# Neighborhood: Dynamic Spatial Alignment for Traversable Telepresence Across Remote Spaces


Taehei Kim
KAIST
Daejeon, South Korea
hayleyy321@kaist.ac.kr

Hyeshim Kim
KAIST
Daejeon, South Korea
hyedeep@kaist.ac.kr

Jihun Shin
KAIST
Daejeon, South Korea
jihun.shin@kaist.ac.kr

Hyeonjin Kim
KAIST
Daejeon, South Korea
hyeonjin@kaist.ac.kr

Sung-Hee Lee
KAIST
Daejeon, South Korea
sunghee.lee@kaist.ac.kr


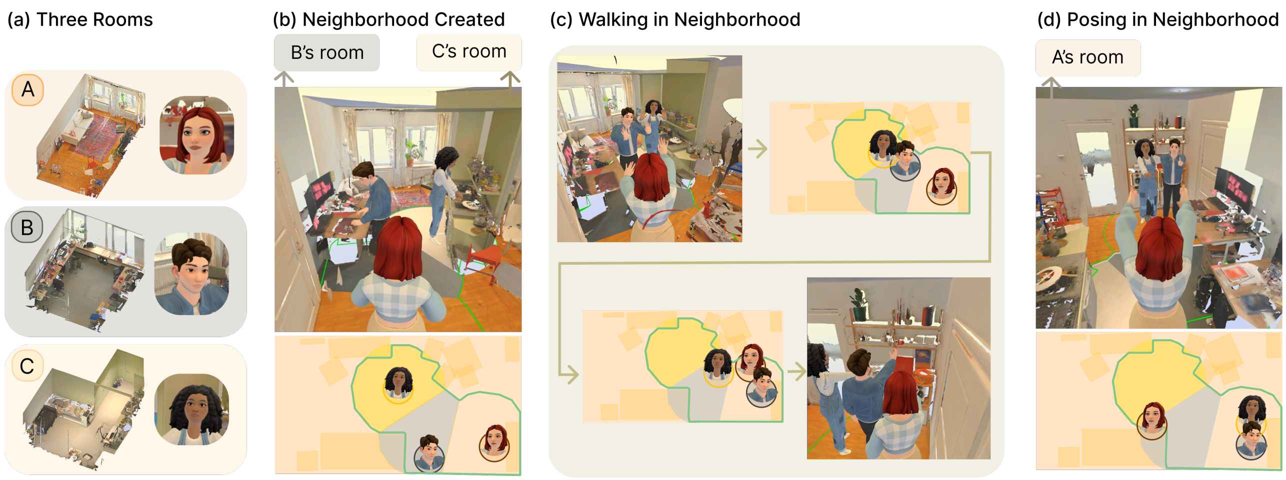


**Figure 1: Our system, Neighborhood, allows users in physically remote spaces to traverse into each other's spaces as if they are adjoined together. (a) Three users' rooms and avatars. (b) The upper row shows the initial view when Neighborhood is created. Users B and C are performing activity within their spaces while user A is observing. The lower row shows a simplified top-down view of three users' positions and room visibility. (c) User A's egocentric third-person perspective and top view of observing user B and C physically walking from B's room and C's room. Arrows indicate the traversal sequence. (d) The upper row shows the resulting view of users B and C posing within A's room. Green lines indicate the safe traversable boundary within each space which in this case, A's. The lower row shows a top-down view.**


## Abstract

We propose *Neighborhood*, a mixed reality (MR) telepresence system that allows users to experience an illusion of walking between local and remote spaces as if they are physically connected. Prior studies transmit either the avatar or the space alone to the remote space. Works that transmit both limit either remote space's visibility or traversability to specific areas. Our work builds upon prior works by transmitting both the avatar and the space, while also supporting full visibility and traversability of local and remote spaces with user's activities grounded with their real furniture and objects, distinct from VR traversal. Our system begins with users moving to the specific area they wish to connect. Using the chosen spaces, the optimization function aligns them to initiate Walking In Mode, visualizing a traversable area with remote spaces augmented. If users wish to move to a different area, the system transits to Panning Mode. Here, the moving user's avatar appears stationary while the background shifts from the observing user's perspective. We quantitatively measure metrics including the size of the neighborhood and ablate each term to evaluate the optimization function across diverse spatial configurations. We also conducted an exploratory user study with 21 participants, showing that traversing within Neighborhood evokes a compelling sense of real spatial adjoin.




*UIST '26, Detroit, MI, USA*



ACM ISBN 979-8-4007-2856-3/2026/11
https://doi.org/10.1145/3830398.3830524


## CCS Concepts

• **Computing methodologies → Mixed / augmented reality**.

## Keywords

Telepresence, Space Optimization, Mixed Realities

**ACM Reference Format:**

Taehei Kim, Hyeshim Kim, Jihun Shin, Hyeonjin Kim, and Sung-Hee Lee. 2026. Neighborhood: Dynamic Spatial Alignment for Traversable Telepresence Across Remote Spaces. In *The 39th Annual ACM Symposium on User Interface Software and Technology (UIST '26), November 02–05, 2026, Detroit, MI, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3830398.3830524

## 1 Introduction

The motivation for this work is to connect physically distant people into one space as if we are living together although physically far away. This desire is not new as researchers have explored ways to enable telepresence, a feeling of co-existence despite distance in several directions. We can see a history of efforts focusing on a perspective of what is transmitted from the local space to the remote space to enable interaction. The first line of work transmits the remote avatar into the local user's space [34]. With more complex space structures taken into consideration, subsequent work addressed spatial differences through space alignment, furniture rearrangement, and motion adaptation [8, 13, 14] while transmitting the avatar.

The second line of work constructs shared virtual environments from users' physical component [28, 40], enabling free traversal across virtual spaces at the cost of losing the users' physical context. The third line, closest to ours, transmits both the remote user's avatar and their space into the local space. Early telepresence work united distant workspaces through a display: this "through-the-window" paradigm allows remote collaborators to see each other and their environments as if attached to or extended from their own space [7, 36]. About a decade later, Maimone et al. [24, 26] captured dynamic 3D scenes in real time and displayed them on wall-sized and autostereoscopic displays, and HMD-based systems like Holoportation [30] began transmitting reconstructed space together with the avatar. Others transmit scanned static scenes: a single user's space can be made accessible to multiple remote users [15, 41], and MRUnion [32] and Voronoi Rooms [19] further provide bidirectional space visibility across all participants. Despite all these advances, telepresence system with full usage of all participating spaces has some limitations. In that, prior systems fall into one of three limitations: (1) only the avatar is transmitted, (2) the remote space is only partially visible, or (3) traversability of the remote space is only possible in VR or is otherwise limited.

To this end, we introduce our mixed reality telepresence system *Neighborhood*. Our method augments part of the real scanned remote space into the local space to create a traversable area. Within this area, users can freely cross over between local and augmented remote spaces, experiencing the illusion of stepping into each other's spaces as if they are connected. Since each user is grounded in their physical environment, activities performed with furniture and objects have real spatial grounding. Through a continuous simultaneous crossover experience within Neighborhood, we aim to demonstrate our interpretation of telepresence space as a shared dwelling.

To construct the neighborhood, our system begins with users moving to a specific area they wish to connect. Using the chosen spaces, the optimization function in Section 4 aligns them to initiate Walking In mode, creating a traversable area. When a user wants to change their individual area, they change to Panning mode. Here, the moving user's avatar appears to walk in place while the background environment shifts from the observing user's perspective. This method helps users to preserve their positional topology despite moving around. We evaluate our method both quantitatively and qualitatively. For qualitative evaluation, we conducted an exploratory user study to gather participants' experience of physical traversal into remote spaces through Neighborhood (Section 5). For quantitative evaluation, we show how our algorithm performs in diverse space sets and ablate each term in the optimization function (Section 6).

As such, our research contributions are as follows:

- We introduce the concept of "walking in" for MR telepresence, where every user experiences the illusion of physically traversing into one another's spaces, evoking a compelling sense of physical crossover, supported by our exploratory user study.
- We propose a space optimization framework that constructs traversable area, Neighborhood, for multi-user telepresence environment. We quantitatively measure and ablate the performance of our optimization terms.
- We develop a fully workable real-time MR system that allows users to dynamically choose and share every part of each space to invite other users.

## 2 Related Work

We organize prior work by what is transmitted from the remote space to enable telepresence interaction. Ohl [29] and Lawrence et al. [21] structure and survey the broader telepresence landscape, from which we draw part of our categorization.

### 2.1 Avatar Transmission

The first line of work transmits remote avatars into the local space; only the avatar is visualized, though its placement may be affected by spatial differences. Room2Room [34] captured remote users and projected life-sized avatars into the local space via projected AR. Spatial differences were then considered more carefully to secure a larger shared space for avatar access. For example, one strategy automatically aligns remote spaces to maximize shared features [22] or optimizes the space through furniture rearrangement [14]. When global alignment is impractical, partial alignment identifies task-relevant sub-areas such as work surfaces [8, 39], or leverages spatial and furniture relationships to position avatars appropriately [12, 17]. Once the avatar is situated, motion adaptation preserves interaction meaning by adjusting avatar gaze, gestures, and body movements to fit the remote environment [3, 4, 13, 18, 35].

Recent systems move beyond static placement toward continuous mapping over the entire space, including neural network-based placement based on interpersonal relations [51], predictive controllers that navigate avatars [47], visual guidance toward optimal positions for gaze and pointing [49], and reinforcement learning for semantically coherent avatar locomotion [23]. Wong et al. [48]

**Table 1: Comparison of multi-user MR telepresence systems. Column Avatar or Avatar+Space means whether the system transmits only the remote user's avatar (1) or also transmits the space (2). Column User Control Sharing means whether space-sharing regions are determined by explicit user designation, implicit system-defined, or are not user-controllable**

| System | Avatar or Avatar+Space | Dynamic or Static Scene (Remote Space) | Scene Style | Bidirectional Space Visibility | Physically Grounded Traversability | Traversability to Other Space | User Control Sharing |
|---|---|---|---|---|---|---|---|
| Lehment et al. [22] | 1 | – | – | – | O | – | – |
| Holoportation [30] | 2 | Dynamic | Reconstructed | Part | O | O | – |
| Loki [43] | 2 | Dynamic | Reconstructed | Part | O | O | – |
| Slice of Light [45] | 2 | Static | Synthetic | Full | X | O | Explicit |
| XSpace [9] | 2 | Static | Simplified Proxy | Full | X | O | Implicit |
| SpaceBlender [28] | 2 | Static | Synthetic* | Part | X | O | Explicit |
| Voronoi Rooms [19] | 2 | Static | Reconstructed | Full | O | X | Implicit |
| **Neighborhood** | **2** | **Static** | **Reconstructed** | **Full** | **O** | **O** | **Explicit** |

1: Avatar only; 2: Avatar and space.
*SpaceBlender incorporates image-based submeshes of user surroundings within a generated environment.

proposed a unified framework bounding the divergent vocabulary used across dissimilar-space telepresence research. Similar to works introduced here, we aim to transmit each user's avatar. We additionally transmit the space surrounding the avatar to provide spatial context.

## 2.2 Virtual Space

Unlike social VR platforms such as VRChat [44] and Rec Room [37], the systems in this section account for users' physical surroundings when constructing a shared virtual space, or visualize multiple virtual environments simultaneously so that users can switch between spaces, or merge multiple spaces. However, the resulting environment is a novel virtual space with no physical referent. Congdon et al. [5] created mappings between VR users' physical spaces to maximize the usable area, while Sra et al. [40] leveraged asymmetries in space size to generate shared virtual environments tailored to each user's surroundings. Slice of Light [45] lets co-located HMD users view and "step" into other users' independent virtual environments, and OVRlap [38] overlays multiple first-person viewpoints in a single view, letting a user monitor several locations and switch between them. Recent work employs generative AI to blend physical surroundings into VR telepresence [28] or fuses environments using Bézier curves [46]. Despite leveraging users' physical layouts, the resulting environment is fully virtual, and users can no longer share their physical context.

## 2.3 Avatar and Space Transmission

The line of work closest to ours visualizes both the remote user's avatar and their residing space in the local space.

*Toward real-time scene transmission.* The earliest visions of spatial telepresence were grounded not in head-worn devices, but in projecting one environment onto another: Raskar et al. [36] envisioned projecting one space onto another through spatially immersive displays, and Gibbs et al. [7] employed wall-sized displays as viewports into virtual extensions of remote participants' spaces. Real-time capture of the actual remote scene was later realized: Maimone et al. [25, 26] captured dynamic 3D scenes in real time and displayed them on head-tracked stereoscopic displays, and Holoportation [30] reconstructed 3D spaces for real-time HMD viewing. Loki [43] integrated video, spatial capture, and mixed-reality presentation for exploring and annotating environments, Project Starline [20] demonstrated realistic face-to-face conversation through an autostereoscopic display without head-worn devices, and volumetric capture now enables AR and VR co-presence [10, 11]. However, real-time reconstruction covers only part of the remote space, leaving the rest inaccessible to the other users.

*Pre-scanned scenes shared unidirectionally.* Other systems use scenes reconstructed offline; the visualized scene is static, yet the avatar transmits the user's movement in real time. Steed et al. [41] reconstructed a local physical space as a virtual model, allowing remote participants to visit and traverse it via immersive VR, and Oyekoya et al. [31] extended this vision by teleporting a visitor captured and rendered in real time across heterogeneous modalities. Teo et al. [42] bridged the two approaches by embedding a live 360 panorama within a pre-scanned 3D reconstruction of the local worker's environment. Kim et al. [16] proposed relative translation gains that adjust a VR client's walking speed per axis to generate a mutually traversable space within the AR host's environment, and multiple remote users can be allocated interactable subspaces within a shared host space by separately considering perceivable and interactable areas [15]. These systems focus on exposing one local space and distribute to multiple remote users.

*Bidirectional space sharing.* A final group exposes the scanned physical spaces of all participating users. XSpace [9] introduced a developer toolkit and authoring tools for constructing shared virtual spaces in distributed AR collaboration, Pabst et al. [32] embedded virtually simplified representations of smaller spaces within larger ones to preserve spatial coherence across heterogeneous environments, and Voronoi Rooms [19] employed scanned reconstructions of each user's physical space, modulating the visibility of overlapping spaces based on user positions. Yet these systems do not allow full traversal of the remote physical space. Table 1 summarizes these differences with additional axes such as scene style (reconstructed from scans/images, synthetically made, or a simplified proxy) and user control sharing (explicitly user-configured or implicitly system-determined). Building on this, our work transmits

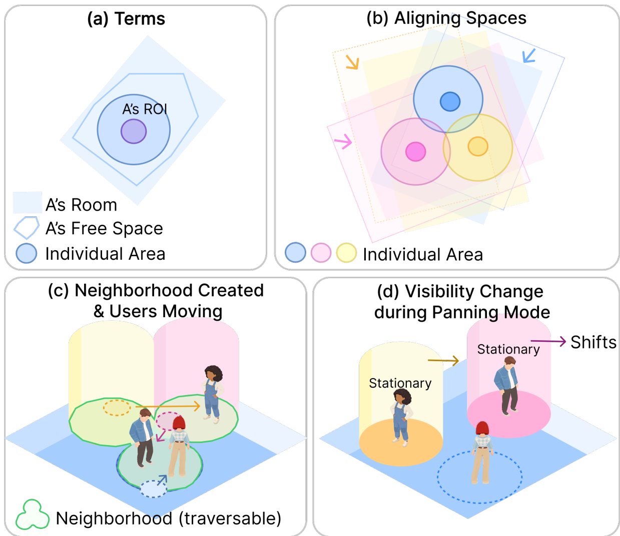


**Figure 2: (a) Terms used within our system. (b) Aligning spaces to create Neighborhood through optimization. Spaces represented in 2D floorplan are translated and rotated to satisfy optimal spatial alignment. (c) Neighborhood Created. Neighborhood is a connected traversable area where remote users' spaces augmented and connected to the local space. Green line represents Neighborhood boundary. Arrows indicate the position change of each user. (d) Visibility during Panning mode from the observer's perspective. When remote users' move, their avatars remain stationary while their background shifts.**

both the avatar and the surrounding space, and further lets users simultaneously cross over into each other's entire physical space.

## 3 Neighborhood Overview

Our system allows users to experience the illusion of visiting and walking into remote spaces as a visitor, while remote users can likewise enter user's own local space. Furthermore, all participants can share every part of their real space. Within our system, there are three spatial regions as depicted in Figure 2 (a). The *Individual Area* represents a specific area including furniture that includes the full context of user actions and spatial surroundings. Remote users can see into these regions as if through a window. Ideally, all individual area may be fully augmented into the remote space. The *Region of Interest (ROI)* can be any area within a user's individual area that they want to share with or be accessed to others and represents the subset of each individual space that must be prioritized when augmented to other users' spaces. It is not restricted to walkable area but encompasses furniture within its circle, capturing the semantic context of each user's immediate environment. The *Neighborhood* is union of connected Individual Areas, creating a traversable zone where users can walk into each other's spaces, not merely observe them. Given a set of physical spaces as 2D floor plans with spatial regions labeled through user placement, we optimize space alignment to create Neighborhood by analyzing individual areas, reflecting the ROI, and maximizing traversable space while minimizing overlap (Figure 2 (b), Section 4). Once the Neighborhood is calculated, the Walking In mode starts where the augmented remote spaces are fixed, allowing users to traverse between spaces as in Figure 2 (c). When users want to relocate to a new area, the system shifts to Panning mode as in Figure 2 (d). Here, the observer's view and the mover's view differ. For the mover, all augmented remote spaces and avatars translate together as if following the user while preserving the topology. From the observer's view, when remote users move, their avatars remain stationary while their background shifts.

*Size of the Individual Area.* We set the initial radius of the individual area to 1.2 meters. This choice is based on principles of workspace enclosure and interpersonal spatial comfort. Alexander et al.'s empirical study of workspace design found that people regard 2.4 meters as an ideal distance between two people when working in the same space [1]. We therefore define a personal space radius of 1.2 meters per user, such that two users standing at their respective centers maintain a 2.4-meter separation. This is enforced only at initialization; once the Neighborhood is created, users may freely enter adjacent zones. We define the Region of Interest as a circle with a 50-centimeter radius, corresponding to the minimum space required to avoid body contact [33].

## 4 Neighborhood Optimization

This section details our optimization function compositions to create Neighborhood. The input to our system is multiple spaces represented as a 2D floor plan. The complete objective function consists of four terms:

$$f = f_R + f_C + W(f_M + f_A) \qquad (1)$$

ROI accessibility cost $f_R$ ensures that ROIs are accessible by all users, and connectivity cost $f_C$ prevents Neighborhood from being fragmented which may result in untraversable space. Once these primary objectives are satisfied, the optimization puts higher weight $W$ toward maximizing walkable area $f_M$, and refining individual area placement $f_A$. $W$ is determined by the ROI metrics; it is assigned a small value when ROI accessibility is not preserved, and grows linearly as the ROI accessibility improves.

### 4.1 Space Identification

Given a 2D floorplan, we begin by identifying free spaces in each space, regions where users can move without colliding with physical obstacles. For each space $i$, we compute the free space $F_i$ by subtracting all furniture and objects from the total space boundary $S_i$. Within each space, the optimizer marks positions and radii of individual areas $\{I_1, I_2, \ldots, I_n\}$ and ROI $\{R_1, R_2, \ldots, R_n\}$ which are used to calculate costs.

### 4.2 Region Of Interest (ROI)

The ROI is the user's immediate surroundings or specific chosen areas prioritized for augmentation within other users' free spaces. ROI cost controls the influence of other terms to guarantee accessibility to each user's ROI while creating Neighborhood. While our framework naturally extends to multiple ROIs per user, we assume at most one ROI per user in this work. In our experiment, each

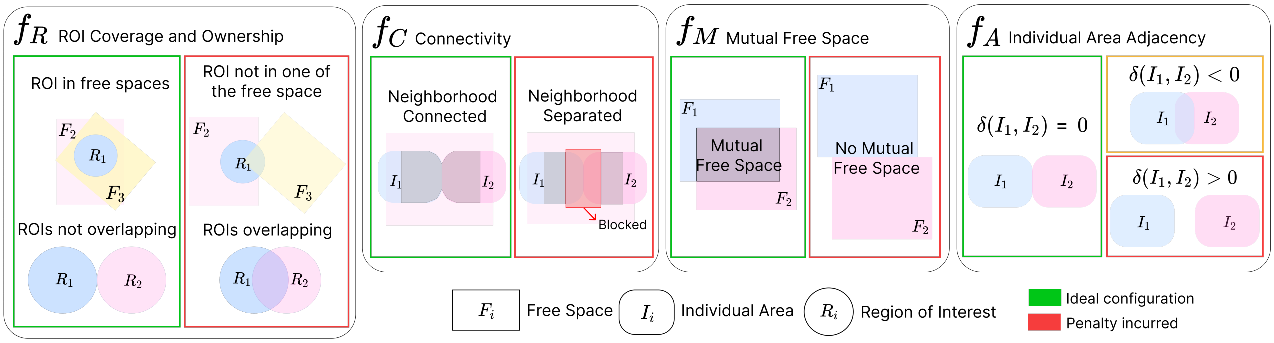


**Figure 3: Role of each term in the optimization function. $f_R$ ensures that all regions of interest (ROI) to cover a minimum amount of free spaces. It also penalizes ROIs from overlapping excessively. Connectivity $f_C$ penalizes Neighborhood from becoming disconnected. The mutual free space $f_M$ works to maximize the free walkable space within Neighborhood. The individual area adjacency $f_A$ encourages $I_1$ and $I_2$ to remain non-overlapping yet spatially adjacent, penalizing configurations where one area intrudes the other.**

user's ROI $R_i$ is represented as a circular area with a 50-centimeter radius around their position.

*Coverage.* ROI Coverage measures how well each ROI is covered by other spaces' free spaces:

$$C_{ij} = \frac{|R_i \cap F_j|}{|R_i|} \quad (i \neq j), \tag{2}$$

where $|R_i|$, the area of user $i$'s ROI, normalizes the coverage. We compute the minimum ROI coverage of user $i$ over all remote spaces, taking the worst-case coverage across all: $C_i^* = \min_{j \neq i} C_{ij}$. Using $C_i^*$, we calculate two metrics: mean coverage $\bar{C}$, and minimum coverage $C_*$.

*Ownership.* ROI ownership prevents ROIs from overlapping each other to secure individual accessibility. When ROIs overlap, each point in the overlapping region is allocated proportionally. We define $n(p)$ as the number of ROIs that contain point $p$ (including user $i$'s own ROI). The ownership for user $i$ is then:

$$O_i = \frac{1}{|R_i|} \int_{R_i} \frac{1}{n(p)}\, dp. \tag{3}$$

Intuitively, a point shared by $n$ ROIs contributes $\frac{1}{n}$ to each user's ownership. Using $O_i$, we get mean ownership $\bar{O}$, and minimum ownership $O_*$.

*ROI Cost.* The total ROI cost is defined to secure sufficient ROI coverage and ownership:

$$f_R = w_R \mathcal{P}_m(\bar{C}) + w_R \mathcal{P}_m(\bar{O}) + \mathcal{P}_c(C_*) + \mathcal{P}_c(O_*) \tag{4}$$

A softplus-based penalty $\mathcal{P}_m$ is designed to approach zero when $\bar{C} \rightarrow 1$ and increases significantly when $\bar{C}$ falls below 0.8. $\bar{O}$ is processed in the same manner. In addition, a linear penalty $\mathcal{P}_c$ is applied to prevent $C_*$ and $O_*$ from falling below $\tau_c = 0.5$.

## 4.3 Connectivity

Connectivity ensures traversable passages within Neighborhood by preventing it from being fragmented. In terms of implementation, we include a morphological connectivity check. We first expand each individual area $I_i$ with a small buffer (5 cm) to account for near-touching regions. We then calculate the union of all buffered zones $I^u$. For each free space $F_i$, we compute its intersection with $I^u$. Disconnection is detected if this intersection results in divided polygon components.

*Connectivity Cost.* As such, the connectivity cost is formulated as follows:

$$f_C = w_C \sum_{i=1}^{n} \chi_i \tag{5}$$

where $\chi_i = 1$ when $F_i \cap I^u$ becomes disconnected and $\chi_i = 0$ otherwise. We assign a high weight $w_C = 100$ to ensure connectivity.

## 4.4 Mutual Free Space

The third term aims to maximize walkable space within Neighborhood, and is weighted more heavily once ROI and connectivity objectives are met. For each space $i$, we compute how much of its free space intersects other users' individual areas. Then we average across all other spaces:

$$M_i = \sqrt{\frac{1}{n-1} \sum_{j \neq i} \frac{|F_i \cap I_j|}{|I_j|}}. \tag{6}$$

*Mutual Free Space Cost.* We want users to have relatively equal amounts of free space to prevent the cases where one user's free space is neglected. We measure imbalance using the squared coefficient of variation $\gamma^2$ of $M_i$ with a penalty weight $\lambda = 2.0$. As such, the cost function is designed as follows:

$$f_M = -w_M \left( \sum_{i=1}^{n} M_i - \lambda \cdot \gamma^2 \right). \tag{7}$$

## 4.5 Individual Area Adjacency

Individual areas should be placed close enough to enable walking traversal between users, but not overlap excessively, which would reduce the size of Neighborhood. To calculate positional relationship between individual areas, we adopt a signed distance function. For each individual area pairs, we define a signed measure $\delta(P, Q)$ that identifies both separation and penetration. When the areas do not overlap, $\delta$ equals the minimum distance between their boundaries. When areas overlap, $\delta$ becomes negative and scales with its magnitude proportional to the degree of penetration, penalizing excessive penetration between individual areas. We map signed measure to penalties using a scoring function $s(d)$ centered on a threshold separation $d^\circ$:

$$s(d) = \begin{cases} |d - d^\circ| - \varepsilon, & \text{if } d < d^\circ - \varepsilon \\ 0, & \text{if } |d - d^\circ| \le \varepsilon \\ (|d - d^\circ| - \varepsilon)^{1.5} & \text{otherwise,} \end{cases} \tag{8}$$

where $\varepsilon$ defines an acceptable tolerance around the ideal distance. The exponents differ for the two directions encouraging the optimizer to prefer overlap than separation.

*Individual Area Adjacency Cost.* The overall adjacency cost averages scores across all valid area pairs:

$$f_A = \frac{w_A}{|\mathcal{P}|} \sum_{(i,j)\in\mathcal{P}} s(\delta(I_i, I_j)) \tag{9}$$

where $\mathcal{P}$ is the set of evaluated pairs. For configurations with more than four spaces, we limit adjacency evaluation to each area's $k$ nearest neighbors (determined by centroid distances) rather than computing all $O(n^2)$ pairs to reduce computational cost while preserving the essential constraint.

## 4.6 Optimization Implementation

We fix the first space at the origin and optimize only the relative positions of remaining spaces. Each moving space is parameterized by three variables: translation in $x$ and $y$ and rotation. For $n$ spaces, this yields a $(n-1) \times 3$ dimensional search space. For the optimization algorithm, we adopted Differential Evolution algorithm. The population size is set to $15 \times d$ where $d$ is the number of dimensions. We run optimization for a maximum of 80 generations with convergence tolerances of $10^{-6}$ for both function value and parameter changes. Term weights are set as $w_R = 10$, $w_C = 100$, $w_M = 1.0$, and $w_A = 1.5$. The secondary weight $W$ grows linearly from 0.5 to 2.0 as ROI coverage approaches to the target threshold $\tau = 0.85$.

# 5 Neighborhood System

We develop a mixed reality system to demonstrate our concept of traversable telepresence. Our system fully supports MR in the sense that remote avatar placement and spatial transformations are computed relative to each user's physical position, using the local space as the coordinate reference frame. Ideally, the local space would be best experienced if the user is physically located in a fully furnished space and observes it as see-through function of HMD. However, due to spatial constraints, we demonstrate our system using a pre-scanned mesh, meaning the local user also sees their space as virtual. The system accepts any physical pre-scanned mesh and we used scenes from ScanNet++ [50] to balance visual fidelity with spatial limitations. We processed these meshes; furniture and wall bounding boxes are manually placed in Unity and exported as JSON, from which object footprints and wall boundaries are projected onto the XY plane to construct a 2D floor plan.

## 5.1 System Modes

Our system consists of three phases: Initialize, Walking In mode, and Panning mode. Once the system turns on, each user moves to their preferred location within their physical space for initialization (Figure 4 (a)). It depicts a scene where user A stands in front of the sofa, user B standing next to the desk, and user C in front of the kitchen. Using the individual areas users chose during initialization that are marked as green circles, the system calculates optimal alignment to create Neighborhood. Once calculated, the system changes to Walking In mode (Figure 4 (b)). In this mode, Neighborhood is constructed and fixed so that users can observe and walk into the other remote spaces. The green outline is fixed and delineates the extent of local objects within the overlapping region. Here, Figure 4 (b) shows two scenes where user A standing on B's room, and user A observing user B and C in C's room. When users wish to relocate their individual area, they switch to Panning mode (Figure 4 (c)). When panning mode starts, the fixed remote augmentation is removed and the green outline follows the user's current position, rendering the area corresponding to their individual area size with that user's space. When a user initiates movement, all augmented remote spaces and avatars translate together with them, preserving the topological relationship. The movement vector is applied uniformly to the entire remote environment. For observers watching a moving user relocate, we employ an inverse transformation: instead of the remote avatar moving through the environment, the entire remote environment appears to shift around the stationary avatar. This creates the view where the remote user explores their space while keeping their position fixed in the observer's view. Figure 4 (c) depicts a scene in which User C introduces C's room to Users A and B. Through Panning mode, users can navigate and share all parts of their local physical environment. When users choose new individual areas through Panning mode, they can re-run optimization to create a new Neighborhood. Likewise, users can continuously alternate between Walking In mode and Panning mode while the system runs.

## 5.2 Implementation

Within our system, one user's desktop serves as the server, managing mode transitions and optimization computation. When all users finish moving to their chosen initial locations as individual areas, the server collects their positions and starts computing. Once finished, the server broadcasts the resulting spatial transforms to all clients, transitioning the system into Walking In mode. Walking In mode re-aligns spaces to create Neighborhood, repositioning remote users' spaces and avatars as if they were teleported. These transforms are expressed respective to the server's local frame. As such, each client transforms the result by treating its own space as the origin to obtain the target position and rotation of remote spaces and avatars in its own local frame. This includes reversing the translation and rotation assigned during optimization. When the system transitions to Panning mode, an individual area is marked around each user based on their position at the moment of transition. Each individual area is rendered as the remote space of the other user. During Panning mode, remote avatars should appear stationary while spaces are shifting. To achieve this, each user captures all avatar chest joint positions at the moment of transition within their

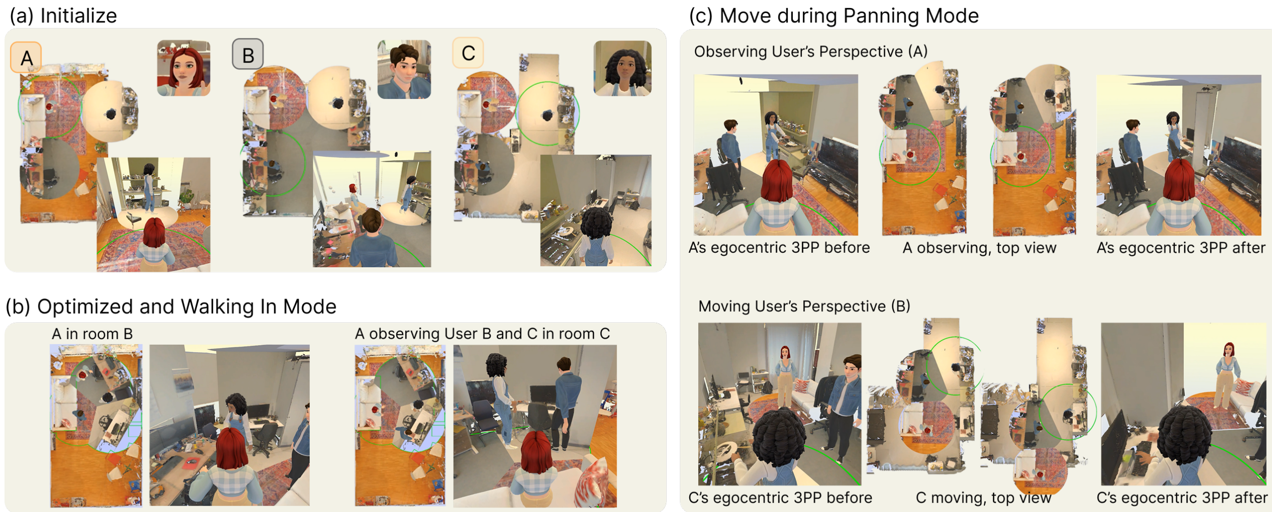


**Figure 4: Overview of our system with three stages. (a) Initialize: Users in three different rooms (A, B, C) navigate to locations they wish to share, defining their individual areas. (b) Walking In Mode: The system computes optimal room alignment, forming a traversable neighborhood. The optimized top-down view (left) shows connected spaces; User A can then walk into B's room or observe other users walking into each other's spaces. All views are presented from A's perspective. (c) Panning Mode: Users can relocate their individual area while maintaining spatial topological relationships. The upper row shows the observing user's experience, which is A: before C's move, during C's move where C's avatar appears stationary while the background shifts, and after C finishes moving to the new position. The lower row shows the moving user's experience, which is C: before relocating, during movement with remote spaces following alongside, and after C finishes moving to the new position.**

local world. The relative offset vectors from the local avatar to each remote avatar are computed and used as invariants throughout the session. As the local user moves through their physical space, the system calculates per-frame positional deltas from the local chest joint and applies these deltas to each remote avatar's root transform, maintaining the fixed relative offsets in the horizontal plane. Here, we applied a filter to the chest position and to each remote avatar's target position to reduce tracking noise. Regarding space shift, remote spaces are continuously re-positioned by inverting the remote avatar's position within their local coordinate frame. This is because remote avatars are moving in their respective spaces, yet from the observing user's perspective, spaces instead of avatars should be shifting.

To render remote environments, each user stores pre-scanned room meshes of all users locally. We then apply a custom shader with zone-based visibility to each mesh. Each room mesh is assigned a zone mode (Local or Remote), and the shader clips fragments based on whether they fall within users' individual areas. Remote zone fragments outside the assigned user's individual area are clipped entirely. Local zone fragments are always rendered unless they fall within overlapped regions: when multiple individual areas overlap, the shader computes distances to all user positions and assigns each fragment to its nearest user via Voronoi tessellation. User positions are passed to the shader in real time for boundary updates.

For system development, we used Unity engine (2022.3.57f1) along with the Meta XR Core SDK. We used avatars from Meta Avatar SDK and implemented multiplayer networking with Photon. For the hardware setup, we used Meta Quest 3. For system control and rendering, we used two laptops and one desktop, each with a minimum CPU of Ryzen Threadripper 3970X and a 2080 Ti graphics card.

### 5.3 Exploratory User Study

We conducted an exploratory user study to examine the experiential qualities of traversable telepresence. The study aimed to gather experiences from two aspects: (1) the experience of physically traversing into remote spaces and (2) the experience of one's own space being shared with remote users.

*5.3.1 Task.* We designed a collaborative scene-creation scenario following the idea of the circumplex model of group tasks [27]. Participants were asked to create three collaborative scenes using the physical affordances of their rooms. Participants engaged in brainstorming to discuss how they would make use of each allocated local space, then physically enacted the scene through bodied poses and actions. Each participant was provided with a brief situational context describing their living circumstances. We asked participants to regard the virtual room as if they were physically residing in it. This generative task encouraged participants to engage with the room's spatial context, such as furniture interaction, as they were encouraged to propose scene ideas based on their available rooms, select which scenes to pursue, collaborate and communicate whose room and which furniture to use, and build scenes together.

*5.3.2 Participants.* Our experiment was reviewed by the Institutional Review Board. We recruited 21 participants (12 male, 9 female; age 22–33, M=26.57, SD=2.61) through campus announcements and obtained written consent forms. We recruited participants who had at least 10 sessions of VR experience, of which seven participants had experience conducting VR-related research. Participants were compensated for their participation.

*5.3.3 Procedure.* Our system operated with three users at a time, situated in different physical spaces. Upon arrival, participants signed a written consent form and listened to experiment details including the procedure, time taken, and task rules. The experiment began with participants entering their respective physical spaces. Participants walked around their given virtual spaces for five minutes to become accustomed to the room structure and furniture. Participants walked throughout any area that appeared as walkable space within the virtual environment just as if they do in physical space. The physical lab space (floor area) was matched 1:1 in scale to the virtual room. Once ready, participants started the practice session, which lasted for 15 minutes. The practice session was designed for participants to familiarize themselves with the system before performing the main task. During this session, participants explored and learned how to use the two modes until they felt comfortable with the system. When participants reported that they understood how to use the system, the task session started. Here, participants collaborated to construct three consecutive scenes for 15 minutes. When they finished, the interview session began with a starting question asking for their general impression of the system. Including the 20-minute interview, the total experiment took around 75 minutes.

## 5.4 Results and Discussion

We recognize that interview data may not reflect universal truths, but rather the situated realities of individual participants, mediated by their subjective experiences. Nonetheless, through cross-group thematic analysis, we aim to identify patterns of experience that recurred with sufficient consistency to provide meaningful insight into our system. We performed reflexive thematic analysis following Braun and Clarke's six-phase process for theme development [2]. Through this analysis, we identified five themes, each evidenced across five or more of the seven groups.

*5.4.1 Spatial Traversal and Object Reference.* Participants frequently reported that the system evoked a genuine sense of being physically present within each other's space: "Because the spaces came together and we could go into them ... we started gesturing like we'd actually met in person." (G5-P2), "Because we could actually walk around together." (G6-P1), "actually being together ... watching someone make coffee right beside me" (G5-P3), "I'm here cooking, they're there reading ... yet it feels like we're all somehow together in my room." (G7-P2). One participant summarized the entire experience as "virtual residential intrusion" (G6-P2). This sense of adjoin appeared to intensify at moments of object-level contact. Participants also reported that co-presence increased when another person's hand reached toward or engaged with physical furniture and objects in the local space: "Watching someone come into my space and interact with my objects ... that's when it felt real" (G4-P3), and "seeing my hand and his object in the same space ... I really felt like we were together" (G4-P2). Along these lines, many participants requested object referencing functions such as object highlighting (G1-P1), moving virtual objects (G3-P2, G2-P1), or placing new shared virtual objects (G5-P2), suggesting that object-level interaction capabilities would further enrich the system experience. Additionally, seven participants expressed a desire for higher-quality scene meshes, as they believed mesh quality affected their experience, particularly when referring to objects with visible artifacts.

*5.4.2 Spatial Mixture and Clarity.* Participants across all seven groups described the spatial mixture itself as a novel and engaging experience. Reactions such as "remarkable" (G1-P1, G2-P1, G2-P2, G3-P1, G3-P3, G5-P2, G6-P3, G7-P1) and "fun" (G1-P3, G2-P2, G2-P3, G4-P1, G4-P2, G6-P1, G7-P2, G7-P3) emerged often, suggesting that the act of watching physically separate spaces merge in real time carries experiential value. However, this novelty coexisted with a need for clear explanation of the merged space: participants simultaneously desired clearer cues about where one space ends and another begins. Proposed solutions ranged from visual boundary indicators and zone dimming to auditory room differentiation. Furthermore, participants often expressed that the merged space felt small (eight out of 21). We believe this occurs due to a fundamental constraint of the system whereby remote spaces can only be augmented within the physically walkable local space. As the number of rooms increases, the intersection of all users' free space shrinks, resulting in the splitting of spaces. Theoretically, the system functions as long as more than 50% of the ROI is augmented into the remote space. However, this minimum threshold may fall short of supporting meaningful collaboration since participants report spaces feel condensed when augmented areas are small. From this insight, we suggest that spaces share an overlapping region of approximately $0.8\text{m}^2 \times$ number of spaces as minimum requirement.

*5.4.3 Using Space to Communicate.* Showing the local space seems to work as an additional communication tool. Participants consistently recalled video call analogies to describe what they had experienced: "like a dynamic video call" (G2-P2), "the VR version of a video call" (G2-P1), and "instead of video calling distant friends" (G1-P2). One participant described an explicit evolutionary path: "text → call → video call → this" (G5-P3). Participants actively used augmented physical environment to convey their movement context. With this experience, participants suggested many use cases such as virtual cafes, remote fashion reviews, and collaborative cooking which all leveraged the physical environment as context.

*5.4.4 Audio Feedback.* Several participants reported a need for spatial audio to increase spatial presence. Requests ranged from ambient differentiation, "if the sound shifted when I crossed into their room, I'd know I'd crossed over" (G1-P2), to precise spatialization, "if the reverberation of their room came through to me, it would feel so much more real" (G6-P2). Others requested more immediate feedback such as footstep sounds during traversal (G3-P1) or interaction sound effects (G6-P1). This suggests that light audio cues may substantially reinforce the spatial boundaries that visual geometry alone fails to convey.

*5.4.5 User Controllable Space Optimization.* Participants acknowledged that the green circle with a 1.2 meter radius would be prioritized to be augmented in remote spaces. However, except for the near proximity, which is the region of interest with a 0.5 meter radius, the area outside the immediate circle has the possibility of being cut due to space limitations. As such, participants expressed a desire for the ability to control the shape and size of the space they want to share. Furthermore, participants reported their desire to assign certain areas to be private. Mentioned examples included drawing custom zones prior to adjoin (G6-P1, G6-P2), adjusting the radius of the individual area on demand (G1-P1), and excluding areas they preferred not to expose (G6-P1). Some also mentioned an explicit object marking function (G5-P1) so that specific objects are guaranteed to be included.

*5.4.6 Panning Mode.* Four participants reported dizziness during Panning mode. But at the same time, many found the experience novel, with several noting diverse use cases such as watching sports broadcasts from a specific vantage point or describing the interaction as a form of a game. Given these examples, the applicability of the Panning can be further investigated.

## 6 Quantitative Metrics and Results

To show and evaluate how our optimization function objectively performs with diverse sets of spaces, we devised several metrics reflecting the quality of the generated zones. For the space dataset, we used Procthor [6]. We used 340 houses with 918 spaces, where each space corresponds to a single room. Regarding quantitative measures, there are two factors that affect the performance. The first is the size of the space. Our framework augments remote spaces in the free space within the local space. As such, the size of the augmented area heavily depends on the space size. If one of the spaces within the combination is small, possible alignments will be restricted. The second is the clutter level of the space. For a similar reason as with the space size, if the space is heavily cluttered, the free space would diminish. As such, we provide four groups which are the combinations of two levels of size and two levels of clutter. For space size, we divided spaces at a $22.65m^2$ threshold. Spaces below this are classified as small (S) and spaces above as large (L). For clutter level, we divided at a 22.82% threshold. Spaces below this are sparse and spaces above are dense. We set the thresholds based on the median value of the aggregated dataset reflecting the distributional center of the data. Details of the combination are provided in Table 2. Regarding the table, Row 1 is the sample size, Row 2 is the space size (mean ± std), and Row 3 is the clutter level (mean ± std). Figure 9 shows the five spaces used in groups 2 and 3 and their alignment results. For the experiment, user positions were sampled at random from the walkable floor area of each room with a 50 cm wall margin and ROIs were designated inside individual areas.

### 6.1 Size of the Neighborhood

Figure 5 presents the distribution of mean neighborhood area across four groups with varying space configurations (2–5 spaces). In general, large spaces with fewer objects create larger Neighborhood (group 3). Also, the size of the neighborhood increases with an increasing number of spaces.

**Table 2: Group statistics by size and clutter level. Each cell reports sample count (n), mean space area ± std ($m^2$), and mean object clutter level (%) ± std.**

| | | Clutter level | |
|---|---|---|---|
| | | Sparse | Dense |
| Size | Small | **Group 1**<br>n=247<br>15.18 ± 4.41<br>15.16 ± 4.52 | **Group 2**<br>n=212<br>14.87 ± 4.85<br>29.72 ± 3.91 |
| | Large | **Group 3**<br>n=212<br>31.99 ± 6.16<br>16.04 ± 5.14 | **Group 4**<br>n=247<br>31.06 ± 8.65<br>29.73 ± 5.66 |

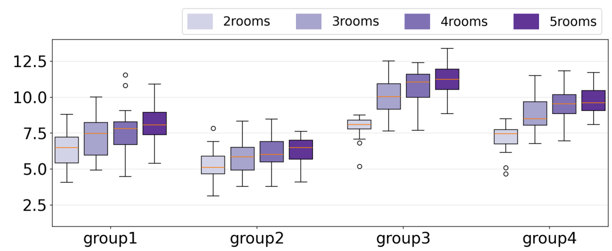


**Figure 5: Size of the Neighborhood ($m^2$)**

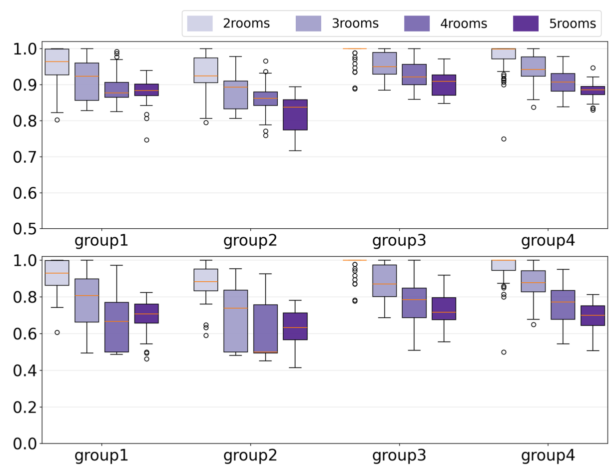


**Figure 6: Region of interest (ROI) Coverage ratio (%) across 2–5 space configurations. Mean coverage (top) stays consistently high, while minimum coverage (bottom) degrades with increasing spatial complexity.**

### 6.2 Region of Interest Accessibility

ROI minimum and mean coverage ratio measure how accessible each user's ROI is from other users' remote free spaces. We set the critical minimum threshold at 50%, as below this point remote users cannot co-exist within the ROI due to small area size to stand on. Figure 6 presents the ROI coverage performance across four groups with varying space configurations (2–5 spaces). The upper

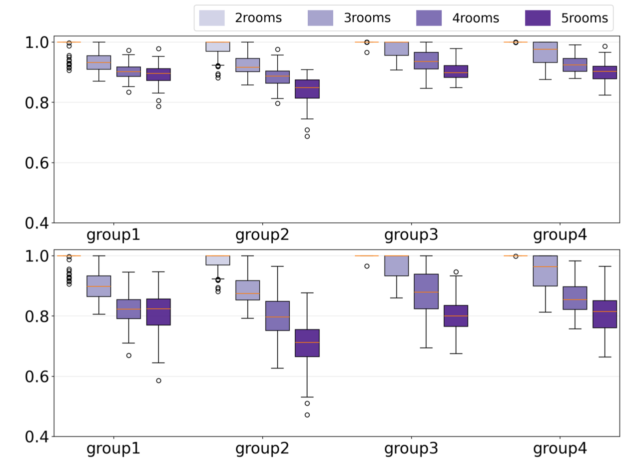


**Figure 7: Region of Interest (ROI) Ownership ratio (%). Mean ownership (top) remains consistently high, while minimum ownership (bottom) degrades with increasing spatial complexity.**

graph shows the mean ROI coverage, while the lower graph shows the minimum ROI coverage, indicating the worst-case accessibility among all spaces. The mean ROI coverage remains consistently high (0.80–1.0) across all conditions. However, the minimum ROI coverage shows a decreasing trend as spatial complexity increases. Group 2 shows the most significant degradation, dropping from approximately 0.85 (2 spaces) to 0.60 (4–5 spaces), suggesting that satisfying ROI constraints for all spaces simultaneously becomes increasingly challenging in small and cluttered spatial configurations. The gap between mean and minimum metrics indicates that while overall ROI accessibility is preserved, individual spaces may experience reduced coverage.

Usability here refers to whether each user's intent can be performed and observed within the neighborhood. If a user's ROI share (ownership) falls below 50%, it means that less than half of their action radius remains exclusively theirs, at which point we consider the user's intent can no longer be reliably conveyed to remote participants. Figure 7 shows the preserved ROI ratio. The upper graph shows the mean value and the lower graph shows the minimum value. The mean ROI ownership stays above 80% across all groups and space configurations, indicating that the optimization term prioritizes preserving the ROI.

## 6.3 Physical Space Overlap

We measure the physical space overlap ratio to calculate how much each Neighborhood is walkable from each user. We show four graphs depending on the space size and the clutter level. Figure 8 illustrates the mean furniture penetration ratio within Neighborhood across four groups, with lower values indicating better walkability. Group 3 demonstrates the lowest penetration ratios (approximately 0.07–0.18), suggesting that larger spaces with lower clutter levels secure minimal furniture intrusion into Neighborhood. In contrast, Group 2 exhibits the highest penetration (0.30–0.39), indicating greater difficulty in avoiding furniture placement within

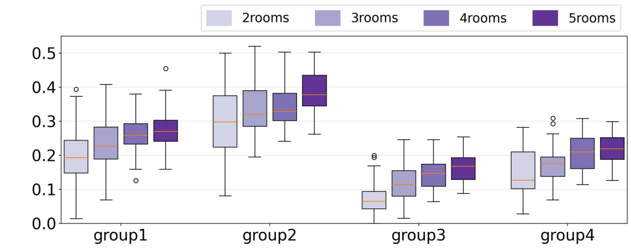


**Figure 8: Physical space overlap ratio (%). Lower values indicate better walkability, with larger and less cluttered spaces achieving lower penetration.**

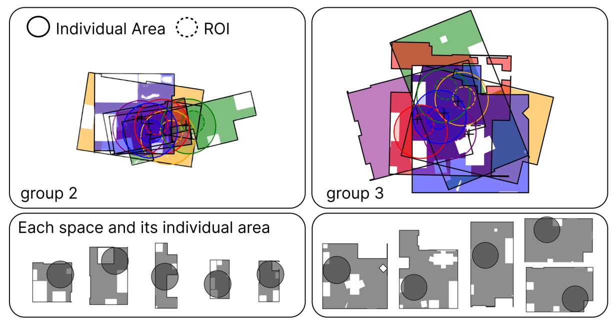


**Figure 9: Visualization of five spaces in groups 2 and 3. The top row depicts five aligned spaces, and the bottom row shows the floor plan of each space.**

Neighborhood. The penetration ratio generally increases slowly as the number of spaces increases.

## 6.4 Ablation

We evaluate the contribution of each objective term through ablation studies across space configurations of varying complexity: two to four spaces ($n$=2–4), with 80 ($N = 80$) sets per space count. Our objective function comprises four terms: ROI coverage and ownership ($f_R$), connectivity ($f_C$), mutual free space maximization ($f_M$), and individual area adjacency ($f_A$). We find that the effect of each term varies with spatial complexity: secondary terms ($f_M$ and $f_A$) have greater impact in configurations with fewer spaces, while primary terms ($f_R$ and $f_C$) become critical as the number of spaces grows, where spatial constraints are tighter. We discuss each term in turn.

**Table 3: Ablation results for $f_M$ and $f_A$ ($n = 2$, $N = 80$). Bold indicates better value (↑ higher is better ↓ lower is better).**

| Term | Metric | Ours | w/o |
|---|---|---|---|
| $f_M$ | Space Overlap (%) ↓ | **0.143 ± 0.086** | 0.180 ± 0.086 |
| $f_A$ | Ideal Pair (%) ↑ | **91.25%** | 6.25% |

Table 3 reports ablation results for the two secondary objective terms in two-space configurations. With a lower number of spaces participating, the primary constraints (ROI coverage and connectivity) are readily satisfied, and the secondary terms serve to incrementally refine the spatial layout. Removing $f_M$ increases mean physical space overlap ratio from 0.143 to 0.180, indicating that without an explicit incentive to maximize traversable space, individual areas are more likely to include non-traversable area. For $f_A$, we calculate the percentage of individual area pairs within ideal distance as a measure of adjacency. The difference between 91.25% and 6.25% suggests that in two-space configurations, the optimizer is less likely to leave individual area adjacent without an explicit penalty.

**Table 4: Ablation results for $f_R$ and $f_C$ across space configurations ($N = 80$ per configuration) in ratios. Bold indicates better value (↑ higher is better, ↓ lower is better).**

| Term | $n$ | Metric | Ours | w/o |
|---|---|---|---|---|
| $f_{\text{cov}}$ | 3 | Mean Coverage ↑ | **0.935 ± 0.051** | 0.676 ± 0.314 |
| | | Min Coverage ↑ | **0.829 ± 0.146** | 0.496 ± 0.388 |
| | 4 | Mean Coverage ↑ | **0.883 ± 0.046** | 0.496 ± 0.308 |
| | | Min Coverage ↑ | **0.698 ± 0.151** | 0.280 ± 0.330 |
| $f_{\text{own}}$ | 3 | Mean Ownership ↑ | **0.954 ± 0.043** | 0.896 ± 0.105 |
| | 4 | Mean Ownership ↑ | **0.900 ± 0.032** | 0.760 ± 0.107 |
| $f_C$ | 3 | Disconnected (%) ↓ | **0%** | 21.25% |
| | 4 | Disconnected (%) ↓ | **0%** | 17.50% |

Table 4 reports ablation results for the two objective terms, $f_R$ and $f_C$, across space configurations. In three and four-space configurations, the minimum ROI coverage might not be secured, shown by the minimum (min) ROI coverage drop. The effect is most severe in challenging space configurations (group 2, 3 spaces), removing ROI Coverage ($f_{\text{cov}}$) causes the ROI coverage minimum to fall below 0.5 in 18 out of 20 cases, with the minimum ROI coverage collapsing to near zero in the majority. Removing ROI Ownership ($f_{\text{own}}$) also reduces the ownership score, meaning each user's allocated ROI is reduced. This indicates that the optimizer reduces accessibility for one or more users without these explicit ROI terms. Connectivity follows a similar scaling pattern. The percentage of disconnected Neighborhood occurs occasionally in configuration of three and four spaces. Even for the small occurrence, the disconnected neighborhood fails to serve its intended purpose of traversal between spaces.

## 7 Limitations and Future Works

Our system has a few points that call for further research. First, the optimizer may only secure a small Neighborhood size or fail to secure the ROI in small and compact spaces. This was also noted during the exploratory user study, where one participant suggested allowing users to choose whether to join the system with their local space as the number of users increases. Future systems could dynamically adapt individual area sizes based on participant count, trading coverage for connectivity. Second, our rearrangement process preserves topological relationships but not viewing directions. After re-optimization, users maintain their relative positions but may face different orientations than before. This can momentarily disrupt conversational context, suggesting that gaze direction may be considered as an additional optimization objective. Third, our user study was conducted using three virtual rooms due to physical space constraints. Although we encouraged participants to imagine the given spaces as physical, the absence of real objects that would prevent participants from colliding reduced their immersion. If system aims to transit to MR, it will require perceiving local space with video see-through and establishing a spatial anchor to align the Unity coordinate space with the physical room. Beyond these integration steps, reliable real-time walkable-area detection, and semantic scene understanding remain open challenges. Regarding scene update, our current system uses a static mesh of each space, so changes such as object movement cannot be reflected once the space has been scanned. Incorporating real-time scanning technologies such as SLAM could enable continuous scene updates that reflect such dynamic scene changes. We evaluated our system under a single condition. Comparisons with other methods such as portal-based approaches or with more diverse room settings and increased number of users will provide further insight. Furthermore, increased user controllability over ROI selection would increase usability. Our split criterion for quantitative measure may itself introduce bias. Thresholds are defined relative to our own house set reflecting the relative compactness and clutter within this sample. Finally, the computation time of the optimization function could be reduced. During the exploratory user study, participants experienced around 20 seconds delay when transitioning to Walking In mode due to optimization computation. Replacing the current approach with supervised learning trained on offline-generated solutions could enable faster inference.

## 8 Conclusion

We present Neighborhood, a traversable MR telepresence system. Users can experience the illusion of physically walking into remote spaces as they are augmented within the local space. Once users choose the area they wish to share (individual area), the system optimizes the spatial alignment to create the Neighborhood and starts Walking In mode. In this mode, users can simultaneously cross over to each other's augmented spaces. When users want to change the area they wish to share, the system can switch to Panning mode. Here, users can either move to a new area or observe other users moving within their own environment. Through continuous mode changes, users can share and invite all parts of their physical space bidirectionally. We hope that our system shows one vision of a future telepresence experience where users can experience spatial connection as if they are dwelling together.

## Acknowledgments

This work was supported by KOCCA, MCST, South Korea (RS-2026-25525207), IITP, MSIT, South Korea (RS-2026-25507551), and the InnoCORE program, MSIT, South Korea (AI Meta-Scientist, N10260110). We also sincerely thank our reviewers for their valuable feedback.

## References

[1] Christopher Alexander, Sara Ishikawa, and Murray Silverstein. 1977. *A Pattern Language: Towns, Buildings, Construction.* Oxford University Press, New York.

[2] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4 (2019), 589–597. arXiv:https://doi.org/10.1080/2159676X.2019.1628806 doi:10.1080/2159676X.2019.1628806
[3] Soojin Choi, Seokpyo Hong, Kyungmin Cho, Chaelin Kim, and Junyong Noh. 2023. Online Avatar Motion Adaptation to Morphologically-similar Spaces. *Comput. Graph. Forum* 42, 2 (May 2023), 13–24. https://doi.org/10.1111/cgf.14740
[4] Youjin Choi, Jeongmi Lee, and Sung-Hee Lee. 2020. Effects of Locomotion Style and Body Visibility of a Telepresence Avatar. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 1–9. doi:10.1109/VR46266.2020.00017
[5] Ben J. Congdon, Tuanfeng Wang, and Anthony Steed. 2018. Merging environments for shared spaces in mixed reality. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology* (Tokyo, Japan) *(VRST '18)*. Association for Computing Machinery, New York, NY, USA, Article 11, 8 pages. doi:10.1145/3281505.3281544
[6] Matt Deitke, Eli VanderBilt, Alvaro Herrasti, Luca Weihs, Kiana Ehsani, Jordi Salvador, Winson Han, Eric Kolve, Aniruddha Kembhavi, and Roozbeh Mottaghi. 2022. ProcTHOR: Large-Scale Embodied AI Using Procedural Generation. In *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Eds.), Vol. 35. Curran Associates, Inc., 5982–5994. https://proceedings.neurips.cc/paper_files/paper/2022/file/27c546ab1e4f1d7d638e6a8dfbad9a07-Paper-Conference.pdf
[7] Simon J. Gibbs, Constantin Arapis, and Christian J. Breiteneder. 1999. TELEPORT— towards immersive copresence. *Multimedia Syst.* 7, 3 (May 1999), 214–221. doi:10.1007/s005300050123
[8] Jens Emil Sloth Grønbæk, Ken Pfeuffer, Eduardo Velloso, Morten Astrup, Melanie Isabel Sønderkær Pedersen, Martin Kjær, Germán Leiva, and Hans Gellersen. 2023. Partially Blended Realities: Aligning Dissimilar Spaces for Distributed Mixed Reality Meetings. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 456, 16 pages. doi:10.1145/3544548.3581515
[9] Jaylin Herskovitz, Yi Fei Cheng, Anhong Guo, Alanson P. Sample, and Michael Nebeling. 2022. XSpace: An Augmented Reality Toolkit for Enabling Spatially-Aware Distributed Collaboration. *Proc. ACM Hum.-Comput. Interact.* 6, ISS, Article 568 (Nov. 2022), 26 pages. doi:10.1145/3567721
[10] Andrew Irlitti, Mesut Latifoglu, Thuong Hoang, Brandon Victor Syiem, and Frank Vetere. 2024. Volumetric Hybrid Workspaces: Interactions with Objects in Remote and Co-located Telepresence. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 802, 16 pages. doi:10.1145/3613904.3642814
[11] Andrew Irlitti, Mesut Latifoglu, Qiushi Zhou, Martin N Reinoso, Thuong Hoang, Eduardo Velloso, and Frank Vetere. 2023. Volumetric Mixed Reality Telepresence for Real-time Cross Modality Collaboration. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 101, 14 pages. doi:10.1145/3544548.3581277
[12] Dongsik Jo, Ki-Hong Kim, and Gerard Kim. 2015. SpaceTime: Adaptive control of the teleported avatar for improved AR tele-conference experience. *Computer Animation and Virtual Worlds* 26 (04 2015). doi:10.1002/cav.1645
[13] Jiho Kang, Dongseok Yang, Taehei Kim, Yewon Lee, and Sung-Hee Lee. 2023. Real-time Retargeting of Deictic Motion to Virtual Avatars for Augmented Reality Telepresence. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 885–893. doi:10.1109/ISMAR59233.2023.00104
[14] Mohammad Keshavarzi, Allen Y. Yang, Woojin Ko, and Luisa Caldas. 2020. 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 Computer Society, Los Alamitos, CA, USA, 353–362. doi:10.1109/VR46266.2020.00055
[15] Dooyoung Kim, Seonji Kim, Selin Choi, and Woontack Woo. 2024. Spatial Affordance-aware Interactable Subspace Allocation for Mixed Reality Telepresence. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 1256–1265. doi:10.1109/ISMAR62088.2024.00142
[16] Dooyoung Kim, Jae-eun Shin, Jeongmi Lee, and Woontack Woo. 2021. 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)*. 653–660. doi:10.1109/VR50410.2021.00091
[17] Seonji Kim, Dooyoung Kim, Jae-Eun Shin, and Woontack Woo. 2024. 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)*. 796–805. doi:10.1109/VR58804.2024.00099
[18] Taehei Kim, Hyeshim Kim, Jeongmi Lee, and Sung-Hee Lee. 2025. Evaluating user perception toward physics-adapted avatar in remote heterogeneous spaces. *Computers & Graphics* 128 (2025), 104185. doi:10.1016/j.cag.2025.104185
[19] Taehei Kim, Jihun Shin, Hyeshim Kim, Hyuckjin Jang, Jiho Kang, and Sung-Hee Lee. 2025. Voronoi Rooms: Dynamic Visibility Modulation of Overlapping Spaces for Telepresence. *ACM Trans. Graph.* 45, 2, Article 17 (Dec. 2025), 15 pages. doi:10.1145/3777900
[20] Jason Lawrence, Danb Goldman, Supreeth Achar, Gregory Major Blascovich, Joseph G. Desloge, Tommy Fortes, Eric M. Gomez, Sascha Häberling, Hugues Hoppe, Andy Huibers, Claude Knaus, Brian Kuschak, Ricardo Martin-Brualla, Harris Nover, Andrew Ian Russell, Steven M. Seitz, and Kevin Tong. 2021. Project starline: a high-fidelity telepresence system. *ACM Trans. Graph.* 40, 6, Article 242 (Dec. 2021), 16 pages. doi:10.1145/3478513.3480490
[21] Jason Lawrence, Ye Pan, Dan B. Goldman, Rachel McDonnell, Julie Robillard, Carol O'Sullivan, Yaser Sheikh, Michael Zollhoefer, and Jason Saragih. 2022. State of the art in telepresence. In *ACM SIGGRAPH 2022 Courses* (Vancouver, British Columbia, Canada) *(SIGGRAPH '22)*. Association for Computing Machinery, New York, NY, USA, Article 14, 159 pages. doi:10.1145/3532720.3535673
[22] Nicolas H. Lehment, Daniel Merget, and Gerhard Rigoll. 2014. Creating automatically aligned consensus realities for AR videoconferencing. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 201–206. doi:10.1109/ISMAR.2014.6948428
[23] Yi-Jun Li, Hao-Zhong Yang, Wen-Tong Shu, and Miao Wang. 2025. Semantics-Aware Avatar Locomotion Adaption for Indoor Cross-Scene AR Telepresence. *IEEE Transactions on Visualization and Computer Graphics* (2025), 1–16. doi:10.1109/TVCG.2025.3525697
[24] Andrew Maimone, Jonathan Bidwell, Kun Peng, and Henry Fuchs. 2012. Enhanced personal autostereoscopic telepresence system using commodity depth cameras. *Computers & Graphics* 36, 7 (2012), 791–807. doi:10.1016/j.cag.2012.04.011 Augmented Reality Computer Graphics in China.
[25] Andrew Maimone and Henry Fuchs. 2011. Encumbrance-free telepresence system with real-time 3D capture and display using commodity depth cameras. In *Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR '11)*. IEEE Computer Society, USA, 137–146. doi:10.1109/ISMAR.2011.6092379
[26] Andrew Maimone and Henry Fuchs. 2011. A First Look at a Telepresence System with Room-Sized Real-Time 3D Capture and Life-Sized Tracked Display Wall. In *ICAT 2011*. https://api.semanticscholar.org/CorpusID:30511926
[27] Joseph E. McGrath. 1984. *Groups: Interaction and Performance*. Prentice-Hall, Englewood Cliffs, NJ.
[28] Nels Numan, Shwetha Rajaram, Balasaravanan Thoravi Kumaravel, Nicolai Marquardt, and Andrew D Wilson. 2024. SpaceBlender: Creating Context-Rich Collaborative Spaces Through Generative 3D Scene Blending. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24)*. ACM, 1–25. doi:10.1145/3654777.3676361
[29] Stephan Ohl. 2018. Tele-Immersion Concepts. *IEEE Transactions on Visualization and Computer Graphics* 24, 10 (Oct. 2018), 2827–2842. doi:10.1109/TVCG.2017.2767590
[30] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L. Davidson, Sameh Khamis, Mingsong Dou, Vladimir Tankovich, Charles Loop, Qin Cai, Philip A. Chou, Sarah Mennicken, Julien Valentin, Vivek Pradeep, Shenlong Wang, Sing Bing Kang, Pushmeet Kohli, Yuliya Lutchyn, Cem Keskin, and Shahram Izadi. 2016. Holoportation: Virtual 3D Teleportation in Real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST '16)*. Association for Computing Machinery, New York, NY, USA, 741–754. doi:10.1145/2984511.2984517
[31] Wole Oyekoya, Ran Stone, William Steptoe, Laith Alkurdi, Stefan Klare, Angelika Peer, Tim Weyrich, Benjamin Cohen, Franco Tecchia, and Anthony Steed. 2013. Supporting interoperability and presence awareness in collaborative mixed reality environments. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST*, 165–174. doi:10.1145/2503713.2503732
[32] Michael Pabst, Linda Rudolph, Nikolas Brasch, Verena Biener, Chloe Eghtebas, Ulrich Eck, Dieter Schmalstieg, and Gudrun Klinker. 2025. MRUnion: Asymmetric Task-Aware 3D Mutual Scene Generation of Dissimilar Spaces for Mixed Reality Telepresence. *IEEE Transactions on Visualization and Computer Graphics* 31, 5 (2025), 3354–3364. doi:10.1109/TVCG.2025.3549878
[33] Julius Panero and Martin Zelnik. 1979. *Human dimension and interior space: A source book of design reference standards*. Watson-Guptill.
[34] Tomislav Pejsa, Julian Kantor, Hrvoje Benko, Eyal Ofek, and Andrew Wilson. 2016. Room2Room: Enabling Life-Size Telepresence in a Projected Augmented Reality Environment *(CSCW '16)*. Association for Computing Machinery, New York, NY, USA, 1716–1725. doi:10.1145/2818048.2819965
[35] Thammathip Piumsomboon, Gun A. Lee, Jonathon D. Hart, Barrett Ens, Robert W. Lindeman, Bruce H. Thomas, and Mark Billinghurst. 2018. Mini-Me: An Adaptive Avatar for Mixed Reality Remote Collaboration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3173574.3173620
[36] Ramesh Raskar, Greg Welch, Matt Cutts, Adam Lake, Lev Stesin, and Henry Fuchs. 1998. The office of the future: a unified approach to image-based modeling and spatially immersive displays. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '98)*. Association for Computing Machinery, New York, NY, USA, 179–188. doi:10.1145/280814.280861

[37] RecRoom. 2016. Rec Room. https://recroom.com/

[38] Jonas Schjerlund, Kasper Hornbæk, and Joanna Bergström. 2022. OVRlap: Perceiving Multiple Locations Simultaneously to Improve Interaction in VR. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 355, 13 pages. doi:10.1145/3491102.3501873

[39] Ludwig Sidenmark, Tianyu Zhang, Leen Al Lababidi, Jiannan Li, and Tovi Grossman. 2024. Desk2Desk: Optimization-based Mixed Reality Workspace Integration for Remote Side-by-side Collaboration. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 44, 15 pages. doi:10.1145/3654777.3676339

[40] Misha Sra, Aske Mottelson, and Pattie Maes. 2018. Your Place and Mine: Designing a Shared VR Experience for Remotely Located Users. In *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong, China) *(DIS '18)*. Association for Computing Machinery, New York, NY, USA, 85–97. doi:10.1145/3196709.3196788

[41] Anthony Steed, William Steptoe, Wole Oyekoya, Fabrizio Pece, Tim Weyrich, Jan Kautz, Doron Friedman, Angelika Peer, Massimiliano Solazzi, Franco Tecchia, Massimo Bergamasco, and Mel Slater. 2012. Beaming: An Asymmetric Telepresence System. *IEEE Computer Graphics and Applications* 32, 6 (2012), 10–17. doi:10.1109/MCG.2012.110

[42] Theophilus Teo, Mitchell Norman, Gun Lee, Mark Billinghurst, and Matt Adcock. 2020. Exploring interaction techniques for 360 panoramas inside a 3D reconstructed scene for mixed reality remote collaboration. *Journal on Multimodal User Interfaces* 14 (07 2020). doi:10.1007/s12193-020-00343-x

[43] Balasaravanan Thoravi Kumaravel, Fraser Anderson, George Fitzmaurice, Bjoern Hartmann, and Tovi Grossman. 2019. Loki: Facilitating Remote Instruction of Physical Tasks Using Bi-Directional Mixed-Reality Telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 161–174. doi:10.1145/3332165.3347872

[44] Inc. VRChat. 2014. VRChat — hello.vrchat.com. https://hello.vrchat.com/

[45] Chiu-Hsuan Wang, Chia-En Tsai, Seraphina Yong, and Liwei Chan. 2020. Slice of Light: Transparent and Integrative Transition Among Realities in a Multi-HMD-User Environment. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20)*. Association for Computing Machinery, New York, NY, USA, 805–817. doi:10.1145/3379337.3415868

[46] Miao Wang, Yi-Jun Li, Jinchuan Shi, and Frank Steinicke. 2024. SceneFusion: Room-Scale Environmental Fusion for Efficient Traveling Between Separate Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics* 30, 8 (2024), 4615–4630. doi:10.1109/TVCG.2023.3271709

[47] Xuanyu Wang, Hui Ye, Christian Sandor, Weizhan Zhang, and Hongbo Fu. 2022. Predict-and-Drive: Avatar Motion Adaption in Room-Scale Augmented Reality Telepresence with Heterogeneous Spaces. *IEEE Transactions on Visualization and Computer Graphics* 28, 11 (2022), 3705–3714. doi:10.1109/TVCG.2022.3203109

[48] Emily Wong, Adélaïde Genay, Jens Emil Sloth Grønbæk, and Eduardo Velloso. 2025. Spatial Heterogeneity in Distributed Mixed Reality Collaboration. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 980, 19 pages. doi:10.1145/3706598.3714033

[49] Dongseok Yang, Jiho Kang, Taehei Kim, and Sung-Hee Lee. 2024. Visual Guidance for User Placement in Avatar-Mediated Telepresence Between Dissimilar Spaces. *IEEE Transactions on Visualization and Computer Graphics* 30, 12 (2024), 7558–7570. doi:10.1109/TVCG.2024.3354256

[50] Chandan Yeshwanth, Yueh-Cheng Liu, Matthias Nießner, and Angela Dai. 2023. ScanNet++: A High-Fidelity Dataset of 3D Indoor Scenes. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. 12–22. doi:10.1109/ICCV51070.2023.00008

[51] Leonard Yoon, Dongseok Yang, Jaehyun Kim, ChoongHo Chung, and Sung-Hee Lee. 2020. Placement Retargeting of Virtual Avatars to Dissimilar Indoor Environments. *IEEE Transactions on Visualization and Computer Graphics* 28, 3 (March 2020), 1619–1633. doi:10.1109/tvcg.2020.3018458