a quantitative evaluation, we used two metrics to evaluate our method *Union* and compared it with the *Intersect* method. We are particularly interested in investigating the increase in available space and the number of objects to interact with, compared to the *Intersect* method.

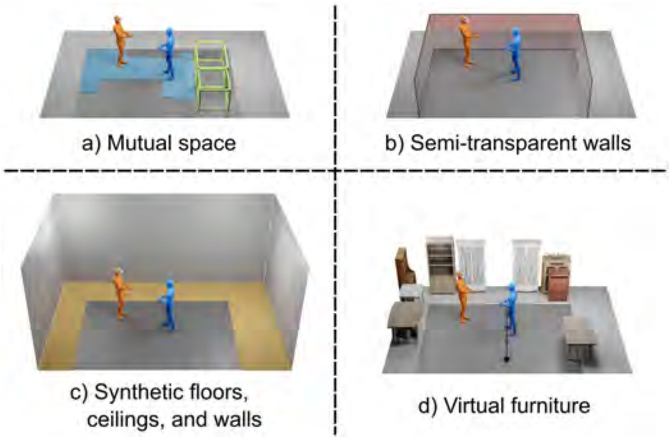

Fig. 3: Taxonomy of mutual space visualization: We show a) mutual space, b) walls made semi-transparent to indicate physical boundaries, c) added synthetic floors, ceilings, and walls to visually extend the space for the orange user in the smaller room, and d) added virtual furniture to mark inaccessible space.

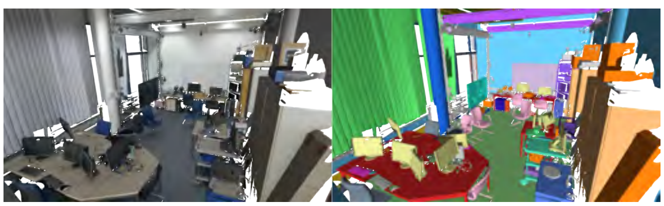

Fig. 4: Simplified mesh on the left and the semantically annotated mesh on the right.

5.1 Evaluation design and quantitative metrics

We compare our approach, the *Union* method, with the *Intersect* method, which takes the intersection of aligned rooms for collaboration and ignores the inaccessible parts of the scene for both users. Therefore, users in larger spaces can only use the intersection area, not the rest of their space.

For the quantitative evaluation, we used two metrics: First, we measured the resulting room size, which is the area accessible to the user in their room. Second, we measured the number of total interactable objects. Our objective was to investigate how many objects could serve as potential sources of physical tools or information artifacts required for the collaborative task.

5.2 Evaluation setup

To evaluate our method for different combinations of scenes, a dataset consisting of a large number of scenes is required. For the scene dataset we used ScanNet [7]. We used SceneCAD [3] to obtain the room layout, including floor, walls, and ceiling. In our evaluation, we used 14 spaces from the scene dataset, which we grouped into seven large scenes, $L = 7$ ($\mu = 37.18$ m^2, $\sigma = 12.96$ m^2), and seven small scenes, $S = 7$ ($\mu = 10.52$ m^2, $\sigma = 2.05$ m^2). We selected different types of scenes for our evaluation, including living room, lobby, classroom, dining room, computer cluster, workshop, office, and conference room. L1-S1 refers to one large space and one small space. Furthermore, we evaluated configurations for L2-S2 (two large, two small), L1-S2 (one large, two small) and L1-S3 (one large, three small). The total number of combinations is calculated using $\binom{L}{n_{large}} \times \binom{S}{n_{small}}$, where n_{large} is the number of large scenes, and n_{small} is the number of small scenes. The total combinations resulting are shown in Table 1. To this end, we evaluated our approach on 882 space combinations.

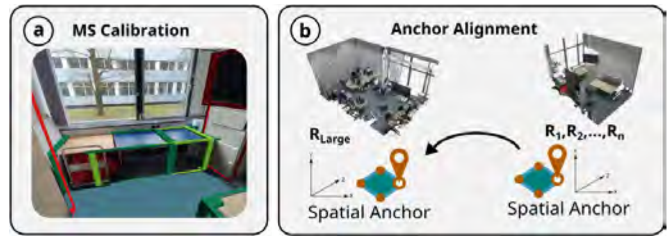

Fig. 5: a) scene calibration (the green and red bounding boxes must be aligned with the physical objects in the environment), b) anchor alignment

Table 1: Number of combinations.

Combination	Total Combinations
L1-S1: $n_{Large} = 1, n_{Small} = 1$	49
L1-S2: $n_{Large} = 1, n_{Small} = 2$	147
L1-S3: $n_{Large} = 1, n_{Small} = 3$	245
L2-S2: $n_{Large} = 2, n_{Small} = 2$	441

5.3 Results

Room Size For all combinations, we compared the original room sizes of the individual rooms with the resulting room sizes for both the *Union* and *Intersect* method using the Wilcoxon signed-rank test (Fig. 6). This test showed that the resulting room size for the *Union* method ($\mu = 44.63m^2, \sigma = 12.83$) was significantly larger than the original room space ($\mu = 21.26m^2, \sigma = 15.54$, $p < 0.001$), while the *Intersect* method ($\mu = 7.53m^2, \sigma = 1.49$) resulted in significantly smaller sizes than the original room ($p < 0.001$). We found the same results when looking at small and large rooms separately: For the small rooms, *Union* resulted in significantly larger spaces ($\mu = 43.62m^2, \sigma = 12.93$) than the original room ($\mu = 10.52m^2, \sigma = 2.05$, $p < 0.001$), and *Intersect* ($\mu = 7.43m^2, \sigma = 1.50$), in significantly smaller spaces than the original room ($p < 0.001$). Also, for the large rooms, *Union* resulted in significantly larger spaces ($\mu = 46.13m^2, \sigma = 12.53$) than the original room ($\mu = 37.18m^2, \sigma = 12.96$, $p < 0.001$), and *Intersect* ($\mu = 7.68m^2, \sigma = 1.48$), in significantly smaller spaces than the original room ($p < 0.001$). The same patterns were also observed when looking at the different combinations (L1-S1, L1-S2, L1-S3, L2-S2) separately.

Interactable Objects For the *Union* method, all interactable objects of the original rooms remain present in the shared space. However, this is not the case for the *Intersect* method. Therefore, we compared for all combinations the total number of interactable objects in a room with the number of objects that are still present in the shared space (Fig. 7). The Wilcoxon signed-rank test showed that there were significantly fewer objects after applying the *Intersect* method ($\mu = 2.01, \sigma = 2.16$) than originally present in the room ($\mu = 8.67, \sigma = 5.81$, $p < 0.001$). We found the same results when looking at the small and large rooms separately: For small rooms, applying the *Intersect* method ($\mu = 2.84, \sigma = 2.36$) resulted in significantly less objects than the ones originally present ($\mu = 7.29, \sigma = 4.56$, $p < 0.001$). Also, for large rooms, applying the *Intersect* method ($\mu = 0.78, \sigma = 0.89$) resulted in significantly less objects than originally present ($\mu = 10.71, \sigma = 6.78$, $p < 0.001$). The same patterns were also observed when looking at the different combinations (L1-S1, L1-S2, L1-S3, L2-S2) separately.

6 Exploratory User Study

Our user study aims to evaluate our method for asymmetry-aware scene generation in a real MR telepresence scenario. We opted for an exploratory user study, and we did not aim to prove any hypothesis; rather, we are interested in analyzing how the room layout *Union* compares to the *Intersect* with respect to user experience and if there are differences between the modes (*MR* and *VR*). Therefore, we explore the following research questions:

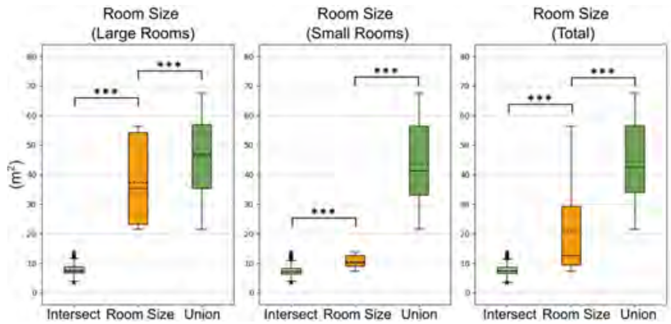

Fig. 6: Room Size results. The number of stars specifies the significance levels between the factors indicated by: *** <0.001, ** <0.01, * <0.05.The dotted line indicates the mean.

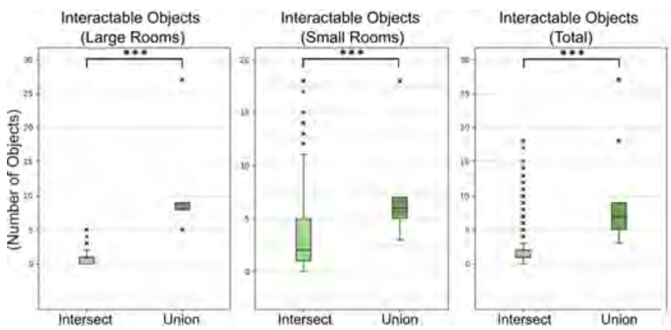

Fig. 7: Interactable Objects results. The number of stars specifies the significance levels between the factors indicated by: *** <0.001, ** <0.01, * <0.05.

RQ1 How does the layout (*Union* vs. *Intersect*) of a synthetic mutual space in an asymmetric MR remote collaboration task affect spatial presence, social presence, and perceived workload?

RQ2 How does the mode of mediation (*MR* vs. *VR*) of a synthetic mutual space in an asymmetric MR remote collaboration task affect spatial presence, social presence, and perceived workload?

RQ3 How are user preferences influenced by dissimilar user spaces during an asymmetric MR remote collaboration task, and how does this influence vary based on the mode of mediation and the layout of the synthetic mutual space?

6.1 Study design

In the user study, we compare two layouts (*Union* and *Intersect*) and two modes of mediation (*MR* and *VR*) using a within-subject design. This results in four conditions, which we evaluate in an asymmetric collaborative assembly task setup using previously generated scenes for MR, which are incorporated into the real environment (see Fig. 1). For the *Intersect* condition in MR, we additionally augment an intersection wall in the larger space. For VR, we merge all virtual furniture into one scene and adapt the boundary based on the *Intersect* or *Union* layout. The four conditions are given below.

MR+Intersect In this condition, both users only use the intersection space for collaboration and need to solve the task in MR, allowing them to view their respective real-world spaces.

VR+Intersect In this condition, the setup is similar to *MR+Intersect* except that users are completely in virtual space.

MR+Union In this condition, both users once again solve the task in MR. This situation differs from the *MR+Intersect* case in that the union of the users' rooms is utilized for collaboration. This setup enables both users to utilize the full space in their respective rooms for collaboration. The setup is illustrated in Fig. 9.

VR+Union This condition is similar to *MR+Union*, except that users are again completely in the virtual reality environment.

6.2 Task

We designed an asymmetric collaborative brick-building task for a telepresence setup (Fig. 8), aimed at teamwork between two users: the instructor (user B) and the operator (user A). The instructor (who is always in the smaller room) has a blueprint and a parts list to build a virtual brick structure. Each brick in the blueprint is identified by numbers visible to the instructor, but without specifying the color. The operator, who lacks access to the blueprint, performs the assembly process and knows the color-number mapping of the bricks. Each task involves a unique number-color mapping to ensure uniqueness and prevent any memory effect. The assembly task consists of 9 to 12 bricks in two distinct shapes and different colors. For each condition, we designed a unique brick assembly task, all of which are similar in difficulty. The first brick for each condition is already placed on the brick board to provide initial assistance in starting. In the conditions *MR+Intersect* and *VR+Intersect*, the bricks are confined to the intersection space. In contrast, in the *MR+Union* and *VR+Union* conditions, the bricks are spread throughout the full user rooms, forcing users from the larger space to move beyond the intersection area to use the entire space. In addition, the number-color mapping information is placed outside the intersection area to encourage the operator to move continuously beyond it during the session, gathering the required information to solve the asymmetric task. The bricks are only visible to the user responsible for collecting them, but they become visible to both once collected in the *mutual workable space*. The assembly process must occur in the *mutual walkable space*. This arrangement enables users to move around the construction, allowing them to perceive it from multiple viewpoints. To complete the assembly task, the operator must follow the instructor's instructions until the instructor is satisfied with the result. The roles do not change during the study, ensuring consistent interaction dynamics.

6.3 Participants

We recruited 24 participants divided into the following age groups: one participant aged 18-21, 13 participants aged 22-25, seven participants aged 26-29, and three participants aged 30 or older. The group consisted of 16 men (66. 67%) and eight women (33.33%). Of the participants, 14 (58.33%) identified as European, eight (33.33%), as Asian, one (4.17%), as Turkish, and one (4.17%), as Russian. Most of the participants, ten (41.7%), reported German as their primary language, followed by nine (37.5%) who reported English, and five (20.8%) who reported other languages. In seven of the 12 sessions, the participants knew each other. Among the participants, 19 (79.2%) reported having prior AR/VR experience.

6.4 Metrics

For all study conditions, we used the following metrics: spatial presence, social presence, workload, and user preference.

Spatial presence For spatial presence, we used the igroup presence questionnaire (IPQ) [45]. This questionnaire comprises 14 items and is divided into four subscales: involvement, realism, general presence, and spatial presence.

Social presence Social presence encompasses three dimensions: co-presence, psychological involvement, and behavioral engagement, and it is measured using the questionnaire from the networked minds measure of social presence [4]. For psychological involvement, we focused on mutual attention and mutual understanding. We omitted empathy, as the task does not involve interaction on an emotional level.

Workload The NASA task load index (TLX) questionnaire [18] was utilized to evaluate workload. This questionnaire is divided into six subjective subscales: mental demand, physical demand, temporal demand, performance, effort and frustration.

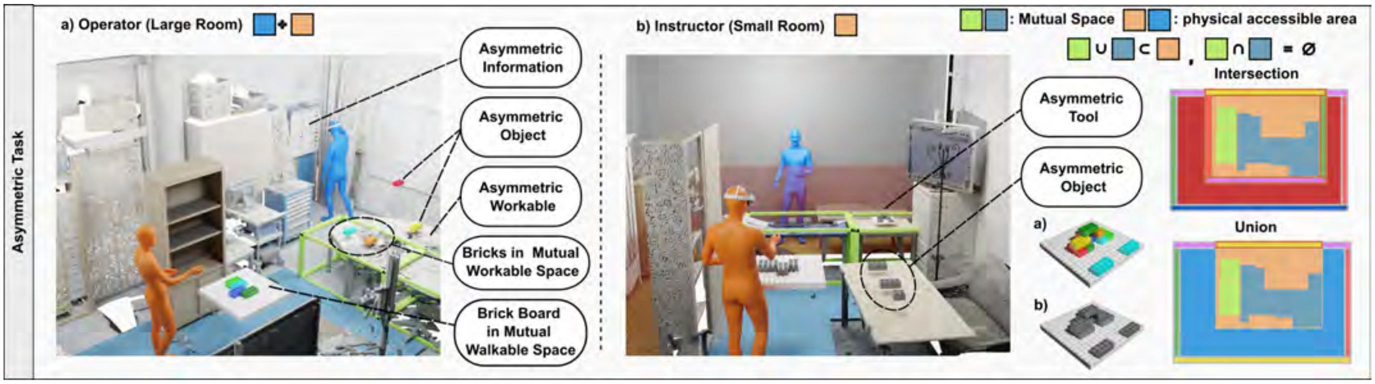

Fig. 8: Exploratory user study: It involves the operator (user A, located in the large room) and the instructor (user B, located in the small room). Both need to collect the bricks (asymmetric objects), which are partially placed in the asymmetric workable space. After collection, the bricks appear in the *mutual workable space*. The instructor provides assembly instructions to the operator based on the assembly plan (asymmetric tool). The operator checks the number-color mapping (asymmetric information) and assembles the bricks on the brick board, which is placed in the *mutual walkable space*. The operator sees the bricks in color, whereas the instructor sees them only in gray. In the intersection condition, both the operator and the instructor are confined to the intersection area, while in the union case, the operator can also utilize the asymmetric interaction area shaded in blue.

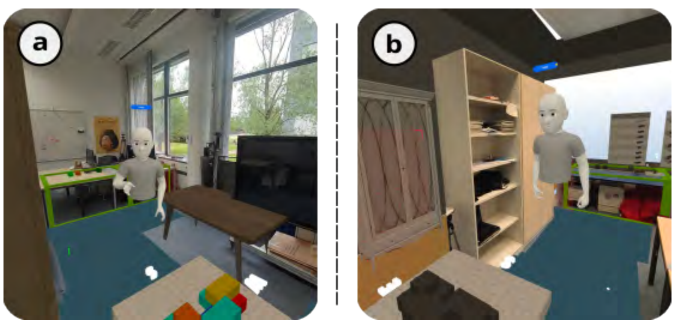

Fig. 9: a) user A in *Union* Layout, b) user B in *Union* Layout

User Preference In the user preference questionnaire, participants were asked to select their preferred conditions, mode, and layout. The preference questionnaire is categorized into the following seven categories: user comfort (which condition did you find the most comfortable?), sense of presence assessment (which condition gave you the strongest sense of being present in the remote space?), visual experience (which condition most effectively enhanced your visual experience?), user experience (which condition did you find the most enjoyable to use?), safety (which condition made you feel the safest?), collaborative task support (which condition most effectively supported your collaboration tasks?) and general feedback (which condition did you find most preferable overall?).

6.5 Procedure

The study was carried out according to the Declaration of Helsinki [57]. As the study posed no physical or psychological risks, IRB approval was not required by our institute. First, participants received an overview of the study, covering the study procedure, safety measures, and the underlying scientific investigation. Subsequently, they were asked to complete a consent form, which confirmed that they were informed of the scientific investigation and that the data about the participant would be collected and recorded anonymously. The participants then completed a demographic questionnaire and were introduced to the study environment, the Meta Quest 3, along with an overview of the tasks they would be undertaking. For each study session, two individuals were assigned the roles of instructor or operator. Throughout the study, participants maintain their initial roles and remain in the same room. The operator was located in room A, which was larger, and the instructor, in room B, which was smaller. The study consisted of two sequential phases: a single tutorial phase followed by a trial phase with four conditions. The tutorial phase aimed to equip each participant with the necessary knowledge and skills for the task. The tutorial is split into MR and VR segments to prepare participants for substantially different experiences and ensure that they become familiar with the environment. In the trial phase, all four conditions are tested. Before each condition, a small introductory video was shown explaining the condition. For each condition, they had one minute to move and perceive the environment. After that, the task material appeared, including the bricks, the brick board, and all required information, and they were asked to start the task. After each condition, participants were required to complete a questionnaire on spatial presence, social presence, and perceived workload. We used Latin squares to counterbalance the order of conditions, thereby mitigating order effects. After completion of the trial, the study participants were asked to complete a questionnaire which included user preferences and open questions. In the final step, a short interview with the participants was conducted.

6.6 Results

We report repeated measures of spatial presence, social presence, and workload responses. An overview of the user preferences across conditions is also presented. To analyze the social and spatial presence and workload, we used the Shapiro-Wilk test [47] ($\alpha = 0.05$) to check for data normality and use the aligned rank transform (ART) [55] to perform two-way repeated measures ANOVA. Furthermore, we tested for equivalence using TOST [46]. We used the χ^2 goodness-of-fit test [37] to compare the conditions regarding the user's preferences. Group A refers to Users A located in the larger room A, while Group B refers to Users B located in the smaller room B.

6.6.1 Spatial presence

The average spatial presence ratings for all conditions are shown in Table 2. A significant main effect of spatial presence was found for the mode ($F(1,23) = 10.626$, $p < .001$, $\eta_p^2 = 0.316$) such that spatial presence was significantly higher in *VR* ($\mu = 5.17$, $\sigma = 0.74$) than in *MR* ($\mu = 4.60$, $\sigma = 0.76$). We found equivalence for the layouts (*Intersect*: $\mu = 4.85$, $\sigma = 0.81$, *Union*: $\mu = 4.91$, $\sigma = 0.80$) with $p < 0.001$ in both cases (Fig. 10).

When looking at the results of groups A and B separately, we only found a significant main effect of the mode for group B ($F(1,11) = 5.544$, $p = 0.038$, $\eta_p^2 = 0.335$), such that spatial presence was significantly higher in VR ($\mu = 5.38$, $\sigma = 0.59$) than in *MR* ($\mu = 4.76$, $\sigma = 0.74$). Equivalence regarding the layouts was only found in group A (*Intersect*: $\mu = 4.74$, $\sigma = 0.83$, *Union*: $\mu = 4.64$, $\sigma = 0.84$), the larger values $p = 0.002$.

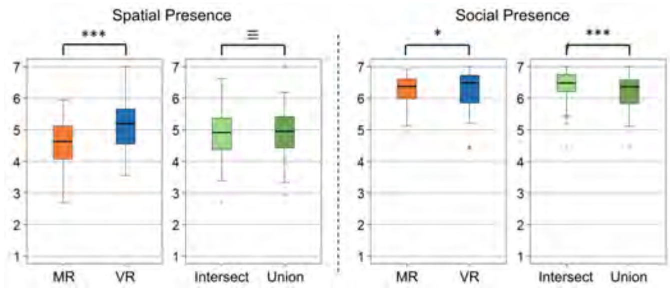

Fig. 10: Spatial and social presence results. The number of stars specify the significance levels between the factors as indicated by the RM-ANOVA: *** <0.001, ** <0.01, * <0.05. Equivalences are indicated with ≡.

Table 2: Mean μ and standard deviation σ of the three measures for the four conditions

Condition	Spatial presence	Social presence	Workload
MR+Intersect	$\mu = 4.58$ $\sigma = 0.79$	$\mu = 6.36$ $\sigma = 0.48$	$\mu = 17.22$ $\sigma = 12.19$
VR+Intersect	$\mu = 5.13$ $\sigma = 0.74$	$\mu = 6.29$ $\sigma = 0.67$	$\mu = 20.83$ $\sigma = 15.54$
MR+Union	$\mu = 4.61$ $\sigma = 0.73$	$\mu = 6.14$ $\sigma = 0.60$	$\mu = 21.11$ $\sigma = 13.46$
VR+Union	$\mu = 5.21$ $\sigma = 0.76$	$\mu = 6.27$ $\sigma = 0.63$	$\mu = 17.99$ $\sigma = 14.49$

6.6.2 Social presence

Table 2 shows the average social presence ratings for all conditions. A significant main effect of the layout on social presence was found ($F(1,23) = 16.5558$, $p < .001$, $\eta_p^2 = 0.419$) such that *Intersect* ($\mu = 6.33$, $\sigma = 0.58$) lead to significantly higher ratings than *Union* ($\mu = 6.21$, $\sigma = 0.56$). There was also a significant main effect of the mode ($F(1,23) = 4.637$, $p < .05$, $\eta_p^2 = 0.168$) such that *VR* resulted in significantly higher ratings ($\mu = 6.28$, $\sigma = 0.64$) than *MR* ($\mu = 6.25$, $\sigma = 0.50$). In the results of group A and B separately, we found the significant main effect of layout for both group A ($F(1,11) = 7.993$, $p = 0.016$, $\eta_p^2 = 0.420$) and group B ($F(1,11) = 5.642$, $p = 0.037$, $\eta_p^2 = 0.339$), with *Intersect* (A: $\mu = 6.21$, $\sigma = 0.66$, B: $\mu = 6.45$, $\sigma = 0.46$) resulting in a significantly higher social presence than *Union* (A: $\mu = 6.07$, $\sigma = 0.64$, B: $\mu = 6.35$, $\sigma = 0.45$). However, instead of a main effect of mode, we found an equivalence between *MR* (A: $\mu = 6.13$, $\sigma = 0.58$, B: $\mu = 6.37$, $\sigma = 0.37$) and *VR* (A: $\mu = 6.14$, $\sigma = 0.72$, B: $\mu = 6.43$, $\sigma = 0.52$) for both group A (larger value $p = 0.002$) and B (larger value $p = 0.001$). This could be an effect of the much smaller sample size of individual groups.

6.6.3 Workload

Table 2 shows the average task workload for all conditions. A significant interaction effect between the two factors (layout and mode) was found ($F(1,23) = 4.405$, $p < .05$, $\eta_p^2 = 0.161$). *MR-Union* ($\mu = 21.11$, $\sigma = 13.46$) resulted in a slightly higher score than *MR-Intersect* ($\mu = 17.22$, $\sigma = 12.19$), while it is the opposite for *VR*, where *VR-Union* ($\mu = 17.99$, $\sigma = 14.49$) results in a slightly lower score than *VR-Intersect* ($\mu = 20.83$, $\sigma = 15.54$), but post hoc tests show no significant differences between any pairs.

6.6.4 User preferences

We collected data on user preferences for the four conditions for seven categories (see Section 6.4). Among all conditions, 37.5% of users preferred *MR+Union*, followed by *MR+Intersect* (33.33%), *VR+Intersect* (20.83%) and *VR+Union* (8.33%).

Intersection vs union Regarding the layout, 54.17% of participants preferred *Intersect* overall. Looking at the individual categories, *Intersect* was preferred with regard to safety (75%), its ability to support collaboration (70.83%) and its comfort (58.33%). The *Union* scene layout seemed more enjoyable (54.17%) and enhanced the visual experience (58.33%). In both layouts, being present in the remote space was rated equally. Of 13 participants who enjoyed the *Intersect* layout more, participants such as S12-A (Study 12, User A) claimed, "*I liked that I could look out the window but could also see the other person's limits*". Reasons for participants who preferred the *Union* case more included how ... *in the Union layout, I could clearly identify the area where I could move and the area where my partner could move*" (S6-B). The χ^2 goodness of fit test showed that *Intersect* was significantly more preferred than *Union* with regard to both the support of the collaborative task ($\chi^2 = 4.17, p = .041$) and safety ($\chi^2 = 6.00, p = .014$). However, there is no significant difference between the *Intersect* and *Union* scene layouts in terms of general feedback. When looking at the results of groups A (large room) and B (small room) separately, these significant differences between layouts were only found for group B, which preferred *Intersect* with regard to the support of collaborative task (83. 33%, $\chi^2 = 5.33, p = .001$) and safety (91.67%, $\chi^2 = 8.33, p = .004$).

MR vs VR For the mode of mediation, users had an overall preference of *MR* over *VR* (70.83%) for both user groups (user A: 75%; user B: 66.67%). The highest user preference rating for *MR* was in the safety category, with a rating of 87.5%. In favor of the *MR* condition, S8-B said, "*being able to see the real environment was important to me because I felt that I wasn't going to hit any obstacles"*. Among all categories, the only category in which *VR* had a higher rating was the feeling of being present in remote space (66.67%) (user A: 66.67%, user B: 66.67%). The participants supported this result by claiming "*VR felt more enclosed"* (S12-B) and that "*MR makes me feel less immersed"* (S11-B). The χ^2 goodness of fit test showed that *MR* was significantly more preferred than *VR* with regard to being more comfortable (75%, $\chi^2 = 6.00, p < .05$); user (S6-A) describes it as a "... *comfortable feeling because the virtual elements are projected solely into the room, so I am still aware that I am located in the physical world.*". Additionally, *MR* was significantly preferred in terms of safety ($\chi^2 = 13.5, p < .001$) and general preference ($\chi^2 = 4.17, p = .041$). When looking at the results of groups A and B separately, the preference for *MR* was only significant in terms of safety (A: 91.67%, $\chi^2 = 8.33, p = .004$, B: 83.33%, $\chi^2 = 5.33, p = .021$).

Room size The room size also potentially impacts user preferences. Users in the larger room (user A) preferred *Union* (n = 7), while users in the smaller room (user B) preferred *Intersect* (n = 8). User S8-A reasoned that " *I don't want to lose space just because my partner's room is smaller*." when considering the *Intersect* mutual scene layout. In open text replies, the *Union* case was often praised for being "*large"* (n = 9) and more *"comfortable"* (n = 3), with participant claiming that "*it was still clear which area my partner and I could use together"* (S6-A). Conversely, S8-B felt it was "*more risky to crash into some objects due to the additional space"*. The preference of Group B for the intersection case is also reflected in their responses, where (S5-B) liked how they "*saw my own room's walls"* (S7-B) and found it "*natural* " and "*accessible*".

7 Discussion

When evaluating mutual scene generation for a variety of rooms, we saw that our *Union* method results in significantly larger room sizes and more interactable objects than the *Intersect* method. In addition, we have shown that we can generate feasible mutual spaces not only for two rooms but also for three or four rooms of different sizes.

Through our exploratory user study, we found that in most ratings, the *Union* method was comparable with the *Intersect* method. In certain cases, the two methods were equivalent.

For spatial presence, we found equivalence between the two layouts. Additionally, *VR* resulted in significantly higher ratings. We argue that this finding is because users feel more involved in a VR environment, since the real environment is tuned out. It is also underlined by the subcategory involvement, which is significantly higher for the *VR* condition compared to the *MR* condition. Furthermore, the interviews support this finding, where many users stated that the VR environment felt more immersive and that they were completely encapsulated by it.

For social presence, we identified significant main effects for both the layout and the mode of mediation with *VR* and *Intersect* leading to higher scores. We believe that users wearing a VR HMD can focus more on each other since the real environment is tuned out. In terms of layout, we think that the layout of the intersection contributed to a greater sense of social presence because the portion of the *mutual walkable space* and *mutual workable space* is larger in the *Intersect* layout than in the *Union* layout, and the mutual space strongly contributes to social presence, as it is available to all users in the same way, ensuring proximity throughout the interaction. Moreover, the study was designed so that the users in the *Intersect* condition could stay within the mutual space to solve the task. In contrast, in the *Union* layout, the user in the larger space had to constantly move out of the intersection area to access the required information.

Neither the mode of mediation nor the layout had any significant effect on perceived workload. There was a significant interaction effect between these two factors, but the post hoc tests did not show any significant differences. We speculate that the task used in the study was too easy, as it was designed to be solved quickly by users without prior experience with MR or VR.

Regarding preferences, users reported a significantly higher sense of safety and support for the collaborative task in the *Intersect* layout. This preference may be traced back to the intersection space being accessible to both users in the same way. Furthermore, we designed the task for the *Intersect* layout to be solvable within the intersection space, so no additional room space was needed. User S12-A also supported this assumption in interviews by stating that *...the intersection was sufficient for the task, and I had everything I needed close by*". It seems that safety is better supported if less movement is required to solve the task. Therefore, the *Intersect* layout felt more secure to many users. For users in the smaller room, the *Union* layout provided additional space where they needed to pay more attention to which areas they could access. The preferences across the other categories were balanced for both layouts, including the overall preference. Users significantly preferred *MR* over *VR* in the general feedback category, as well as for safety and comfort. The ability of *MR* users to see the real world seems to help their environmental awareness, making *MR* safer and more comfortable to navigate.

Through our quantitative evaluation, we observed that our approach demonstrated advantages in space availability and accessibility of interactable objects. However, our user study revealed that, in terms of user experience, our approach is limited in aspects such as social presence, safety, and support of collaboration. Interestingly, the limitations concerning safety and support of collaboration were observed among users in Group B, who were situated in the smaller room. We believe that the user experience heavily depends on the task setup. A task setup requiring the user in the larger room to leave the intersection area within the *Intersect* layout could potentially impact the outcome. However, we chose to neglect this factor and designed the task for the intersection conditions such that the users do not need to leave the intersection area to ensure a fair comparison between the two layouts.

7.1 Design implications

Based on our analysis, we propose specific design implications for synthetic mutual spaces in asymmetric task-based MR telepresence scenarios.

Telecollaboration with asymmetric mutual spaces Our study demonstrates that the *Union* room layout can effectively support asymmetric task-based telecollaboration. Therefore, we suggest incorporating the *Union* layout in MR telepresence setups when tools and information are distributed throughout the entire room.

Dynamic adaption of room layouts Users in larger spaces preferred *Union* to maintain their expansive area, while those in smaller rooms favored *Intersect* to better suit their limited space. To find a suitable compromise, we suggest dynamically adapting the room layout as preferred by the user and required by the task.

Visualization of spatial limitations The visualization of the mutual space greatly helped users understand the collaborative area accessible to both. We suggest having the possibility for users to visually highlight the mutual space in telepresence scenarios. In the *Union* room layout, where parts of the room are not accessible to everyone, these areas should also be highlighted through semi-transparent walls, synthetic walls, ceilings and floors, and virtual furniture to clearly visualize the spatial limitations for both users.

7.2 Limitations

There are limitations in our work that need to be addressed in future research. With our method, we are currently limited to processing room layouts with four walls, typically rectangular or square in shape. More complex room layouts must be explored in future efforts. In our current experimental prototype, scene generation is calculated offline. This setup was designed to evaluate our method, but would need further adaptation to allow dynamic scene generation with specific user preferences. Therefore, dynamic mutual scene generation is currently not possible with our prototype.

In the user study, we focused on single-person room setups due to space and time constraints, with only one person present in each room. Furthermore, the study involved a very simple asymmetric collaborative task designed to accommodate participants without any MR or VR experience. It would be interesting to explore more challenging asymmetric tasks that allow participants to move and interact more within the room, which was not addressed in the current study. In our study, the instructor was always in the smaller room, which could introduce a potential confounding factor. However, this was determined by the nature of our task. Future efforts should explore this variable with different task setups. Moreover, 79.2% of the participants reported prior AR/VR experience. Therefore, the potential influence of the level of experience on the outcome should be further investigated in future work with a more diverse participant group. Furthermore, in our study, we defined the information required for the asymmetric task beforehand, including the position in the room. Consequently, it is limited in terms of dynamically adjusting the space based on the tools and information the user needs. Finally, we only evaluated the setup with two rooms, yet it would also be interesting to explore collaborative tasks among people in three or more rooms.

8 Conclusion and Future Work

In this work, we presented MRUnion, a system for asymmetric-task aware 3D mutual scene generation of dissimilar spaces, and introduced a novel mutual scene layout called *Union*. Our quantitative evaluation and user study showed that the proposed room layout is a valid alternative to conventional *Intersect* methods for asymmetric tasks that require full room access. However, our study results indicated that further exploration of our proposed layout is needed, particularly in terms of social presence, safety, and support of collaboration. We propose using the *Union* layout for collaborative asymmetric tasks that require full room access. We also propose allowing users to dynamically adjust the room layout based on preferences and activities within an MR telepresence scenario, and we suggest visualizing spatial limitations.

In future work, it would be interesting to investigate the transition between intersection and union layouts more closely. This plan includes examining when and how the room should be virtually extended to allow full-room access. This question is particularly relevant for users in smaller rooms, as it could improve their experience while ensuring that all users can still access the full room. Additionally, generative models, such as diffusion models, may be explored in future efforts to improve the scene generation process by creating virtual furniture, textures, and other 3D content adapted to the user's physical environment.

Acknowledgments

The authors wish to thank Towsif Zahin Khan for the help with the user studies, Chandan Yeshwanth for the support with the 3D scans of the environment. We would like to acknowledge the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting Michael Pabst. We thank the Alexander von Humboldt

Foundation funded by the German Federal Ministry of Education and Research.

References

[1] D. Akselrad, C. DeVeaux, E. Han, M. R. Miller, and J. N. Bailenson. Body crumple, sound intrusion, and embodiment violation: Toward a framework for miscommunication in vr. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*, pp. 122–125, 2023. 2

[2] A. Avetisyan, M. Dahnert, A. Dai, M. Savva, A. X. Chang, and M. Niessner. Scan2cad: Learning cad model alignment in rgb-d scans. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 3

[3] A. Avetisyan, T. Khanova, C. Choy, D. Dash, A. Dai, and M. Nießner. Scenecad: Predicting object alignments and layouts in rgb-d scans. In *The European Conference on Computer Vision (ECCV)*, August 2020. 3, 5

[4] F. Biocca, C. Harms, and J. Gregg. The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In *4th annual international workshop on presence, Philadelphia, PA*, pp. 1–9, 2001. 6

[5] C. Choy, J. Gwak, and S. Savarese. 4d spatio-temporal convnets: Minkowski convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3075–3084, 2019. 3

[6] B. Congdon, T. Wang, and A. Steed. Merging environments for shared spaces in mixed reality. pp. 1–8, 11 2018. doi: 10.1145/3281505.3281544 2

[7] A. Dai, A. X. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5828–5839, 2017. 5

[8] N. Feld and B. Weyers. Overview of collaborative virtual environments using augmented reality. 2019. 2

[9] C. G. Fidalgo, M. Sousa, D. Mendes, R. K. Dos Anjos, D. Medeiros, K. Singh, and J. Jorge. Magic: Manipulating avatars and gestures to improve remote collaboration. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 438–448. IEEE, 2023. 2

[10] 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. 2

[11] H. Fu, R. Jia, L. Gao, M. Gong, B. Zhao, S. Maybank, and D. Tao. 3d-future: 3d furniture shape with texture, 2020. 3

[12] L. Gao, H. Bai, G. Lee, and M. Billinghurst. An oriented point-cloud view for mr remote collaboration. In *SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications*, pp. 1–4. 2016. 2

[13] M. Garland and P. S. Heckbert. Surface simplification using quadric error metrics. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pp. 209–216, 1997. 4

[14] D. Gasques, J. G. Johnson, T. Sharkey, Y. Feng, R. Wang, Z. R. Xu, E. Zavala, Y. Zhang, W. Xie, X. Zhang, et al. Artemis: A collaborative mixed-reality system for immersive surgical telementoring. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pp. 1–14, 2021. 2

[15] S. Gauglitz, B. Nuernberger, M. Turk, and T. Höllerer. World-stabilized annotations and virtual scene navigation for remote collaboration. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pp. 449–459, 2014. 2

[16] J. E. S. Grønbæk, K. Pfeuffer, E. Velloso, M. Astrup, M. I. S. Pedersen, M. Kjær, G. Leiva, and H. Gellersen. Partially blended realities: Aligning dissimilar spaces for distributed mixed reality meetings. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, article no. 456, 16 pages. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544548.3581515 2

[17] J. E. S. Grønbæk, J. Sánchez Esquivel, G. Leiva, E. Velloso, H. Gellersen, and K. Pfeuffer. Blended whiteboard: Physicality and reconfigurability in remote mixed reality collaboration. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–16, 2024. 2

[18] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, vol. 52, pp. 139–183. Elsevier, 1988. 6

[19] A. Irlitti, M. Latifoglu, T. Hoang, B. V. Syiem, and F. Vetere. Volumetric hybrid workspaces: Interactions with objects in remote and co-located telepresence. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–16, 2024. 2

[20] J. G. Johnson, T. Sharkey, I. C. Butarbutar, D. Xiong, R. Huang, L. Sy, and N. Weibel. Unmapped: Leveraging experts' situated experiences to ease remote guidance in collaborative mixed reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–20, 2023. 2

[21] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. In *Proceedings of the fourth Eurographics symposium on Geometry processing*, vol. 7, 2006. 3, 4

[22] M. Kazhdan and H. Hoppe. Screened poisson surface reconstruction. *ACM Transactions on Graphics (ToG)*, 32(3):1–13, 2013. 3, 4

[23] M. Keshavarzi, F. C. Reyes, R. Shrivastava, O. Afolabi, L. Caldas, and A. Y. Yang. Contextual scene augmentation and synthesis via gsacnet, 2021. 3

[24] M. Keshavarzi, A. Y. Yang, W. Ko, and L. Caldas. Optimization and manipulation of contextual mutual spaces for multi-user virtual and augmented reality interaction. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Mar. 2020. doi: 10.1109/vr46266.2020.00055 2

[25] M. Keshavarzi, M. Zollhoefer, A. Y. Yang, P. Peluse, and L. Caldas. Mutual scene synthesis for mixed reality telepresence, 2022. 2, 3

[26] D. Kim. Mutual space generation with redirected walking for asymmetric remote collaboration. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 985–986, 2023. doi: 10.1109/VRW58643.2023.00335 2

[27] D. Kim, H.-I. Kim, and W. Woo. Mutual space generation with relative translation gains in redirected walking for asymmetric remote collaboration. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 674–677, 2022. doi: 10.1109/ISMAR-Adjunct57072.2022.00140 2

[28] D. Kim, S. Kim, S. Choi, and W. Woo. Spatial affordance-aware interactable subspace allocation for mixed reality telepresence. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 1256–1265. IEEE, 2024. 2

[29] D. Kim, J.-e. Shin, J. Lee, and W. Woo. Adjusting relative translation gains according to space size in redirected walking for mixed reality mutual space generation. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 653–660, 2021. doi: 10.1109/VR50410.2021.00091 2

[30] D. Kim and W. Woo. Edge-centric space rescaling with redirected walking for dissimilar physical-virtual space registration. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 829–838. IEEE, 2023. 2

[31] S. Kim, D. Kim, J.-E. Shin, and W. Woo. Object cluster registration of dissimilar rooms using geometric spatial affordance graph to generate shared virtual spaces. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 796–805. IEEE, 2024. 2

[32] Y.-J. Kim, A. D. Wilson, J. Jacobs, and T. Höllerer. Reality distortion room: A study of user locomotion responses to spatial augmented reality effects. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 1201–1210, 2023. doi: 10.1109/ISMAR59233.2023.00137 3

[33] R. Komiyama, T. Miyaki, and J. Rekimoto. Jackin space: designing a seamless transition between first and third person view for effective telepresence collaborations. In *Proceedings of the 8th Augmented Human International Conference*, pp. 1–9, 2017. 2

[34] N. H. Lehment, D. Merget, and G. Rigoll. Creating automatically aligned consensus realities for ar videoconferencing. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 201–206, 2014. doi: 10.1109/ISMAR.2014.6948428 2, 3

[35] P. Mohr, S. Mori, T. Langlotz, B. H. Thomas, D. Schmalstieg, and D. Kalkofen. Mixed reality light fields for interactive remote assistance. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2020. 2

[36] S. Ohl. Tele-immersion concepts. *IEEE Transactions on Visualization and Computer Graphics*, 24(10):2827–2842, 2018. doi: 10.1109/TVCG.2017.2767590 1

[37] K. Pearson. X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of*

Science, 50(302):157–175, 1900. 7

[38] T. Piumsomboon, A. Dey, B. Ens, G. Lee, and M. Billinghurst. The effects of sharing awareness cues in collaborative mixed reality. *Frontiers in Robotics and AI*, 6:5, 2019. 2

[39] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 652–660, 2017. 3

[40] C. R. Qi, L. Yi, H. Su, and L. J. Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017. 3

[41] S. Razzaque. *Redirected walking*. The University of North Carolina at Chapel Hill, 2005. 2

[42] 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. 2

[43] D. Roth, K. Yu, F. Pankratz, G. Gorbachev, A. Keller, M. Lazarovici, D. Wilhelm, S. Weidert, N. Navab, and U. Eck. Real-time mixed reality teleconsultation for intensive care units in pandemic situations. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 693–694. IEEE, 2021. 2

[44] 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. 2

[45] T. Schubert, F. Friedmann, and H. Regenbrecht. The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, 10(3):266–281, 2001. 6

[46] D. J. Schuirmann. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of pharmacokinetics and biopharmaceutics*, 15:657–680, 1987. 7

[47] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4):591–611, 1965. 7

[48] J.-E. Shin, B. Yoon, D. Kim, H.-I. Kim, and W. Woo. The effects of device and spatial layout on social presence during a dynamic remote collaboration task in mixed reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 394–403, 2022. doi: 10.1109/ISMAR55827.2022.00055 1

[49] R. Strak, K. Yu, F. Pankratz, M. Lazarovici, B. Sandmeyer, J. Reichling, S. Weidert, C. Kraetsch, B. Roegele, N. Navab, et al. Comparison between video-mediated and asymmetric 3d teleconsultation during a preclinical scenario. In *Proceedings of Mensch und Computer 2021*, pp. 227–235. 2021. 2

[50] T. Teo, G. A. Lee, M. Billinghurst, and M. Adcock. 360drops: Mixed reality remote collaboration using 360 panoramas within the 3d scene. In *SIGGRAPH Asia 2019 Emerging Technologies*, pp. 1–2. 2019. 2

[51] H. Thomas, C. R. Qi, J.-E. Deschaud, B. Marcotegui, F. Goulette, and L. J. Guibas. Kpconv: Flexible and deformable convolution for point clouds. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6411–6420, 2019. 3

[52] 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. 2

[53] X. Wang, H. Ye, C. Sandor, W. Zhang, and H. Fu. Predict-and-drive: Avatar motion adaption in room-scale augmented reality telepresence with heterogeneous spaces. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3705–3714, 2022. doi: 10.1109/TVCG.2022.3203109 2

[54] N. L. Williams, A. Bera, and D. Manocha. Redirected walking in static and dynamic scenes using visibility polygons. *IEEE transactions on visualization and computer graphics*, 27(11):4267–4277, 2021. 2

[55] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 143–146, 2011. 7

[56] E. Wong, J. E. S. Grønbæk, and E. Velloso. The jamais vu effect: Understanding the fragile illusion of co-presence in mixed reality. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*, pp. 2227–2246, 2024. 2

[57] World Medical Association. WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects. https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/, 2022. [Accessed: 8 May 2024]. 7

[58] D. Yang, J. Kang, T. Kim, and S.-H. Lee. Visual guidance for user placement in avatar-mediated telepresence between dissimilar spaces. *IEEE Transactions on Visualization and Computer Graphics*, 2024. 2

[59] C. Yeshwanth, Y.-C. Liu, M. Nießner, and A. Dai. Scannet++: A high-fidelity dataset of 3d indoor scenes. In *Proceedings of the International Conference on Computer Vision (ICCV)*, 2023. 4

[60] B. Yoon, J.-e. Shin, H.-i. Kim, S. Y. Oh, D. Kim, and W. Woo. Effects of avatar transparency on social presence in task-centric mixed reality remote collaboration. *IEEE Transactions on Visualization and Computer Graphics*, 2023. 1

[61] L. Yoon, D. Yang, C. Chung, and S.-H. Lee. A full body avatar-based telepresence system for dissimilar spaces. *arXiv preprint arXiv:2103.04380*, 2021. 2

[62] L. Yoon, D. Yang, J. Kim, C. Chung, and S.-H. Lee. Placement retargeting of virtual avatars to dissimilar indoor environments. *IEEE Transactions on Visualization and Computer Graphics*, 28(3):1619–1633, mar 2022. doi: 10.1109/tvcg.2020.3018458 2

[63] K. Yu, G. Gorbachev, U. Eck, F. Pankratz, N. Navab, and D. Roth. Avatars for teleconsultation: Effects of avatar embodiment techniques on user perception in 3d asymmetric telepresence. *IEEE Transactions on Visualization and Computer Graphics*, 27(11):4129–4139, 2021. 2

[64] K. Yu, D. Roth, R. Strak, F. Pankratz, J. Reichling, C. Kraetsch, S. Weidert, M. Lazarovici, N. Navab, and U. Eck. Mixed reality 3d teleconsultation for emergency decompressive craniotomy: An evaluation with medical residents. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 662–671. IEEE, 2023. 2

[65] 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. IEEE, 2023. 2

[66] F. Zaman, C. Anslow, and T. J. Rhee. Vicarious: Context-aware viewpoints selection for mixed reality collaboration. In *Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–11, 2023. 2