Automatic Formation Generation based on Scene Awareness for Guided Group Navigation in VR

Jian Wu, Lili Wang*, Zhikai Wen, Yanzhou Chen, Qianwen Wang, Bin Hu, Xuehuai Shi, and Xiaolong Liu

Abstract—Group navigation is a virtual reality (VR) technology designed to replace personal navigation, enhancing efficiency and aiding guided tours in path planning for other users. Group navigation techniques require users to have a clear understanding of navigation and travel modalities. Tour guides must ensure that users do not intersect with the environment or other users' models and must provide reasonable tour formations. To address these needs, we propose a scene-aware automatic formation generation method for fast and easy-to-use guided tours. First, we generate a limited number of candidate points for the exhibition and initialize the visiting group formation. Next, we optimize the formation to maximize view quality using our proposed viewpoint observation score. Finally, we match the visitors to the optimized formation to ensure a minimal view deflection angle. Moreover, we conducted a user study to evaluate the performance of our approach. Compared to the current method, our approach significantly increased navigation efficiency and view quality. Additionally, it substantially decreased the task load and improved the system usability for both tour guides and visitors.

Index Terms—Virtual reality, Group navigation, Formation generation, Scene Awareness, Viewpoint quality

1 Introduction

Increasingly, virtual reality (VR) navigation technologies [1]–[7] allow users to navigate freely and quickly through content in virtual scenes; however, for many scenes (e.g., museums, cultural tourism scenes), guided navigation is essential. Tour guides are able to provide the historical origin of exhibits, plan the best path for scene touring, and provide users with a better navigation experience. To realize guided virtual scene roaming, group navigation techniques have emerged. Group navigation technology [8], [9] aims to overcome the limitations of individual navigation and enable group members to unite even when only one person is responsible for motion control. In a group navigation system, the navigation instructions are the responsibility of one member, and the other members move according to the instructions. This approach improves navigation efficiency and reduces the repetition of navigation instructions.

Lili Wang is the corresponding author: wanglily@buaa.edu.cn
Jian Wu, Lili Wang, Zhikai Wen, and Xiaolong Liu are with State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing, 100191.
Yanzhou Chen, Qianwen Wang, and Bin Hu are with the School of Computer Science and Engineering, Beihang University, Beijing, 100191.
Xuehuai Shi is with the School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, 210023.
Manuscript received XX XX, 2024; revised XX XX, 2024.

Group navigation can help VR users interact and explore the virtual world more effectively and reduce various problems associated with individual navigation. For example, (1) group navigation can provide tourists with a more immersive and interactive travel experience [10]. Users can navigate together in a virtual environment, enhancing their interaction with their surroundings and cultural heritage. (2) Group navigation can facilitate navigation and interaction for individuals with limited mobility [8], [9]. Users with limited mobility can follow the navigation to roam the scene efficiently and fully, reducing the physical burden of the real world. (3) Group navigation provides interactive and experiential learning opportunities for education [11]. Students can interact with classmates and teachers in virtual environments for a more engaging and immersive learning experience. Overall, group navigation in VR has many applications, ranging from enhancing the travel experience to improving crowd management, accessibility, and educational interaction. It enables users to explore virtual environments together, facilitating collaboration and a more immersive shared VR experience.

Group navigation relies heavily on the tour guide to determine the user's viewing position during navigation [9]. However, such an approach places a significant burden on the tour guide. On the one hand, the tour guides must ensure that the selected viewing locations for the users do not cause conflicts between users or between users and the scenes. On the other hand, they need to ensure that each user gets a good viewing position and is not too close or far away from the exhibit. These requirements greatly increase the tour guide's operational burden in planning navigation routes and user viewing positions, causing fatigue and affecting the efficiency of group navigation. Overall, minimizing the burden of tour guides, improving the efficiency of group navigation, and providing users with better navigation and observation positions to enhance the effectiveness of group navigation are challenges that need to be solved urgently.

To address these challenges, we propose an automatic formation generation method for guided group navigation. Once the guide identifies the exhibits, our method first generates a limited number of candidate points for the exhibition and initializes the visiting group formation. Then, the formation is optimized to obtain the maximized formation view quality based on our proposed viewpoint observation score. Finally, the visitors are matched to the optimized formation to ensure a minimum view deflection angle. Compared to current methods, our approach significantly reduces task completion time, decision-making time for tour guides, overall viewpoint quality scores, the number of teleportations, and user

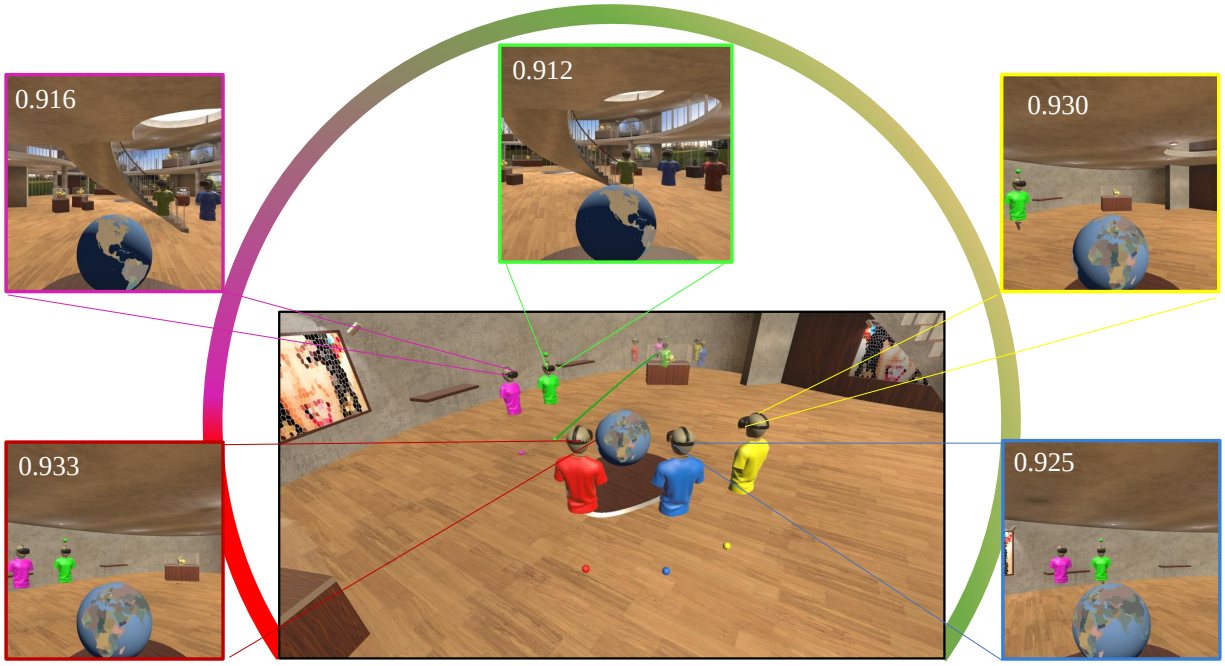


Fig. 1. The scenario of the *Museum* group navigation task. The bottom middle panel shows the visiting formation for the five users optimized with our scene-aware group navigation method. The images observed by each user and the corresponding observation scores are visualized in a small window.

deflection angles. Our method significantly reduces the task load for tour guides. Fig. 1 illustrates our method in the scenario of the *Museum* group navigation task. The bottom middle panel shows the visiting formation for the five users optimized by our method. The images observed by each user and the corresponding observation scores are visualized in a small window.

In summary, our main contributions are as follows:

- A guided group navigation pipeline that enables automatic formation generation based on scene awareness and allows a group of visitors to navigate scenes efficiently and with high view quality.
- Viewpoint Observation Score - A new metric that measures the viewpoint quality in observing exhibition areas and guides group spatial formation generations.
- A visiting formation optimization method that maximizes the total quality of the visitor's viewpoints and supports multilayer formation with obstacle avoidance.

2 Related Work

The utilization of VR in online socialization has advanced the evolution of multi-person collaborative VR technology. Group navigation represents a distinctive instance within multi-person collaborative VR technology involving the tour guide's navigation operations and route decisions. As virtual teaching and the virtual cultural tourism industry progress, group navigation is poised to become an essential technology in this domain. This section provides a concise overview of the relevant research on group navigation (subsection 2.1) and scene awareness (subsection 2.2).

2.1 Group Navigation in VR

Prior to the application of group navigation techniques, most multiplayer navigation virtual scene roaming systems provided independent virtual navigation for each user, including projection-based systems [12]–[16], and head-mounted display-based systems [17]–[23]. Early research on social VR was primarily based on distributed collaboration frameworks, where users operated in separate workspaces with limited support for coordinated navigation [24]. Subsequent work began to explore group navigation techniques for pairs (duo-based group navigation). While dyadic collaboration showed advantages over individual navigation in acquiring survey knowledge [25], it also revealed challenges such as users struggling to maintain co-location, easily becoming separated, and experiencing inconsistencies in understanding spatial references [21]. To improve dyadic collaboration, Weissker et al. proposed a two-person jumping technique that allows a leading user to carry a partner during teleportation, enhancing mutual understanding of the navigation path [26]. Beyond jumping techniques, other researchers introduced group World-in-Miniature (group WiM) methods, enabling guides to macroscopically control team positioning through a miniature replica of the environment [27], [28]. Weissker et al. also experimented with temporary group formation via virtual hand-holding, reporting positive initial usability in two-user scenarios [29]. Further extending group navigation research, they investigated formation strategies using five-user group configurations and proposed a generalized framework that scales to arbitrary group sizes [30], while Brument et al. compared the effects of different group locomotion techniques on spatial awareness and user preference [31]. However, these approaches primarily focus on dyadic scenarios or emphasize macroscopic control of group positioning, paying insufficient attention to how group members collectively perceive navigation actions and target exhibits in dynamic environments. In contrast, our proposed method aims to dynamically interpret environmental semantics and automatically generate adaptive formations, systematically enhancing group co-location experience and shared situational

awareness.

To keep multiple users involved in navigation and maintain an understanding of each other's actions, group navigation techniques have begun to emerge in distributed VR systems. In a review of commercial social virtual reality applications, Kolesnichenko et al. reported mechanisms that allow users to form groups to switch between different virtual environments [32]. Later, Weissker et al. introduced a framework stating that group navigation techniques in distributed virtual reality should allow users to form navigational groups (Forming), distribute navigational responsibilities (Norming), navigate together (Performing), and eventually split up again (Adjourning) [8]. Building on this, Weissker et al. extended the number of users by predefining jump locations and having a guide adjust the parameters of the formation tour instead of each individual, effectively meeting the needs of understandability, obstacle avoidance, and view optimization when navigating in a group [9]. They presented a group navigation technique based on short-distance teleportation ("jumping") for guided tours in distributed virtual environments. A guide controls the movement of the entire group—including position, formation, and orientation—using a handheld controller. The system enhances predictability and mutual awareness through preview avatars, personal target rays, and collision detection. It supports both formation-preserving and formation-changing jumps, offering predefined formations like circle, horseshoe, grid, or line to suit different spatial and viewing needs. Recently, Rasch et al. also explored shared visualizations to support the understanding of motion intent, capable of increasing group feeling during navigation [33]. In the process of multi-person, *Multi-ray Jumping* group navigation [9], the guide has a heavy burden: it is necessary to ensure that the formation is reasonable. On the one hand, tour guides need to ensure that the formation of the visiting group does not result in intersections between users, as this would significantly impact the users' navigation experience. On the other hand, guides also need to ensure that each user obtains a favorable position for the tour. These requirements cause a high burden for the tour guide in planning the navigation target formation, which requires constant manual adjustment of the formation.

In this paper, we present a group navigation technique in virtual reality that relies on scene awareness. This technique can autonomously generate a more sensible recommended formation based on scene awareness, thereby alleviating the tour guide's workload and enriching the overall user tour experience.

2.2 Scene Awareness

In virtual reality, there is a close relationship between scene awareness and viewpoint quality [34]. Scene awareness refers to the user's perception and understanding of the surrounding environment and events in the virtual environment. Viewpoint quality, on the other hand, pertains to the quality and clarity of the visual information observed by the user from a specific viewpoint [35]. When users navigate in virtual reality, viewpoint quality plays a crucial role in their scene awareness. High-quality viewpoints provide clearer and more realistic visual information, enabling users to perceive and understand the details and elements of the virtual environment better. This enhances their perception and cognition of the scene, resulting in a more immersive experience. Conversely, low-quality viewpoints may result in blurry, distorted, or unclear visual information. This can impact the user's perception of details in the scene, leading to a reduced understanding and awareness of the environment. Users may miss important details or struggle to accurately judge spatial relationships and object positions within the virtual environment. Therefore, providing high-quality viewpoints is essential in virtual reality to support users' scene awareness.

A good viewpoint for an object or scene can be defined as providing a clear and unobstructed view of the object or scene while highlighting important visual features [36]. Additionally, a good viewpoint should allow the viewer to focus on the most relevant or interesting aspects of the object or scene, such as its shape, texture, or color, and provide a sense of depth and perspective. The concept of the general position of the viewpoint proposed by Kamada et al. refers to a specific location or angle from which an object can be viewed in a way that provides the maximum amount of shape information in the rendering image [37]. Plemenos et al. proposed a new constraint for selecting a viewpoint that provides a good visualization of an object [38]. According to their paper, a viewpoint that maximizes the angular deviation between the view direction and the surface normal of the object provides the best visualization of its geometric details. Vázquez et al. proposed a method for automatically exploring scenes based on viewpoint entropy [34]. It takes into account both the projection area and the number of visible faces of the object. Sokolov et al. proposed the concept of "viewpoint quality" [35]. They suggested that in the context of global world exploration, the quality of a viewpoint could be determined by two factors: total curvature for meshes and the projection area of the visible region of the objects. Later, they proposed a method for calculating viewpoint quality in automatic 3D scene exploration, which considers three factors: the size of the object bounding box, the observation quality, and the fraction of visible area of the object [39]. Recently, Freitag et al. [40] proposed a method to normalize the viewpoint quality values according to the viewpoint quality of the whole scene and used the normalized viewpoint quality to adjust the travel speed when traversing large scenes automatically. Then they proposed an interactive assist interface based on automatic analysis of object visibility and viewpoint quality to support exploration and guide the user to the interesting parts of the scene [36]. Key aspects of the viewpoint quality included the object's uniqueness and the visual size of the object. Wang et al. constructed a viewpoint quality function and evaluated the viewpoints of multiple users by calculating its three components: the visibility of the object needs to be manipulated, the visibility of the target, the depth and distance combined of the target [41]. While these methods provide valuable metrics for static and individual viewpoint assessment, they exhibit critical limitations. First, they are predominantly rooted in geometric and information-theoretic principles (e.g., visible area, curvature, entropy), fundamentally ignoring the semantic context of the scene. Second, and most critically, these approaches are inherently designed for a single user. They compute a single, optimal view, utterly failing to address the core challenge in group navigation: how to simultaneously satisfy the visual and navigational needs of multiple users.

Besides, for tasks such as museum navigation, since visitors aim to obtain a clear and feature-rich viewing angle of the target exhibit, the visibility of the target exhibit is not sufficient to evaluate the user's viewpoint quality. Therefore, in this paper we introduce a novel approach to create adaptable viewing formations based on scene awareness automatically. Whenever a guide establishes a navigation position, the method dynamically analyzes the scene in real time using our defined *Viewpoint Observation Score*

(subsubsection 3.3.1). Leveraging this information, the method can autonomously modify the position and posture of each user, ensuring a sensible formation and an optimal viewing experience.

3 METHOD

Previous work has implemented a group navigation approach based on short-distance teleportation that is easy to understand and well-suited for conducting guided tours [9]. However, the tour guide must manually control and adjust the visitors' tour formations during the navigation process. This creates a significant workload for the tour guide. Our method automatically generates formation after the tour guide selects an exhibit, thus reducing the guide's workload and enriching the tour experience.

3.1 Overview

Our method proceeds according to the pipeline shown in Fig. 2. The overall approach involves establishing a viewpoint quality assessment function. Once the guide identifies the exhibits, the system automatically optimizes and calculates the best viewing formation and position for each visitor, avoiding collisions between visitors and maximizing the overall quality of the visual experience within the guided tour group. The method liberates the guide from arranging the formation and positions of individual visitors, thereby reducing the guide's workload and allowing them to focus on introducing exhibits to the visitors.

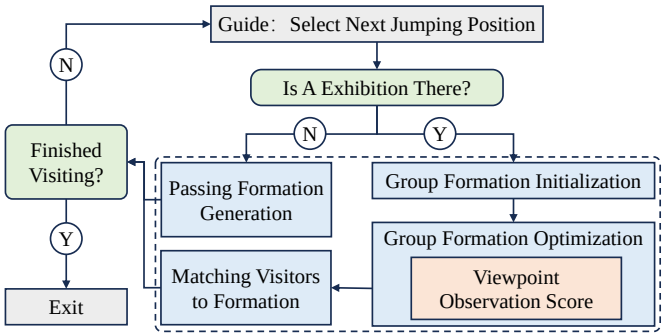


Fig. 2. Overview of our automatic formation generation method for guided group navigation in VR.

Whenever the tour guide selects a teleport/jump location, our system will automatically detect the presence of exhibits at the target location. If an exhibit exists, the system will start our scene-aware group navigation method. The method first initializes the group spatial formation near the target exhibit (subsection 3.2). Then, it executes the group formation optimization algorithm to optimize the formation shape using a scene awareness function - Viewpoint Observation Score - so that the formation maximizes the quality of the user's viewpoint while satisfying the obstacle avoidance and social distance constraints (subsection 3.3). Finally, the visitors are matched with the optimized formation, so the user jumps to the formation with a minimum total deflection angle (subsection 3.4).

If no exhibit exists in the selected location, the system generates a passing formation and automatically jumps to it when the tour guide confirms. Our approach adds scene-aware functionality to the grid formation and adopts an easy-to-use strategy when generating the passing formation. The implementation details are presented in subsection 3.5. Finally, after the guide ends the navigation activity, the system stops and dismisses the group.

3.2 Group Formation Initialization

When the guide fires a ray to apply a potential jump, if the ray intersects the exhibit's bounding box (pre-defined in the scene), the exhibit will be selected after the guide confirms by pressing the confirm button on the controller. Once the guide has selected the target exhibit, the task of the method is to generate a spatial formation of visitors in the vicinity of the exhibit. In this task, we require the following rules to be satisfied: 1. All visitors in the formation face the center of mass of the exhibit. 2. The distance between any two visitors should be kept above the "social distance" to avoid a collision. 3. All visitors in the formation do not collide with the obstacles in the scene.

As described in Fig. 2, the first step in applying the visiting is to initialize the group formation. We first automatically generate an initial circular or arc-shaped formation (which satisfies rules 1 and 3) from sampled candidate visiting points in the walkable area based on whether the exhibits are located on the walls of the scene and make the visiting positions in the initial formation spaced beyond the "social distance" (which satisfies rule 2 and was set as 0.5m in practice). Algorithm 1 shows the detailed initialization process.

Algorithm 1: Group Formation Initialization

input : visitors number n, minimum social distance d, formation type T_f, arc angle range $[\phi_s, \phi_e]$, exhibit position $e.p$
output : initial position $P[0 \sim n-1]$

1 **if** $T_f == CIRCLE$ **then**
2 $r = d/2sin(\pi/n)$
3 $\Delta = \delta$
4 **for** $i = 0$ to $n-1$ **do**
5 $P[i] = e.p + r \cdot (cos(2\pi i/n), sin(2\pi i/n))$
6 **while** $P[i].isCollided()$ **do**
7 $P[i] = e.p + (r-\Delta) \cdot (cos(2\pi i/n), sin(2\pi i/n))$
8 $\Delta += \delta$
9 **if** $T_f == ARC$ **then**
10 $r = d/sin[(\phi_e - \phi_s)/2(n-1)]$
11 $\Delta = \delta$
12 **for** $i = 0$ to $n-1$ **do**
13 $P[i] = e.p + r \cdot (cos(i(\phi_e - \phi_s)/(n-1)), sin(i(\phi_e - \phi_s)/(n-1)))$
14 **while** $P[i].isCollided()$ **do**
15 $P[i] = e.p + (r-\Delta) \cdot (cos(i(\phi_e - \phi_s)/(n-1)), sin(i(\phi_e - \phi_s)/(n-1)))$
16 $\Delta += \delta$
17 **return** $P[0 \sim n-1]$

The initialization follows the same rules for both circular (lines 1-8) and arc-shaped formations (lines 9-16). The radius r of the circle/arc is calculated according to the visitors' number n and the minimum social distance d (line 2/line 10). We then distribute the initial positions of the formation evenly on the circle/arc (lines 4-5/lines 12-13). If the initial position collides with an obstacle, we find a feasible alternative position by contracting (Δ for each iteration) the radius of that position (lines 6-8/lines 14-16). The iteration step is set as δ. Such an approximately uniformly distributed initial formation is useful for subsequent optimization.

3.3 Group Formation Optimization

The following steps involve optimizing the initial formation obtained in subsection 3.2, where the positions are adjusted to improve the overall viewpoint quality. The three rules proposed in subsection 3.2 must be satisfied during the optimization process. To quantitatively measure the viewpoint quality, we introduce the concept of viewpoint observation score and propose an algorithm for optimizing the group formation based on it.

3.3.1 Viewpoint Observation Score

In order to optimize the calculation of visitors' observation positions for exhibits during group roaming, it is necessary to quantify the visual quality when visitors observe exhibits from different locations. In the context of group navigation, a favorable visitor viewpoint often implies a larger proportion of the target exhibit's image in the view. Additionally, the image should present the exhibit more completely rather than just a partial view. Furthermore, efforts should be made to minimize instances where exhibits are obstructed by other objects. Simultaneously, exhibits should not be positioned too far or close to the visitors.

We employ five factors to construct the quantitative indicator of viewpoint quality, called viewpoint observation score (VOS). They consist of the ratio of the exhibit's portion within the field of view to the entire field of view, the ratio of the exhibit's portion within the field of view to the total exhibit, the proportion of the exhibit's visible part within the field of view, the color quality of the exhibit (which measures the colorfulness of the exhibit's image from the current position), and the depth quality of the exhibit (a comprehensive metric of the depth and distance of the exhibit from the current viewpoint). These factors are weighted with coefficients k_1, k_2, k_3, k_4, and k_5 (0.41, 0.30, 0.20, 0.06, 0.03), respectively. The values of these coefficients were determined using the analytical hierarchy process (AHP) [42] following a subjective experiment in which participants compared different images and rated the five factors. We recruited 8 participants who were all familiar with VR experiences. We showed them many comparison images of the exhibits with coefficients ranging from k_1 to k_5 and explained how these coefficients affect the images. Then, they were asked to rate each coefficient pair, following the analytical hierarchy process (AHP). The study generated a total of 3 (exhibits) × 10(coefficient pairs) × 8 (participants) = 240 data points. We organized them into ten groups, each containing 24 ratings. After calculating the geometrical average, we got the average rating matrix in Table 1. Finally, we ran the AHP analysis with SPSSAU software. The

TABLE 1
AHP rating matrix.

	k_1	k_2	k_3	k_4	k_5
k_1	1.00	2.34	2.59	5.41	8.34
k_2	0.43	1.00	3.23	4.95	7.06
k_3	0.39	0.31	1.00	5.68	8.44
k_4	0.18	0.20	0.18	1.00	3.41
k_5	0.12	0.14	0.12	0.29	1.00

results are as follows:

$$\{k_1, k_2, ..., k_5\} = \{0.41, 0.30, 0.20, 0.06, 0.03\} \quad (1)$$

The consistency ratio $C_r = 0.08 < 0.1$, so the rating matrix (Table 1) passes the consistency test.

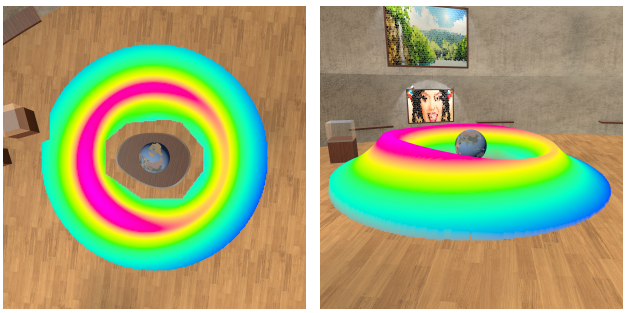
Algorithm 2 uses multiple images rendered by the camera at the viewpoint to measure the VOS. The algorithm's input is the exhibit e to be measured and the current measured viewpoint v. The algorithm's output is the VOS *score* of the exhibit at the current viewpoint. The resolution of all the images rendered in the algorithm is 1024×1024.

Algorithm 2: Viewpoint Observation Score Calculation

input : the target exhibit e, the current viewpoint v
output : viewpoint observation score s

1 $I_a \leftarrow$ RenderAll(v, $FOV = 60°$)
2 $I_e \leftarrow$ RenderExhibit(v, e, $FOV = 60°$)
3 $I_l \leftarrow$ RenderExhibit(v, e, $FOV = 170°$)
4 $A_e \leftarrow$ Area(I_e)
5 $A_l \leftarrow$ Area(I_l)
6 $A_u \leftarrow$ Unobscured(I_a, I_e)
7 $s_1 \leftarrow A_e / A_{screen}$
8 $s_2 \leftarrow$ Integrity(A_e,A_l)
9 $s_3 \leftarrow A_u / A_e$
10 $s_4 \leftarrow$ ColorfulnessQuality(I_e)
11 $s_5 \leftarrow$ DepthQuality(I_e)
12 $s \leftarrow k_1 * s_1 + k_2 * s_2 + k_3 * s_3 + k_4 * s_4 + k_5 * s_5$

In this algorithm, we first render the image of the exhibit by rendering all visible objects at the current viewpoint position I_a, the image of only the current exhibit I_e, and the image of the exhibit with a larger FOV I_l. The FOV of I_a and I_e are set to both 60°, the eccentric angle of the near peripheral region [43], while the FOV of I_l is set to 170° to ensure that the camera can see the complete exhibit (lines 1-3). Then we calculate the number of pixels A_e of the exhibit in I_e, the number of pixels A_l of the exhibit in I_l, and the number of pixels A_u of the exhibit that is not obscured by other objects by comparing I_a and I_e pixel by pixel (lines 4-6). After that, we can calculate the exhibit area percentage s_1, the exhibit completeness s_2 (the pixel projection area ratio for different FOV is taken into account), and the exhibit unobscured proportion s_3 (lines 7-9). Then we compute the color quality s_4 according to the algorithm proposed by David Hasler [44] (line 10) and calculate the depth quality s_5 according to the algorithm proposed by Wang [45] (line 11). Finally, we obtained the final viewpoint observation score s by weighted summation of coefficients (line 12). Fig. 3 shows the VOS field around the exhibit *Earth*, calculated by Algorithm 2 with sampled points.



(a) (b)

Fig. 3. Visualization of the VOS of the exhibit *Earth*, the more red the color, the higher the VOS, and the more blue the color, the lower the VOS.

(a) (b) (c)

Fig. 4. Trajectory diagram of the Group Formation optimization process. The colored avatars show the final optimized positions and the curves on the ground show the optimizing trajectories. *Left:* Visitors can perceive detailed information in the exhibit and look for a more colorful view of the exhibit with our optimization. *Middle:* The algorithm can automatically line up the formation in two layers to stagger the viewpoints when there are too many visitors (16 in the image). *Right:* The trajectories of two visitors (leftmost and rightmost) show that our optimization can sense obstacles in the scene and bypass them.

3.3.2 Optimization

The introduction of the VOS metric allows us to use the camera to quantitatively perceive the quality of the viewpoints in the scene, thus enhancing the overall visit experience. We have devised an iterative algorithm (Algorithm 3) based on the greedy approach, whereby each viewpoint seeks and moves toward the location with the highest viewpoint score in its vicinity. The algorithm's input is the exhibit e to be visited and the sequence of current positions $P[0 \sim n-1]$ of the Group Formation. The output is the optimized Group Formation $Q[0 \sim n-1]$.

Algorithm 3: Group Formation Optimization

input : exhibit e, initial formation positions $P[0 \sim n-1]$
output : optimized formation positions $Q[0 \sim n-1]$

1 $Q[0 \sim n-1] \leftarrow P[0 \sim n-1]$
2 **while** iter $<$ max & !oscillating **do**
3 $P[0 \sim n-1] \leftarrow Q[0 \sim n-1]$
4 **for** $i = 0$ to $n-1$ **do**
5 $VP[0 \sim m-1] \leftarrow$VPNearby$(P[i])$
6 **for** $j = 0$ to $m-1$ **do**
7 **if** TestValidation$(VP[j], P[i+1 \sim n-1], Q[0 \sim i-1])$ **then**
8 $VP[j].score \leftarrow$ VOSCalc$(e, VP[j])$
9 $Q[i] = VP[j].score > Q[i].score ? VP[j] : Q[i]$
10 $iter++$
11 **return** $Q[0 \sim n-1]$

At the beginning of the Algorithm 3, the optimized formation $Q[0 \sim n-1]$ is initialized as the initial formation $P[0 \sim n-1]$ (line 1). The final Group Formation is obtained through multiple rounds of iteration (line 2). In each iteration, we first traverse all initial positions in P and generate m candidate visiting points around each position $P[i]$. For example, for one position $P[i]$, m candidate positions around $P[i]$ are tested if they are legal based on Rule 2 and Rule 3 in subsection 3.2 (line 4-7). Each candidate position is tested with the original formation $P[i+1 \sim n-1]$ and optimizing formation $Q[0 \sim i-1]$ in the current iteration loop. In practice, we set m as 9, which means that a 3×3 grid (in practice, the side length of each grid is set as 0.2m) around $P[i]$ on the floor is constructed to generate the candidate positions. Then, according to Algorithm 2, we measure the VOS of these points VP and select the highest one as the next position for the corresponding one $Q[i]$ (lines 8-9). After each iteration, we get a new group formation Q, and the next round of iteration begins. There are two conditions for ending the iteration: reaching the maximum number of iterations (in practice, we set it to 40) or encountering an "oscillation" phenomenon (line 2), where the difference of the group formation Q between the $iter$-th and $(iter+2)$-th iterations is small enough.

For algorithm 3, the final optimized formation layout is affected by the optimization order. The first position to be processed always gets the best candidate position, while later positions may be limited by the update formation and cannot be optimized to the best candidate positions around them. Our algorithm is based on a greedy strategy that, while it may not result in an optimized formation, ensures that the computation time is not too long and does not confuse the user.

Fig. 4 shows the Group Formation optimization process trajectories for three exhibits in the scene *Museum*. In the left image, the final positions enable users to perceive more detailed and colorful views of the exhibit *Earth*. In the middle image, the optimizing trajectories show that the algorithm can automatically line up the formation in two layers to stagger the viewpoints when there are too many visitors (16 in the image). In the right image, the initial formation is an arc-shaped formation. The trajectories of the two avatars (leftmost and rightmost) indicate that our optimization can sense obstacles in the scene and bypass them.

3.4 Matching Visitors to Formation

When the optimized final group formation is reached, the next step is to perform target matching, which involves individually matching the current group formation positions with the visitors. When teleporting in a virtual environment, virtual rotation (especially unexpected or large rotations) is a major cause of spatial

disorientation and cognitive load [46]–[48]. Given the consistent research showing that rotation is disruptive, we chose to minimize the rotation displacement of each user during teleportation, which is a straightforward and scientifically proven strategy to maintain user spatial orientation and reduce discomfort. Therefore, we have employed a target-matching algorithm (Algorithm 4) to achieve the minimum deflection angle.

Algorithm 4: Matching Visitors to Formation

input : visitor sequence $V[0 \sim n-1]$, visiting formation $Q[0 \sim n-1]$

output : reordered visitor sequence $V'[0 \sim n-1]$, reordered preview avatars formation $Q'[0 \sim n-1]$

1 $V_{sorted}[0 \sim n-1] \leftarrow V[0 \sim n-1]$ sorted by Vec2toAngle($V[i].f.xz$)
2 $Q_{sorted}[0 \sim n-1] \leftarrow Q[0 \sim n-1]$ sorted by Vec2toAngle($Q[i].f.xz$)
3 **for** $i = 0$ to $n-1$ **do**
4 $sum_{angle} = \sum_{j=0}^{n-1}$ GetDeltaAngle($V_{sorted}[(j+i)\%n].f.xz$, $Q_{sorted}[j].f.xz$)
5 **if** $sum_{angle} < angle_{min}$ **then**
6 $angle_{min} = sum_{angle}$
7 $front = i$
8 $V'[0 \sim n-1] \leftarrow V_{sorted}[front, front+1, ..., n-1, 0, 1, ..., front-1]$
9 $Q'[0 \sim n-1] \leftarrow Q_{sorted}[0 \sim n-1]$

The basic idea of Algorithm 4 is to sort the current visitor sequence according to its forward angle within the previously optimized visitor formation. This results in two circular sequences of forward angles—one determined by the sorted current visitor sequence, and the other by the sorted visitor formation generated by Algorithm 3. We then perform a clockwise matching search on these two circular sequences of forward angles to develop a matching method. By iterating through all possible matches without changing the relative order within each cycle, we can obtain the optimal solution with the goal of minimizing the deflection angle.

The algorithm's inputs are the current visitor sequence $V[0 \sim n-1]$, and the visiting formation $Q[0 \sim n-1]$ obtained in the previous optimization step. where $V[i].f$ denotes the forward direction of the $i-th$ visitor and $Q[i].f$ denotes the forward direction of the $i-th$ visiting position (toward the corresponding exhibit centroid). The algorithm's output is an array of visitors sequence $V'[0 \sim n-1]$ and an array of visiting formation $Q'[0 \sim n-1]$, both with the changed order. $V'[i]$ corresponds to $Q'[i]$ one by one, satisfying that the sum of visitors' deflection angles before and after the jump is minimized in this correspondence.

In this algorithm, we first sort the current visitor sequence and the visiting formation, using the orientation projection angle of visitors/visiting positions on the floor as the sorting keyword. The orientation indicates the direction from the positions to the exhibit e, and its projection yields an angle value between $[0, 2\pi]$ with respect to the positive x-axis (lines 1-2). We then traverse all possible matching cases while preserving the relative order of visitors, i.e., by rotating $V_{sorted}[0 \sim n-1]$ n times (lines 3-4). For each visitor-visiting position pair, we can calculate a deviation angle between two positions on the floor, which is between $[0, \pi]$. Then, we sum all pairs to obtain the total deviation angle sum_{angle} (line 4). The corresponding rotation index $front$, i.e., the number of iterations required to obtain the minimum deviation angle $angle_{min}$, is recorded (lines 5-7). Finally, the corresponding visitors sequence and visiting formation are returned, according to the rotation index $front$ (lines 8-9).

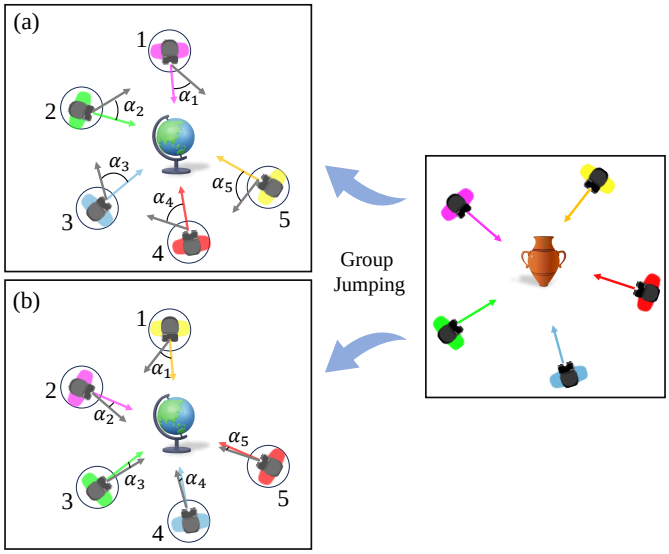


Fig. 5. Five visitors are jumping from the vase to the globe. The five gray circles around the globe show the visiting formation generated by our method. Two arrows are displayed to show the pre-jumping forward direction (gray) and the post-jumping forward direction (clothing color). The angles between them are the visitors' deflection angles ($\alpha_1 \sim \alpha_5$). The visitors are matched to the formation according to their serial number (a)/the minimum total deflection angle (b).

Fig. 5 compares the results of the two matching methods after the jumping operation. The gray arrows in (a) and (b) indicate the pre-formation's forward directions, which were pointing to the pre-exhibit (the vase). It can be seen that directly using the serial number to match results in a larger total deflection angle for the visitors (a), while the deflection angle calculated using our method is smaller (b).

3.5 Implemention Details

3.5.1 Passing Formation Generation

When the distance between the exhibit and the next exhibit is too far, we need to jump to the non-exhibit area for transition, then the requirement of "view optimization" for group navigation may decrease, but the requirement of "obstacle avoidance" may increase, as exhibit aisles may be irregular and narrow. Therefore, we still need to satisfy Rule 2 and Rule 3 in subsection 3.2, but not necessarily Rule 1 (in fact, we set everyone in the passing formation to face the same direction).

Our approach to generating the passing formation is similar to the previous method [9], but calculated automatically. The preview passing formation of the visitors is pre-defined as a grid shape according to the group size N. The width of the grid n_w can be calculated as follows:

$$n_w = \lfloor \sqrt{N-1} \rfloor + 1 \tag{2}$$

For example, when $N \in [5, 9]$, $n_w = 3$, means three people in a row in the target formation. If a position in the formation collides with the scene, it will be skipped until all visitors have available

positions. If there are not enough positions in the candidate grid, a new row will be added to the back end of the grid. Finally, a grid-enabled passing formation is formulated.

3.5.2 GN Settings

In our implementation, the calculated group formation will be shown to the users with a preview avatar before applying the jump. This preview mechanism can alleviate users' spatial disorientation during VR navigation tasks [9]. During the group formation optimization process, the preview avatar will constantly flash to alert guides and visitors that the optimization is in progress. For further details, please also refer to the accompanying video.

4 User Study

4.1 Overview and Hypotheses

To evaluate the effect of our method, we conducted the empirical study. In this study, we aimed to verify the effectiveness of our method in improving navigation efficiency and user experience. Thus, we formulate the following main hypotheses:

H1. Our automatic formation generation method can enable faster navigation and shorter decision time for tour guides.

H2. Our method can provide significantly lower deflection angles than the previous method.

H3. Our method can significantly decrease the task load for tour guides during navigation.

4.2 Paticipants and Apparatus

We recruited 24 volunteers (none of them participated in the last pilot user study), 12 males and 12 females between 20 and 31 years old ($M = 24.58, SD = 3.29$), with normal vision (or corrected-to-normal vision by wearing glasses). They came from academic contexts. Seventeen of them had previous virtual reality experiences. They were, hence, able to provide valuable feedback, discussions, and suggestions regarding our system. Before each participant started the experiment, we measured the inter-pupillary distance (IPD) for them with a millimeter scale and adjusted the IPD of the headset to meet their best visual setting.

We used two Pico 4 head-mounted display (HMD) systems, each with two handheld controllers, allowing two users to point virtual lasers at the virtual environment (VE). Each HMD was connected to its workstation with a 5.00GHz Intel(R) Core(TM) i7-12700KF CPU, 16GB of RAM, and an NVIDIA GeForce GTX 3080 graphics card. The tracked physical space hosting the VR applications is 4m × 4m. We used Unity 2020.1 to implement our VR collaborative manipulation tasks. The virtual environments were rendered at 90fps for each eye.

4.3 User Study Design

We used a 2×3 repeated measures within-subject user study design. We tested two group navigation methods: Weissker's method [9] (*CC*) and our method (*EC*), with three group sizes (5, 10, and 15 visitors). The *CC* method uses a group multi-ray jumping technique with four pre-set formations for different visiting scenarios. During navigation, the guide selects the suitable formation before jumping, adjusts the size and rotation, and transports the visitors. Our *EC* method also uses the group multi-ray jumping technique. The visitor group formation is automatically calculated. The guide only needs to select the target exhibit and confirm the previewed formation. Only two real participants were present during the experiments; the others were simulated avatars. The simulated avatars followed the tour trajectory during each navigation and were programmed to look around occasionally to add realism.

The participants were first randomly split into two groups: the guide group (6 males and 6 females, $M = 25.42, SD = 2.63$) and the visitor group (6 males and 6 females, $M = 23.75, SD = 3.65$). The study task is as follows: one guide should lead a visitor group (only one real visitor) to complete two guided tours using the above two group navigation methods. During the tours, the guide controlled the navigation route, and the visitor just followed and looked around. They are allowed to communicate with each other. Fig. 6 shows the virtual *Museum* scene we used in our user study. We set up six exhibits for the participants, and the guides led the visitors along the dashed line route to view each exhibit.

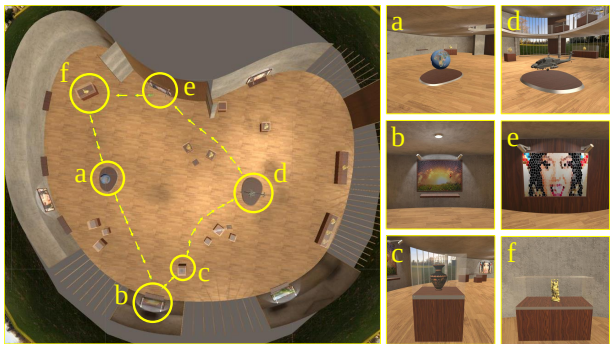


Fig. 6. Left: Top-view of the *Museum* scene. (a) to (f) shows the six target exhibits. The yellow dashed lines show one navigation route.

4.4 Procedure and Metrics

Before each task, the participant pair (the guide and the visitor) signed the information sheet upon arrival at the experimental site. Then we explained and demonstrated the various navigation operations used in the process. After that, they conducted technical training, during which the guide should be familiar with all the operations and navigation routes. The participants acting as guides were informed of the navigation route by the experimenter in advance. Before the formal experiment began, the guides were given sufficient time to familiarize themselves with the route. During the experiment, the experimenter remained present throughout to provide assistance and would remind the guides if they reported forgetting the route. However, no instances where the guides forgot the route occurred during the entire experiment.

For our approach, the preview mechanism that visitors see is similar to that in previous research [9]. This consists of a curved ray projected from the tour guide controller indicating the intended teleportation landing zone, and a semi-transparent preview avatar showing their future position in the intended formation. After determining the new formation, the tour guide instructs the visitor to pay attention to the next exhibit location and to perform the jump when ready.

When they were confirmed to be ready, the study began. First, with a group size of 5, the guide conducted the navigation task following the pre-defined route with two comparing methods. Then, the group size increased from 10 to 15, and the same tasks were conducted. For all three group sizes, the order of comparison methods remains unchanged. The participant pair was asked to rest for at least two minutes after completing a guided tour. The whole process lasted about 20 to 30 minutes. The order in which

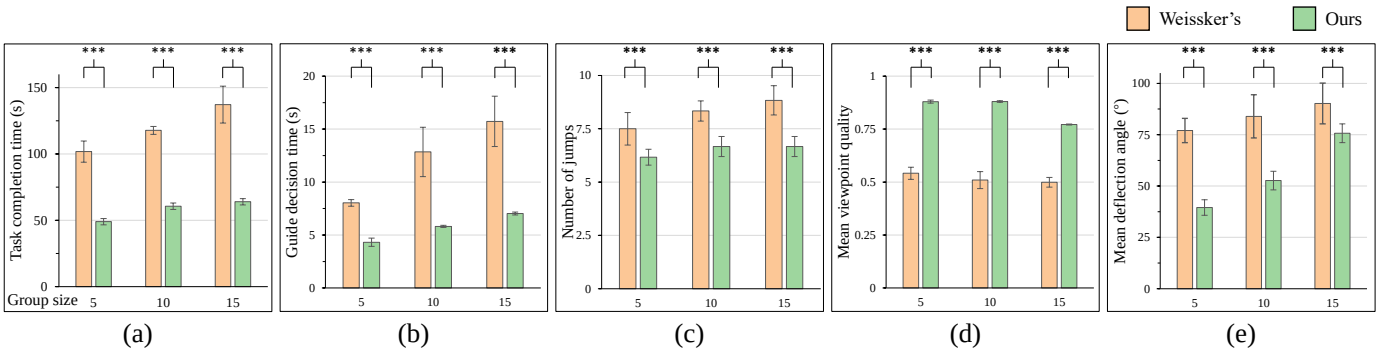


Fig. 7. Plots (a) to (e) show five objective metrics results. The horizontal coordinates indicate the different group sizes. The two compared methods of group navigation are distinguished by different colors. The error line demonstrates the standard deviation, and three asterisks indicate that the results of comparing the two methods are significantly different ($p < 0.001$).

each participant group used the comparison methods (*CC&EC*) was randomly assigned and remained evenly distributed among the 12 groups.

We counted the following five objective metrics:

- Task completion time, the total time it took for a group to walk a tour route and visit all six exhibits.
- Guide decision time, the average duration time from when the guide selected the target exhibit to the successful confirmation of the next visiting formation.
- Number of jumps, the total number of jumps required for a single tour route.
- Mean viewpoint quality, the average of the VOS of the visual field images observed by visitors facing the exhibits during one tour route.
- Mean deflection angle, the average value of the deflection angle in the direction of the line of sight when the visiting team jumped to the exhibit formation.

The mean viewpoint quality and mean deflection angle were calculated from the recorded data at the end of the tour to ensure that participants and guides were not disturbed during the experiment. For subjective metrics, we used the standard NASA TLX questionnaire [49], [50] to measure the task load of participants and the System Usability Score (SUS) [51] to measure the usability of the methods. For each participant pair, both the guide and the visitor filled out the questionnaires after they completed all the tasks.

4.5 Results

4.5.1 Objective Metrics

Fig. 7 gives the results for each objective metric (a for task completion time, b for guide decision time, c for number of jumps, d for mean viewpoint quality, e for mean deflection angle). In each subfigure, the results of two group navigation methods are compared, and the marked asterisks indicate the differences were statistically significant. In the following analysis, a two-factor multivariate analysis of variance (MANOVA) was used to analyze the effects of the group navigation method (GN-Method) and group size on the performance of the above five objective metrics. For each group navigation method and group size condition, the outlier data points were first filtered out (± 3 standard deviation). In total, we removed 2 data points (2.78%).

Before the analysis, the assumptions of the methodology were tested. Scatter plots found a linear relationship between the dependent variables within each group of the independent variables. Pearson's correlation found no multicollinearity between the two dependent variables ($|r| < 0.9$). No one-way outliers were found by box plots, and no multivariate outliers were found by the Mahalanobis distance [52].

Shapiro-Wilk's test showed that the four dependent variables (task completion time, guide decision time, mean viewpoint quality, and mean deflection angle) obeyed a normal distribution ($p > 0.05$), while the number of jumps variable did not ($p = 0.001$). However, since MANOVA is robust to deviations from normality, particularly when group sample sizes are equal or nearly equal, the non-normal distribution does not significantly increase the likelihood of a Type I error. Therefore, we proceeded with the test without additional adjustments.

Box's M test showed that the variance/covariance matrices of the two dependent variables within each group of the independent variables were not equal ($p < 0.006$). So, we used Pillai's criterion statistic because it is more robust to unequal covariance matrices. Levene's test showed that the dependent variables within each group of the independent variables were equal in variance ($p > 0.05$).

Interaction Effect There was a statistically significant effect of the interaction of GN-Method and group size on the dependent variables ($F_{10,126} = 22.168, p < 0.001, \Lambda_{Pillai} = 1.275, \eta_p^2 = 0.638$), i.e., there was a difference in the effect of the group size on the dependent variables between two GN-Method. The MANOVA test revealed statistically significant effects of the interaction of the GN-Method and group size on task completion time ($F_{2,66} = 13.506, p < 0.001, \eta_p^2 = 0.29$), guide decision time ($F_{2,66} = 18.684, p < 0.001, \eta_p^2 = 0.362$), mean viewpoint quality ($F_{2,66} = 26.749, p < 0.001, \eta_p^2 = 0.448$) and mean deflection angle ($F_{2,66} = 15.604, p < 0.001, \eta_p^2 = 0.321$), but not on number of jumps ($F_{2,66} = 3.119, p = 0.051, \eta_p^2 = 0.086$).

Simple Main Effect Because of the statistically significant interaction effects between the GN-Method and group size factors, we report separate simple main effects for each factor here.

GN-Method The simple main effect analysis revealed that the difference between the two group navigation methods was statistically significant for all group size levels on task completion time ($p < 0.001$), number of jumps ($p < 0.001$), guide decision time ($p < 0.001$), mean viewpoint quality ($p < 0.001$) and mean deflection angle ($p < 0.001$). Pairwise comparison tests show that

the differences between the two GN-Method levels were statistically significant for all dependent variables and all group size levels ($p < 0.006$), as shown in Table 2.

Group size The simple main effect analysis revealed that the difference between the three group sizes was statistically significant for all GN-Method levels on task completion time ($p < 0.001$), guide decision time ($p < 0.001$), mean viewpoint quality ($p < 0.001$) and mean deflection angle ($p < 0.001$). On the number of jumps, the difference between the three group sizes was statistically significant for the *CC* method ($p < 0.001$) but not for the *EC* method ($p = 0.059$).

For each dependent variable and each GN-Method level, pairwise comparisons were conducted for the three group size levels, as shown in Table 3. For the task completion time, with the *CC* method, the differences between each group size level pair were statistically significant ($p < 0.001$). When it comes to the *EC* method, the differences were statistically significant ($p \leq 0.001$) for the comparison of 5 and 10, and 5 and 15 group size levels. However, the comparison of the 10 and 15 two group sizes shows $p = 0.021$, indicating that the difference between these two conditions is smaller than the difference observed when compared to the group size of 5.

For the guide decision time, when using the *CC* method, the differences between each group size level pair were statistically significant ($p < 0.001$). When using the *EC* method, the difference between each group size level pair were also statistically significant ($p < 0.001$).

For the number of jumps, when using the *CC* method, the differences in comparing the group size levels: 5 and 10, 5 and 15 were statistically significant ($p = 0.01, p = 0.002$), while the difference between group size levels 10 and 15 was not statistically significant ($p = 0.166$). When using the *EC* method, the differences in comparing the group size levels 5 and 15 was statistically significant ($p = 0.007$), while the differences in comparing the group size levels: 5 and 10, 10 and 15 were not statistically significant ($p = 0.053, p = 1.00$).

For the mean viewpoint quality, when using the *CC* method, the differences in comparing the group size levels: 5 and 10, 5 and 15 were statistically significant ($p = 0.002, p < 0.001$), while the difference between group size levels 10 and 15 was not statistically significant ($p = 0.343$). When using the *EC* method, the differences in comparing the group size levels: 5 and 15, 10 and 15 were statistically significant ($p < 0.001$), while the difference between group size levels 5 and 10 was not statistically significant ($p = 0.676$).

For the mean deflection angle, when using the *CC* method, the difference between group size levels 5 and 15 was statistically significant ($p = 0.005$), while the differences in comparing the group size levels: 5 and 10, 10 and 15 were not statistically significant ($p = 0.4, p = 0.143$). When using the *EC* method, the differences between each group size level pair were statistically significant ($p < 0.001$).

Main Effect Below we report the main effects of each factor.

GN-Method: Multivariate tests show that the main effect of GN-Method on the dependent variables was statistically significant ($F_{5,62} = 2020.475, p < 0.001, \Lambda_{Pillai} = 0.994, \eta_p^2 = 0.994$). Univariate main effect tests revealed statistically significant effects of the GN-Method on task completion time ($F_{1,66} = 1301.425, p < 0.001, \eta_p^2 = 0.952$), number of jumps ($F_{1,66} = 157.776, p < 0.001, \eta_p^2 = 0.705$), guide decision time ($F_{1,66} = 366.379, p < 0.001, \eta_p^2 = 0.847$), mean viewpoint quality ($F_{1,66} = 3461.561, p < 0.001, \eta_p^2 = 0.981$) and mean deflection angle ($F_{1,66} = 253.983, p < 0.001, \eta_p^2 = 0.794$). Pairwise comparison showed significant differences ($p < 0.001$) between the two compared group navigation methods on all five objective metrics.

TABLE 2
Pairwise Comparison tests result for different group navigation method conditions (Cond.) for all dependent variables (Metric) and all group sizes. Significant difference is marked with asterisk.

Metric	Group Size	Cond.	$t_{df=11}$	p	Cohen'd	Effect size
Task completion time	5	*CC* *EC*	20.83	$< 0.001^*$	6.01	Huge
	10	*CC* *EC*	52.48	$< 0.001^*$	15.15	Huge
	15	*CC* *EC*	16.79	$< 0.001^*$	4.85	Huge
Guide decision time	5	*CC* *EC*	22.32	$< 0.001^*$	6.44	Huge
	10	*CC* *EC*	9.77	$< 0.001^*$	2.82	Huge
	15	*CC* *EC*	11.71	$< 0.001^*$	3.38	Huge
Number of jumps	5	*CC* *EC*	4.69	$< 0.001^*$	1.35	Very Large
	10	*CC* *EC*	7.42	$< 0.001^*$	2.14	Huge
	15	*CC* *EC*	8.01	$< 0.001^*$	2.31	Huge
Mean viwpoint quality	5	*CC* *EC*	45.39	$< 0.001^*$	13.10	Huge
	10	*CC* *EC*	32.36	$< 0.001^*$	9.34	Huge
	15	*CC* *EC*	37.82	$< 0.001^*$	10.92	Huge
Mean deflection angle	5	*CC* *EC*	21.65	$< 0.001^*$	6.25	Huge
	10	*CC* *EC*	7.41	$< 0.001^*$	2.14	Huge
	15	*CC* *EC*	3.40	0.006^*	0.98	Large

Group size: Multivariate tests show that the main effect of group size on the dependent variables was statistically significant ($F_{10,126} = 17.437, p < 0.001, \Lambda_{Pillai} = 1.161, \eta_p^2 = 0.581$). Univariate main effect tests revealed statistically significant effects of group size on task completion time ($F_{2,66} = 74.171, p < 0.001, \eta_p^2 = 0.692$), number of jumps ($F_{2,66} = 15.925, p < 0.001, \eta_p^2 = 0.326$), guide decision time ($F_{2,66} = 79.768, p < 0.001, \eta_p^2 = 0.707$), mean viewpoint quality ($F_{2,66} = 67.710, p < 0.001, \eta_p^2 = 0.672$) and mean deflection angle ($F_{2,66} = 67.525, p < 0.001, \eta_p^2 = 0.671$).

Since the group size is a triple categorical variable, we report the multiple comparison results for each dependent variable. For the task completion time, the differences between 5 and 10, 10 and 15, and 5 and 15 group size levels were statistically significant ($p < 0.001$). For the guide decision time, the differences between 5 and 15, and 10 and 15 group size levels were statistically significant ($p < 0.001$), while the difference between the 5 and 10 group size levels was not statistically significant ($p = 0.063$). For the number of jumps, the differences between 5 and 10, and 5 and 15 group size levels were statistically significant ($p = 0.001, p < 0.001$), while the difference between the 10 and 15 group size levels was not statistically significant ($p = 0.303$). For the mean viewpoint quality, the differences between 5 and 10, 10 and 15, and 5 and 15 group size levels were statistically significant ($p < 0.001$). For the mean

TABLE 3
Pairwise Comparison tests result for different group size pairs (GS pair) for all dependent variables (Metric) and all group navigation method conditions (Cond.). Significant difference is marked with asterisk.

Metric	Cond.	GS pair	$t_{df=11}$	p	Cohen' d	Effect size
Task completion time	*CC*	5(10)	-6.45	$< 0.001^*$	1.86	Very Large
		10(15)	-5.43	$< 0.001^*$	1.57	Very Large
		5(15)	-7.02	$< 0.001^*$	2.03	Huge
	EC	5(10)	-12.63	$< 0.001^*$	3.65	Huge
		10(15)	-2.68	0.021	0.77	Medium
		5(15)	-12.20	$< 0.001^*$	3.52	Huge
Guide decision time	*CC*	5(10)	-6.53	$< 0.001^*$	1.89	Very Large
		10(15)	-6.14	$< 0.001^*$	1.77	Very Large
		5(15)	-9.84	$< 0.001^*$	2.84	Huge
	EC	5(10)	-11.30	$< 0.001^*$	3.26	Huge
		10(15)	-22.50	$< 0.001^*$	6.49	Huge
		5(15)	-19.19	$< 0.001^*$	5.54	Huge
Number of jumps	*CC*	5(10)	-3.08	0.01^*	0.89	Large
		10(15)	-1.48	0.166	0.43	Small
		5(15)	-4.00	0.002^*	1.16	Large
	EC	5(10)	-2.17	0.053	0.63	Medium
		10(15)	0.00	1.00	0.00	-
		5(15)	-3.32	0.007^*	0.96	Large
Mean viwpoint quality	*CC*	5(10)	4.01	0.002^*	1.16	Large
		10(15)	0.99	0.343	0.29	Small
		5(15)	4.98	$< 0.001^*$	1.44	Very Large
	EC	5(10)	-0.43	0.676	-0.12	Very Large
		10(15)	145.44	$< 0.001^*$	41.99	Huge
		5(15)	61.53	$< 0.001^*$	17.76	Huge
Mean deflection angle	*CC*	5(10)	-2.33	0.400	0.67	Medium
		10(15)	-1.58	0.143	0.46	Small
		5(15)	-3.50	0.005^*	1.01	Large
	EC	5(10)	-7.02	$< 0.001^*$	2.03	Huge
		10(15)	-10.00	$< 0.001^*$	2.89	Huge
		5(15)	-15.60	$< 0.001^*$	4.50	Huge

deflection angle, the differences between 5 and 10, 10 and 15, and 5 and 15 group size levels were statistically significant ($p < 0.001$).

4.5.2 Subjective Metrics

In Fig. 8, the left half compares the NASA-TLX total scores of Weissker's method and our method. A one-way ANOVA revealed a significant difference among the two methods ($p < 0.001$). Subsequently, post-hoc pairwise t-tests with Holm-Bonferroni correction were employed and revealed that the task load of our method is significantly lower than that of the comparing method for both guides ($t(11) = 17.15, p < 0.001, Cohen'd = 4.95, Effectsize : Huge$) and visitors ($t(11) = 11.39, p < 0.001, Cohen'd = 3.29, Effectsize : Huge$).

The right half compares the SUS scores of Weissker's method and our method. A one-way ANOVA revealed a significant difference among the two methods ($p = 0.004$). The Holm–Bonferroni corrected post-hoc pairwise t-test also revealed that the system usability of our method is significantly higher than that of the comparing method for both guides ($t(11) = -3.87, p = 0.003, Cohen'd = 1.12, Effectsize : Large$) and visitors ($t(11) = -4.54, p < 0.001, Cohen'd = 1.31, Effectsize : VeryLarge$).

4.6 Discussion

We designed the user study to compare two different GN methods for performing group navigation tasks with three different group sizes (5, 10, 15) in VR. The results clearly show that our automatic formation generation for the GN method significantly improved performance. Based on the hypotheses above, we conducted the following discussions.

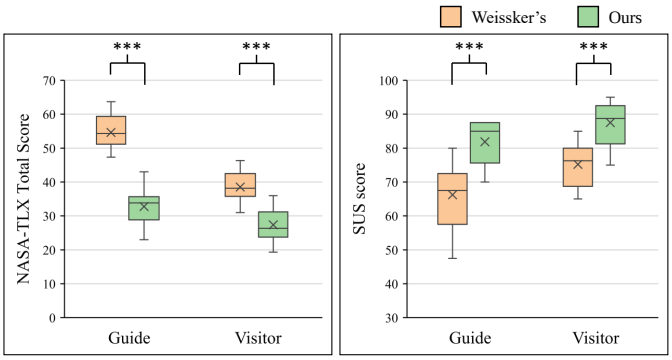


Fig. 8. Box plots for NASA-TLX total scores and SUS scores of the two compared methods for both guide and visitor groups. Three asterisks indicate that the results of comparing the two methods are significantly different ($p < 0.001$).

4.6.1 Efficiency

Above, the results show that for the group navigation task, our method has obtained a significant improvement in efficiency compared to Weissker's method. The main effects analysis showed a 51.37% improvement in task completion time, from 118.92 ± 17.54 to 57.83 ± 6.96 seconds (averaged for three group sizes). Implied from Fig. 7, (a)-(c), the improvement in task completion time is mainly due to less guide decision time (The main effects analysis shows that this metric was reduced by 53.11% for our method, from 12.20 ± 3.77 to 5.72 ± 1.15) in our method, and partially benefited from a reduction in the number of jumps (The results of the main effects analysis showed that this metric was reduced by 20.92% for our method, from 8.22 ± 0.87 to 6.50 ± 0.51). Clearly, **H1** was supported.

We believe this improvement results from reducing the tour guide's operational burden. In Weissker's method, the tour guide must determine the target location, select the formation, adjust its size to avoid collisions, optimize its direction, and finally confirm the jumping. In contrast, our method only requires the guide to select the target exhibit, after which the system automatically calculates the optimal group formation. The guide then confirms the jumping. Therefore, in our method, the tour guide only needs to perform two actions: selection and confirmation, significantly reducing the guide's decision time.

On the other hand, our method significantly reduced the number of jumps. From the user study process and results, we believe this is because, in Weissker's method, the tour guides must first jump to a position where they can better observe the exhibits and their surroundings before determining the next exhibit observation formation. This allows them to conveniently select and adjust the group formation. In contrast, our method eliminates this step. The tour guide only needs to select the exhibits from a distance and then confirm the jump. As a result, the number of jumps is also reduced.

The reduction in the two metrics mentioned above has the combined effect of decreasing the total task completion time.

4.6.2 Navigating Experience

Our method also yielded significant improvements over Weissker's method for the mean deflection angle metrics and the mean viewpoint quality. The main effects analysis showed a decrease of 33.15% in mean deflection angle, from 83.74 ± 10.65 to 55.98 ± 15.78. **H2** was supported.

The viewing direction used by algorithm 4 is the viewing direction after the previous optimization, i.e., the orientation of the exhibits in the previous formation. In practical applications of our method, when the tour guide triggers the formation computation, the users' viewing directions do not exactly match the previously optimized viewing direction. Therefore, the actual viewing deviation perceived by users is greater than the optimal viewing deviation calculated by the algorithm.

To evaluate the impact of the deviation angle, we used each user's actual viewing direction at the moment the tour guide triggered the formation computation when calculating the Mean Deflection Angle in the user experiment. The experimental results indicate that our method outperforms the comparative methods in terms of the deflection angle metric. We believe this is because, although the users' viewing directions are constantly changing, their attention remains more focused on the target exhibit during the task of observing it, which aligns with our design of using the direction toward the exhibit for the new formation computation. Of course, in non-ideal scenarios, such as when users are distracted by external stimuli, a significant increase in the deflection angle may occur if more users deviate from the exhibit's direction.

In terms of the viewpoint quality, the main effects analysis showed an increase of 61.54%, from 0.52 ± 0.04 to 0.84 ± 0.05. It is important to note that the superior viewpoint quality scores achieved by our algorithm were expected by design, as the evaluation metric was identical to the algorithm's internal optimization function. This result primarily serves to validate the computational performance of our implementation. The more significant finding, however, is that this quantitative improvement translated into tangible user benefits, namely reduced task load and higher usability for the guide, as captured by **H1** and **H3**.

4.6.3 Subjective Feedback

As evidenced by the subjective metrics, our method demonstrates lower task load and higher system availability for both guided tours and visitors compared to Weissker's method. For the guides, the average NASA-TLX total score decreased from 54.64 ± 4.79 to 32.75 ± 5.28, and the system availability score improved from 66.25 ± 9.49 to 81.875 ± 5.87. For visitors, the average NASA-TLX total score decreased from 38.53 ± 4.23 to 27.39 ± 4.66, and the system usability score improved from 75.21 ± 6.08 to 87.50 ± 6.04. Hence, **H3** was supported.

After participants completed the subjective questionnaire, we asked them about their experiences with the two methods. All participants reported that they could clearly feel the improvement in experience from the increased efficiency of our method. Visitors particularly noted that the reduced waiting time for the guide to decide the next observation position alleviated boredom. For tour guides, the simplified operation made group navigation easier and eliminated the need to set up the formation, making them feel more relaxed and more inclined to engage in conversation with visitors.

Additionally, when asked about optimizing viewpoint quality and deflection angle, 10 participants reported noticing an improvement in the quality of the field of view of the target exhibit. In contrast, only 4 visitors mentioned an enhanced jumping experience of our method due to the smaller deflection angle, noting that they did not experience a significant loss of their sense of direction before and after the jump. The optimization of the experience resulting from improved viewpoint quality is obvious to visitors. While reducing the deflection angle improves experience, the improvement is not significant enough to be readily noticeable.

4.6.4 General Discussion

Post-experiment interviews with users revealed their subjective experiences. Some users (including tour guides and visitors) expressed a preference for our method's automatic formation calculation feature. However, some users (primarily tour guides) disliked the inability to control their users' tour formations. They wished they could proactively determine their preferred formation for the entire group, rather than simply selecting a target. In group navigation tours, manual and automated formation control should be viewed as complementary tools. Manual control retains distinct advantages: it provides guides with complete situational awareness and precise deployment capabilities, supports exploratory scenarios requiring collaborative positioning, and accommodates subjective preferences beyond algorithmic optimization. Conversely, automated methods excel in efficiency-driven tasks (such as standard guided tours) where the primary objective is to rapidly and consistently deliver quality viewing experiences. Accordingly, future work will focus on developing hybrid interactive systems that allow algorithms to suggest optimal formations while allowing guides to adjust, override, or refine these recommendations before execution – thereby integrating the efficiency of automatic formation generation with human judgment and creativity.

5 Conclusion, Limitation and Future Work

We propose an automatic formation generation method based on scene awareness for guided group navigation in VR. First, a limited number of candidate visiting points are generated for the exhibition, and the visiting formation is initialized. Then, the formation is optimized based on scene awareness (viewpoint observation score), and the visitors are matched to the optimized formation to ensure a minimum view deflection angle. We also applied a pilot user study to define the ratios of the five factors related to the VOS. Finally, we conducted a user study to evaluate the performance of our method. Compared to SOTA group navigation methods, our approach significantly reduces task completion time, guides decision time, the number of jumps in tours, viewpoint quality scores, and user deflection angles. Our method significantly reduces the task load and improves system usability for both guides and visitors.

Our method also suffers from some limitations. One of them is that our approach considers obstacle avoidance only by group members during the transition route to the next exhibit, and does not account for situations in which the route is occluded by obstacles during the transition. Route occlusion often leads tourists to lose their sense of where they are. Future work could focus on alleviating occlusion along the transition route using available approaches, such as occlusion transparency, route enhancement, and multi-view rendering.

Another limitation is the constant flashing of the formation preview, which is not a good indicator for the guides and visitors. Our original intention was to provide a clear visual indicator during the optimization process to reduce the surprise of automatic teleportation. In the future, we should explore a smoother and more predictable animation that would be more effective and less disruptive for both guides and visitors.

Another limitation is that the group navigation team's formation is not fair to all visitors, and there are differences in the quality of viewpoints among visitors. In the future, it is possible to explore how to introduce vision-sharing techniques to guide the design of

visitor group formation, so that different visitors can obtain optimal viewpoints during their visit.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China through Projects 62507002, U25A20441 and 62372026, the National Key R&D plan SQ2025YFE0200118, and the Fundamental Research Funds for the Central Universities.

REFERENCES

[1] D. A. Bowman, E. Kruijff, J. J. LaViola, and I. Poupyrev, "An introduction to 3-d user interface design," Presence, vol. 10, no. 1, pp. 96–108, 2001.

[2] A. Krekhov, S. Cmentowski, K. Emmerich, M. Masuch, and J. Krüger, "Gullivr: A walking-oriented technique for navigation in virtual reality games based on virtual body resizing," in Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play, 2018, pp. 243–256.

[3] T. Weißker, A. Kunert, B. Fröhlich, and A. Kulik, "Spatial updating and simulator sickness during steering and jumping in immersive virtual environments," in 2018 IEEE conference on virtual reality and 3D user interfaces (VR). IEEE, 2018, pp. 97–104.

[4] J. Clifton and S. Palmisano, "Effects of steering locomotion and teleporting on cybersickness and presence in hmd-based virtual reality," Virtual Reality, vol. 24, no. 3, pp. 453–468, 2020.

[5] M. J. Habgood, D. Moore, D. Wilson, and S. Alapont, "Rapid, continuous movement between nodes as an accessible virtual reality locomotion technique," in 2018 IEEE conference on virtual reality and 3D user interfaces (VR). IEEE, 2018, pp. 371–378.

[6] K. Rahimi, C. Banigan, and E. D. Ragan, "Scene transitions and teleportation in virtual reality and the implications for spatial awareness and sickness," IEEE transactions on visualization and computer graphics, vol. 26, no. 6, pp. 2273–2287, 2018.

[7] D. Hayashi, K. Fujita, K. Takashima, R. W. Lindeman, and Y. Kitamura, "Redirected jumping: Imperceptibly manipulating jump motions in virtual reality," in 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE, 2019, pp. 386–394.

[8] T. Weissker, P. Bimberg, and B. Froehlich, "Getting there together: Group navigation in distributed virtual environments," IEEE Transactions on Visualization and computer graphics, vol. 26, no. 5, pp. 1860–1870, 2020.

[9] T. Weissker and B. Froehlich, "Group navigation for guided tours in distributed virtual environments," IEEE Transactions on Visualization and Computer Graphics, vol. 27, no. 5, pp. 2524–2534, 2021.

[10] A. Kunert, A. Kulik, S. Beck, and B. Froehlich, "Photoportals: shared references in space and time," in Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing, 2014, pp. 1388–1399.

[11] A. Kunert, T. Weissker, B. Froehlich, and A. Kulik, "Multi-window 3d interaction for collaborative virtual reality," IEEE transactions on visualization and computer graphics, vol. 26, no. 11, pp. 3271–3284, 2019.

[12] F. Argelaguet, A. Kulik, A. Kunert, C. Andujar, and B. Froehlich, "See-through techniques for referential awareness in collaborative virtual reality," International Journal of Human-Computer Studies, vol. 69, no. 6, pp. 387–400, 2011.

[13] A. J. Fairchild, S. P. Campion, A. S. García, R. Wolff, T. Fernando, and D. J. Roberts, "A mixed reality telepresence system for collaborative space operation," IEEE Transactions on Circuits and Systems for Video Technology, vol. 27, no. 4, pp. 814–827, 2016.

[14] M. Gross, S. Würmlin, M. Naef, E. Lamboray, C. Spagno, A. Kunz, E. Koller-Meier, T. Svoboda, L. Van Gool, S. Lang et al., "blue-c: a spatially immersive display and 3d video portal for telepresence," ACM Transactions on Graphics (TOG), vol. 22, no. 3, pp. 819–827, 2003.

[15] S. Lang, S. Würmlin, J. Borchers, L. Hovestadt, and M. H. Gross, "blue-c: Using 3d video for immersive telepresence applications," Technical Report/ETH Zurich, Department of Computer Science, vol. 484, 2005.

[16] R. Schroeder, A. Steed, A.-S. Axelsson, I. Heldal, Å. Abelin, J. Wideström, A. Nilsson, and M. Slater, "Collaborating in networked immersive spaces: as good as being there together?" Computers & Graphics, vol. 25, no. 5, pp. 781–788, 2001.

[17] A. Hatzipanayioti, A. Pavlidou, M. Dixken, H. H. Bülthoff, T. Meilinger, M. Bues, and B. J. Mohler, "Collaborative problem solving in local and remote vr situations," in 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE, 2019, pp. 964–965.

[18] M. Le Chénéchal, T. Duval, V. Gouranton, J. Royan, and B. Arnaldi, "Help! i need a remote guide in my mixed reality collaborative environment," Frontiers in Robotics and AI, vol. 6, p. 106, 2019.

[19] S. Sharma, S. Jerripothula, S. Mackey, and O. Soumare, "Immersive virtual reality environment of a subway evacuation on a cloud for disaster preparedness and response training," in 2014 IEEE symposium on computational intelligence for human-like intelligence (CIHLI). IEEE, 2014, pp. 1–6.

[20] H. J. Smith and M. Neff, "Communication behavior in embodied virtual reality," in Proceedings of the 2018 CHI conference on human factors in computing systems, 2018, pp. 1–12.

[21] C. Y. Wang, M. Sakashita, U. Ehsan, J. Li, and A. S. Won, "Again, together: Socially reliving virtual reality experiences when separated," in Proceedings of the 2020 chi conference on human factors in computing systems, 2020, pp. 1–12.

[22] 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.

[23] Y. Zhang, H. Nguyen, N. Ladevèze, C. Fleury, and P. Bourdot, "Virtual workspace positioning techniques during teleportation for co-located collaboration in virtual reality using hmds," in 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE, 2022, pp. 674–682.

[24] J. McVeigh-Schultz, A. Kolesnichenko, and K. Isbister, "Shaping pro-social interaction in vr: an emerging design framework," in Proceedings of the 2019 CHI conference on human factors in computing systems, 2019, pp. 1–12.

[25] L. E. Buck, T. P. McNamara, and B. Bodenheimer, "Dyadic acquisition of survey knowledge in a shared virtual environment," in 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE, 2020, pp. 579–587.

[26] T. Weissker, A. Kulik, and B. Froehlich, "Multi-ray jumping: comprehensible group navigation for collocated users in immersive virtual reality," in 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE, 2019, pp. 136–144.

[27] V. Chheang, F. Heinrich, F. Joeres, P. Saalfeld, B. Preim, and C. Hansen, "Group wim: A group navigation technique for collaborative virtual reality environments," in 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). IEEE, 2022, pp. 556–557.

[28] V. Chheang, F. Heinrich, F. Joeres, P. Saalfeld, R. Barmaki, B. Preim, and C. Hansen, "Wim-based group navigation for collaborative virtual reality," in 2022 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR). IEEE, 2022, pp. 82–92.

[29] T. Weissker, P. Bimberg, A. Kodanda, and B. Froehlich, "Holding hands for short-term group navigation in social virtual reality," in 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). IEEE, 2022, pp. 728–729.

[30] T. Weissker, M. Franzgrote, T. Kuhlen, and T. Gerrits, "On the computation of user placements for virtual formation adjustments during group navigation," in 2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). IEEE, 2024, pp. 396–402.

[31] H. Brument, T. Lautenbach, and I. Podkosova, "Spatial awareness and user preferences during group locomotion in virtual reality: A study with four-user groups," in 2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). IEEE, 2024, pp. 403–409.

[32] A. Kolesnichenko, J. McVeigh-Schultz, and K. Isbister, "Understanding emerging design practices for avatar systems in the commercial social vr ecology," in Proceedings of the 2019 on Designing Interactive Systems Conference, 2019, pp. 241–252.

[33] J. Rasch, V. D. Rusakov, M. Schmitz, and F. Müller, "Going, going, gone: Exploring intention communication for multi-user locomotion in virtual reality," in Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 2023, pp. 1–13.

[34] P. Vázquez, M. Feixas, M. Sbert, and W. Heidrich, "Viewpoint selection using viewpoint entropy," in Vision Modeling and Visualization Conference, 2001.

[35] D. Sokolov and D. Plemenos, "Viewpoint quality and scene understanding," in The 6th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST, M. Mudge, N. Ryan, and R. Scopigno, Eds. The Eurographics Association, 2005.

[36] S. Freitag, B. Weyers, and T. W. Kuhlen, "Interactive exploration assistance for immersive virtual environments based on object visibility and viewpoint quality," in 2018 IEEE Conference on virtual reality and 3D user interfaces (VR). IEEE, 2018, pp. 355–362.

[37] T. Kamada and S. Kawai, "A simple method for computing general position in displaying three-dimensional objects," Comput. Vis. Graph. Image Process., vol. 41, pp. 43–56, 1988.

[38] D. Plemenos and M. Benayada, "Intelligent display in scene modelling. new techniques to automatically compute good views." in GraphiCon'96, Saint Petersburg (Russia), July 1996.
[39] D. Sokolov, D. Plemenos, and K. Tamine, "Viewpoint quality and global scene exploration strategies," in GRAPP 2006: Proceedings of the First International Conference on Computer Graphics Theory and Applications, Setúbal, Portugal, February 25-28, 2006, 2006.
[40] S. Freitag, B. Weyers, and T. W. Kuhlen, "Automatic speed adjustment for travel through immersive virtual environments based on viewpoint quality," in 3d User Interfaces, 2016.
[41] L. Wang, X. Liu, and X. Li, "Vr collaborative object manipulation based on viewpoint quality," in 2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2021, pp. 60–68.
[42] O. S. Vaidya and S. Kumar, "Analytic hierarchy process: An overview of applications," European Journal of operational research, vol. 169, no. 1, pp. 1–29, 2006.
[43] S. Ivančić Valenko, V. Cviljušac, S. Zlatić, and D. Modrić, "The impact of physical parameters on the perception of the moving elements in peripheral part of the screen," Tehnički vjesnik, vol. 26, no. 5, pp. 1444–1450, 2019.
[44] D. Hasler and S. E. Suesstrunk, "Measuring colorfulness in natural images," in Human vision and electronic imaging VIII, vol. 5007. SPIE, 2003, pp. 87–95.
[45] L. Wang, X. Liu, and X. Li, "Vr collaborative object manipulation based on viewpoint quality," in 2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). IEEE, 2021, pp. 60–68.
[46] J. W. Kelly, A. G. Ostrander, A. F. Lim, L. A. Cherep, and S. B. Gilbert, "Teleporting through virtual environments: Effects of path scale and environment scale on spatial updating," IEEE Transactions on Visualization and Computer Graphics, vol. 26, no. 5, pp. 1841–1850, 2020.
[47] M. Funk, F. Müller, M. Fendrich, M. Shene, M. Kolvenbach, N. Dobbertin, S. Günther, and M. Mühlhäuser, "Assessing the accuracy of point & teleport locomotion with orientation indication for virtual reality using curved trajectories," in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 2019, pp. 1–12.
[48] L. A. Cherep, A. F. Lim, J. W. Kelly, D. Acharya, A. Velasco, E. Bustamante, A. G. Ostrander, and S. B. Gilbert, "Spatial cognitive implications of teleporting through virtual environments." Journal of Experimental Psychology: Applied, vol. 26, no. 3, p. 480, 2020.
[49] S. G. Hart, "Nasa-task load index (nasa-tlx); 20 years later," in Proceedings of the human factors and ergonomics society annual meeting, vol. 50, no. 9. Sage publications Sage CA: Los Angeles, CA, 2006, pp. 904–908.
[50] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in Advances in psychology. Elsevier, 1988, vol. 52, pp. 139–183.
[51] J. Brooke et al., "Sus-a quick and dirty usability scale," Usability evaluation in industry, vol. 189, no. 194, pp. 4–7, 1996.
[52] G. J. McLachlan, "Mahalanobis distance," Resonance, vol. 4, no. 6, pp. 20–26, 1999.

6 Biography Section

Jian Wu received his Ph.D. degree from the State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing, China. He is currently an assistant professor at the School of Computer Science and Engineering, Beihang University, Beijing, China. His current research focuses on virtual reality, augmented reality, human-computer interaction and visualization.

Lili Wang received her Ph.D. degree from the Beihang University, Beijing, China. She is a professor with the School of Computer Science and Engineering of Beihang University, and a researcher with the State Key Laboratory of Virtual Reality Technology and Systems. Her interests include virtual reality, augmented reality, mixed reality, real-time rendering, and realistic rendering.

Zhikai Wen is now a postgraduate student in the School of Computer Science and Engineering of Beihang University, China. With a passion for virtual reality, Zhikai has demonstrated a strong commitment to academic excellence and professional development.

Xuehuai Shi received his Ph.D. degree from the State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing, China. He is currently a lecturer at the School of Computer Science, Nanjing University of Posts and Telecommunications. His research interests include virtual reality, real-time rendering, and augmented reality.

Xiaolong Liu is currently working toward a PhD degree with the School of Computer Science and Engineering, Beihang University, Beijing, China. His current research focuses on virtual reality, augmented reality, and visualization.

Yanzhou Chen is currently an undergraduate student at Beihang university, pursuing a Computer Science and Engineering degree. With a passion for Virtual Reality, Yanzhou has demonstrated a strong commitment to academic excellence and professional development.

Qianwen Wang is currently an undergraduate student at Beihang university, pursuing a Computer Science and Engineering degree. With a passion for Virtual Reality, Qianwen has demonstrated a strong commitment to academic excellence and professional development.

Bin Hu is currently an undergraduate student at Beihang university, pursuing a Computer Science and Engineering degree. With a passion for Virtual Reality, Bin has demonstrated a strong commitment to academic excellence and professional development.