Avatar360: Emulating 6-DoF Perception in 360° Panoramas through Avatar-Assisted Navigation

Andrew Chalmers* Faisal Zaman† Taehyun Rhee‡

Computational Media Innovation Centre
Victoria University of Wellington

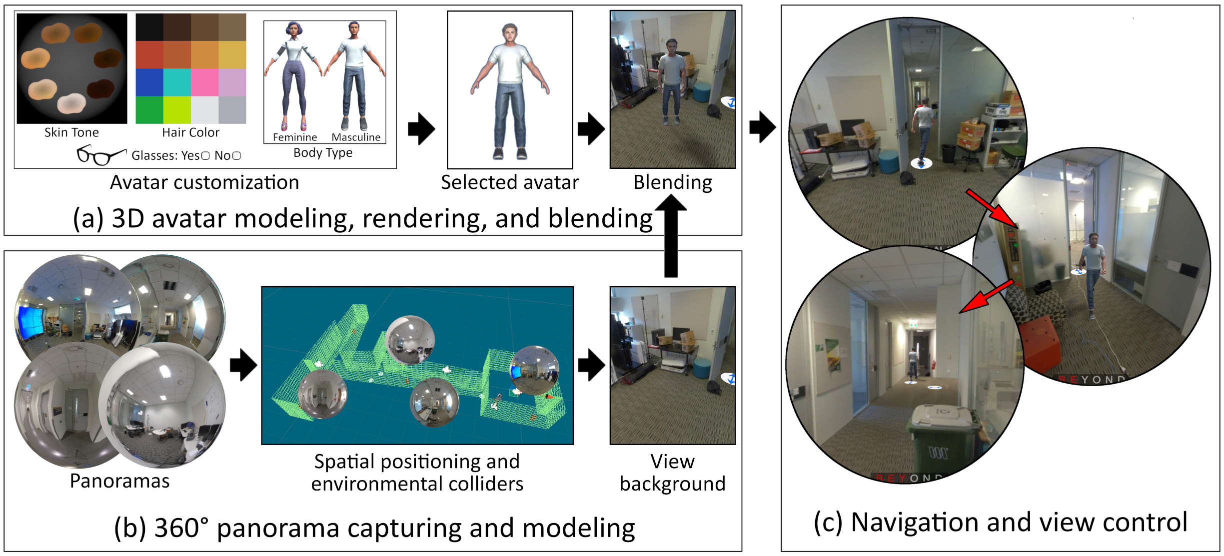


Figure 1: Avatar360 system overview. A user selects their (a) 3D avatar, which is then blended into the background obtained from (b) a collection of spatially connected panoramas. The user can (c) navigate the panoramas using the 3D avatar, providing them with a sense of 6-DoF movement within the environment.

ABSTRACT

360° images offer panoramic views of captured environments, placing users within an egocentric perspective. While users can freely rotate their viewpoint, they don't experience 6-DoF navigation with translational movement. In this research, we introduce Avatar360, a novel method to elicit 6-DoF perception in 360° panoramas, using avatar-assisted navigation combined with an exocentric view of the 360° panorama. We seamlessly integrate a 3D avatar into 360° panoramas, allowing users to navigate a 3D virtual landscape congruent with the 360° background. By aligning the exocentric perspective of the 360° panorama with the avatar's movements, we replicate a sensation of 6-DoF navigation in 360° panoramas. We explore mechanisms for simultaneous avatar and viewpoint controls, as well as procedures for transitions between spatially connected 360° panoramas. A user study was conducted to assess the perception of 6-DoF navigation in 360° panoramas via a 3D avatar, evaluating users' sense of movement, disorientation, and presence. We also gained insight into perspective view controls and transition techniques between panoramas. Statistical analysis shows avatar-assisted navigation elicits a user's sense of movement within 360° panoramas. Our results also provide guidelines for effective view control and transition strategies in avatar-assisted 360° navigation.

*e-mail: andrew.chalmers@vuw.ac.nz
†e-mail: faisal.zaman@vuw.ac.nz
‡e-mail: taehyun.rhee@vuw.ac.nz

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality;

1 INTRODUCTION

360° panoramas offer an expansive view into captured environments, placing users firmly within an egocentric perspective. With a 3-DoF rotation at the fixation of the captured panoramic scene, they afford users a vast field of view. Their versatility has seen application in diverse fields, from indoor and outdoor settings, room and street views, to immersive films, among others.

However, traditional 360° panoramas are intrinsically two-dimensional, lacking depth information. As a result, they do not offer translational movement to move away from the panorama's centre, precluding 6-DoF navigation through the scene. This inherent limitation diminishes immersion in the captured environment. This is particularly important in applications that require a sense of movement and interaction within the environment to maintain users' immersion. These applications include, but are not limited to, navigation at tourist sites, live events, and mixed-reality social platforms where multiple users engage with one another through navigation.

Several approaches have emerged to counter this limitation. One strategy involves recording videos while physically moving the camera [25,29]. Yet, this method deprives users of movement autonomy, and can also lead to vection-induced simulator sickness [5,20,46]. Another category of solutions involves teleportation between multiple, spatially connected 360° panoramas. Users can click to transition between these views, often augmented with an animated zoom effect, to elicit the sensation of movement [2,45,57]. Still, within each panorama, the viewing experience remains restricted to rotations only. Full 3D capturing and scene reconstruction offers a solution by enabling full 6-DoF navigation. However, the 3D capturing process is complex, requiring specialized equipment or sophisticated scene reconstruction algorithms, and the visual artifacts from the current algorithms can compromise the user's immersion [18,42,55]. Recent image-based modeling research in 6-DoF video presents an alternative solution, but challenges remain regarding visual fidelity [7,52], movement range [4,10], requiring sophisticated algorithms, and the need for advanced hardware [33,37], still making widespread application difficult.

In this paper, we present *'Avatar360'*, which emulates 6-DoF perception in 360° panoramas using avatar-assisted navigation providing an exocentric view of 360° panoramas. Our approach sidesteps the need for specialized hardware or intricate algorithms. Our methodology includes 360° panoramic capturing and modeling, 3D avatar modeling and blending, as well as navigation in the 360 panoramas (Figure 1). Moreover, we investigate strategies for view control and transitions between multiple spatially connected 360° panoramas. We conducted a user study to evaluate Avatar360, comparing it to a baseline 360° viewer that does not use an avatar. We measure sense of movement, disorientation, presence, and preference to gauge the effectiveness of Avatar360 in eliciting the sensation of 6-DoF movement. In addition, we also measure three view control styles (coupled, decoupled, static) and view transitions (cut, fade, zoom) between multiple 360° panoramas to gain an insight on how these techniques work with our novel system. Our statistical analysis results from the user study suggest that avatar-assisted navigation can convincingly emulate a user's sensation of movement within 360° panoramas. Users also displayed a strong preference for using an avatar. We found that presence and disorientation were not negatively impacted compared with the baseline egocentric method. The automatic view control that was coupled with the avatar's movements scored higher than the manually controlled decoupled and static view methods across sense of movement and disorientation measures. All three view transition styles were positive across measures, with the zoom transition scoring slightly higher.

The main contributions of this research are:

- **Avatar-Assisted Navigation in 360° Panoramas:** We present Avatar360, a novel method to emulate 6-DoF perception in standard 360° panoramas by integrating a user-controlled 3D avatar. This approach enriches the viewing experience without requiring specialized hardware or complex algorithms.

- **Exocentric Perspective Transformation:** Shifting from the conventional egocentric viewpoint of 360° panoramas, our exocentric perspective, coupled with the avatar's movements, creates a 6-DoF sensation. User studies showed improvements in users' sense of movement within the captured environment when using an avatar.

- **Exploring View Control and Transition Techniques:** Our findings provide insights into perspective view control and view transition styles, offering guidance for optimizing avatar-assisted 360° navigational experiences.

2 Related Work

2.1 View and Navigation of 360° Panoramas

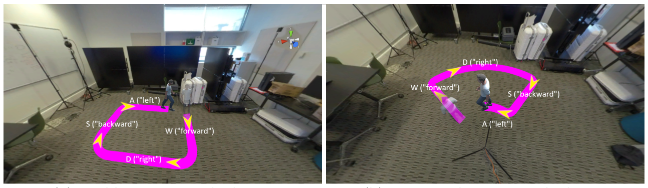
Viewing 360° panoramas of captured real-world locations began with using a desktop monitor and a button press to pan the viewing direction in discrete increments [25,29]. This was later extended to using a mouse to click and drag for a smooth view rotation [8,27,51]. More immersive forms of panoramic viewing include stereo panoramas which introduce a sense of depth [16, 17, 35, 36], as well as viewing tied to head rotations using a head-mounted display [31], which has also been used for viewing live-streaming 360° video [32]. Viewing of panoramic content has since increased in popularity across different device modalities using systems such as Google Street View [2]. While the ways we have viewed and interacted with 360° panoramas have evolved, they still remain from an egocentric perspective, providing only view rotation. We deviate from this paradigm by introducing an avatar to navigate the panorama, providing a sense of 6-DoF to the user.

Multiple 360° panoramas can be spatially connected and navigated between to provide the user with a sense of spatial placement in the environment. Early work involved teleporting between the connected panoramas [8,43], which has been recently adapted to experiencing live events such as a concert from different positions [53]. In 360° film, this teleporting effect is often used in place of a *'cut'* or *'fade to black'* to either move through space or time in the film [26]. Further considerations to reduce disorientation or viewer discomfort from the transition have also been explored [6,23,30,38]. Transition effects between the spatially connected panoramas can introduce a sense of movement in the captured environment. This can come in the form of a video sequence being played between the two panoramas [29], or more recently, an animated image-based interpolation to transition from one panorama to another [2,9,45,57]. These methods allow the user to observe the scene transition process, maintaining their sense of place in the environment while also introducing a sense of movement. However, the sense of movement is only tied to transition events between panoramas, rather than feeling a sense of movement within a given panorama.

There are a number of works to obtain 6-DoF navigation in 360° panoramas. One approach is to create a 3D reconstruction of the environment, however, this often requires specialized hardware setups such as active sensors (LiDAR) mounted onto the 360° camera [3,15,56] or obtaining depth from stereo [21]. These approaches typically represent the captured environment as a textured 3D mesh, where time spent capturing and sensor noise can impact the visual quality of the captured environment. Alternatively, simplified meshes can be generated from either user input or depth from stereo panoramas [19,22,48]. Image-based techniques are another solution for 6-DoF navigation in 360° panoramas, directly rendering a novel view image as opposed to generating a 3D mesh. This can also be obtained through complex hardware setups and sophisticated algorithms such as panoramic light fields [33] or panoramic depth from stereo cameras [37]. Without such hardware, there are depth image-based rendering (DIBR) [50] and neural radiance fields (NeRFs) [28] which have been adapted to 360° panoramas [7,11,52] to produce 6-DoF navigation. However, the visual quality of such processes is still challenging, often requiring a dense set of photographed captures for plausible results. Alternate methods assume a smaller area to capture to improve visual quality [4,10], but this will restrict the user's range of movement. Our avatar-assisted navigation setup circumvents these issues by using simple 360° capturing setups while still providing a sense of 6-DoF navigation.

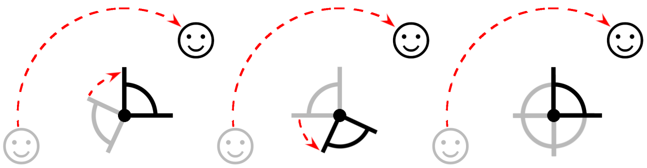
2.2 Avatar Navigation and Blending

Avatars facilitate an embodied navigation experience for users, enabling exploration of the environment from an exocentric perspective. The third-person perspective typically provides a better view to navigate the environment, while the first-person perspective improves immersion [12], though studies have also shown no difference in spatial presence between the two perspectives [14]. Fribourg et al. [13] conducted a study showing the importance of avatar appearance personalization, avatar navigation control, and point of view, all impacting the user's sense of embodiment with the avatar. However, the prior work has not studied the effects of avatar-assisted navigation in 360° panoramas. One obstacle prohibiting such exploration is that blending the avatar into the panorama in a convincing way can be challenging. A separate line of research has shown the blending



(a) World space control scheme (b) Camera relative control scheme

Figure 2: (a) A typical world space control scheme, and (b) our radial camera relative control scheme. The pink line and yellow arrows represent the movement of the avatar. The controls are more intuitive for the user in (b), where for example, from our point of view 'right' goes left in (a), whereas 'right' goes right in (b). We have raised the view angle for illustration purposes.



(a) Coupled with Avatar (b) Decoupled from Avatar (c) Static View

Figure 3: The three perspective view controls (coupled, decoupled, static) in our experiment. The face represents the avatar's position, the angle symbol represents the view position and direction it is facing within the 360° environment. The red dashed arrow indicates the movement of the avatar and view direction. (a) The view follows the avatar, (b) the view is able to go in a different direction from the avatar, specified by the user, and (c) there are fixed view directions that are automatically activated when the avatar is within view.

of digital content into 360° panoramas [40], which has been used to enhance remote collaboration scenarios [41] and live events [53].

While these studies have shown navigating virtual environments with an avatar, and blending virtual content into 360° panoramas, none have explored them together based on our survey.

3 AVATAR360 SYSTEM

In a typical 360° viewer, there are established norms of how a user is able to control their view from an egocentric perspective (e.g., using a mouse or head-mounted display to rotate the view), and how to transition between multiple spatially connected panoramas (e.g., clicking an arrow). Integrating an avatar into 360° environments is a novel concept, and as such, we have outlined the key design features and implementation details behind our experiment. This includes 1) 360° panorama capturing and modeling, 2) 3D avatar modeling, rendering, and blending, and 3) navigation and view control. See Figure 1 for an overview of the Avatar360 system.

3.1 360° Panorama Capturing and Modeling

We use multiple 360° cameras to build up an environment for the avatar to navigate and move between. When transitioning between a pair of panoramas, we position the 360° cameras such that their locations are observable from one another where possible. This will allow users to build a mental model of their spatial location among multiple panoramas. We ensure precise spatial positioning by placing the virtual cameras in the same locations as their physical counterparts. This guarantees that virtual content is rendered from the correct camera perspective. To further enhance realism, 3D plane colliders are positioned around the environment (walls, floor, and planar surfaces), preventing avatars from walking through solid objects. While the placement of the virtual cameras and 3D colliders can be done manually, authoring tools that utilize depth measurements (e.g., 3D scanning or GPS) or depth estimates (see [1] for an overview) can be provided to speed up the process. In our work, we utilize LiDAR depth sensors [1], which guides our placement of the virtual cameras and 3D colliders.

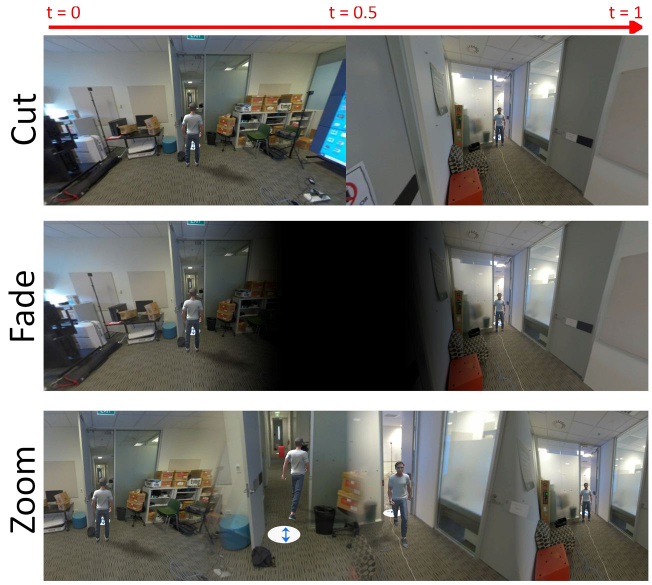


Figure 4: The three view transitions (cut, fade, zoom) in our study, moving from one panorama (left) to another (right) over time t.

In our experiment, we capture the environment with a 360° camera (Insta360 Pro 2), mounted on a tripod at approximately eye level (∼170cm). Placement of the camera is either at the centre, side, or corner of the room for either 360°, 180°, or 45° viewing/navigation respectively.

During runtime, the viewing direction of the virtual camera will create a perspective image (with 80° field of view) from the 360° panorama, which will act as the background and exocentric view that the avatar (and other virtual content) is blended into. See Figure 1b for an overview.

3.2 3D Avatar Modeling, Rendering, and Blending

In our method, the 3D avatar acts as a virtual embodiment of the user within the 360° environment, fostering a heightened sense of movement. To further enhance this immersive experience, users can select their avatar's gender, ethnicity, hair color, and whether they wear glasses. Successfully integrating the avatar into the 360° panorama requires a holistic approach, rather than merely overlaying the avatar onto the background. We ensure the seamless blending of the avatar into the background by considering many facets of avatar design and deployment:

- *Appearance:* We opted for a 3D avatar over 2D to achieve a depth and immersion consistent with 6-DoF navigation into the captured environment. Balancing between realism and artistic representation, our avatars possess a semi-realistic 3D character, ensuring user relatability while avoiding the uncanny valley [13].
- *Modeling:* We use detailed 3D geometric models with skeletal rigging for realistic animations. By leveraging shaders and high-resolution textures, we ensure the avatar looks natural with the lighting conditions of the 360° panorama.
- *Animation:* Natural avatar movement plays an important role in immersion [13]. As such, we provide the avatar with a

[1] https://ouster.com/

natural walk cycle animation to enrich the sense of movement within the environment.

- *Lighting and Rendering:* To achieve congruent shading between the avatar and the panorama's background objects, we utilize the 360° panorama as a light source. This is done using real-time image-based lighting rendering techniques [40].

- *Shadows and Blending:* The avatar's integration hinges on the blending quality, ensuring it feels like an organic component of the panorama, where accurate shadow casting is vital. By analyzing the 360° panorama, we deduce primary light sources (light direction and area) from the brightest pixels in the panorama, and generate lights that produce soft shadows that are composited onto the panorama in real time.

Given the avatar's role in our system to emulate the user's navigation, these careful design and integration choices are pivotal for our effectiveness. See Figure 1a for an overview.

3.3 Navigation and View Control

Avatars serve as a medium for navigating virtual environments, offering users a means to explore and interact with their surroundings from an exocentric perspective. Our aim is to ensure fluidity and alignment with user expectations by employing a radial camera-relative control approach. Considering usability and simplicity, we leverage standard controls: the 'WASD' keys on the keyboard for avatar movement and the mouse for adjusting the view rotation. Specifically, forward and backward movements drive the avatar away and towards the camera, respectively, while lateral movements guide the avatar in an arc around the camera, ensuring an intuitive and predictable navigation experience (refer to Figure 2 for the control scheme). Such an approach mitigates confusion and fosters an enriched and embodied exploration of the 360° panorama.

Avatar-assisted viewpoint manipulation in 360° panoramas remains unexplored based on our literature survey. Our study addresses this by examining various perspective control modes described as follows (see Figure 3):

- **View Coupled with Avatar:** Users move the avatar using the keyboard while the viewpoint auto-aligns with the 3D avatar, with a slight smoothing delay, thereby allowing users to concentrate on the avatar's movement without concerning themselves with controlling the view rotation.

- **View Decoupled from Avatar:** Users move the avatar using the keyboard while using their mouse (click and drag) to rotate their view. This mode offers independent control over their perspective, decoupled from the avatar.

- **Static View:** Users move the avatar using the keyboard while the view remains fixed. When the avatar moves outside the view frustum, the view automatically snaps to another angle that encompasses the avatar's new position. The static mode offers a similar feel to a fixed surveillance camera.

Transitions between spatially connected panoramas are essential for maintaining a sense of place in a large environment. We refer to the points where transitions can occur as 'transition points', and are indicated with 3D virtual door mats which are composited into the 360° panorama at doorways. The transition begins when the mat is stepped on by the avatar, in which user input is paused, and the avatar will automatically start walking over the mat and into the next panorama over time t (mapped from 0 to 1 second). Given two 360° panoramas (A and B), over transition time t, we evaluate the following three types of transitions (see Figure 4):

- **Cut:** The 360° panorama instantly switches from panorama A to panorama B at $t = 0.5$.

- **Fade:** The 360° panorama gradually fades from panorama A to black and then gradually from black to panorama B. The fade occurs linearly over t.

- **Zoom:** The view in the 360° panorama zooms in panorama A toward the transition point (i.e., 3D virtual door mat), the zoom is maximized at $t = 0.5$, then the panorama switches to panorama B, and then the view zooms back out until $t = 1.0$.

The 3D virtual door mats have a two-way arrow symbol indicating that the trigger can work both ways (going from A to B and B to A). To minimize disorientation, the camera will always rotate to face the avatar during the transition.

4 User Study

4.1 Study Design and Conditions

Our primary research question is whether avatar-assisted navigation, combined with an exocentric perspective of the panorama, can elicit the sensation of 6-DoF navigation. There are a number of sub-questions for the implementation details regarding how to control the view and provide natural transitions between multiple panoramas for navigating a large space. For these questions, we set three independent variables: 1) *avatar* (off, on), 2) *view control* (coupled, decoupled, static), and 3) *view transition* (zoom, cut, fade). The avatar has two levels (either avatar is used, or not), while both the view control and view transition have three levels each, comparing against the methods described in the implementation section.

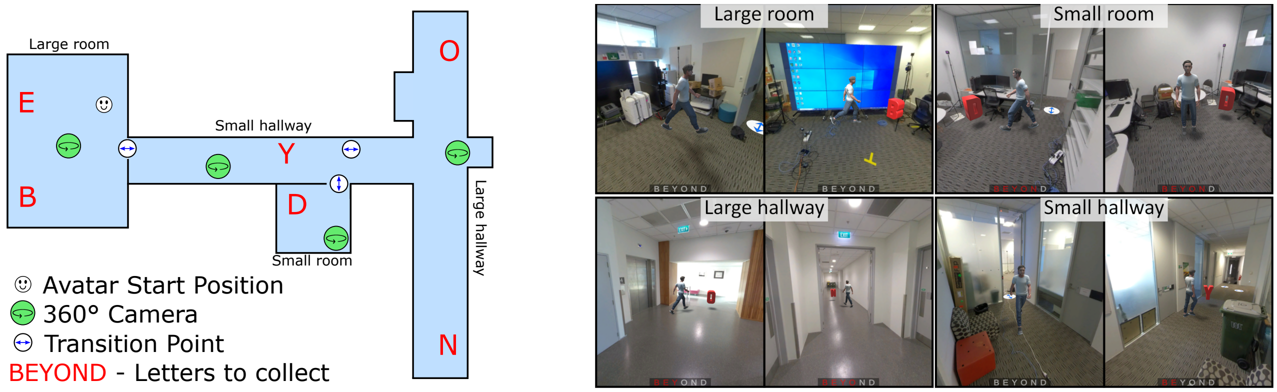
We conducted a within-subjects $2 \times 3 \times 3$ mixed factorial design with post hoc tests comparing groups of conditions. We focused on six out of 18 conditions. We compared against one baseline condition with no avatar using one combination of view control and view transition, so this eliminated eight of the 18 conditions. Furthermore, user input was paused during the view transition, so view transition did not need to be paired with each view control style. This eliminated a further four of the 18 conditions. This resulted in six conditions that we focused on in our study, shown in Table 1.

Table 1: Conditions in the user study.

	Avatar	View Control	View Transition
C1	Off	Decoupled	Zoom
C2	On	Coupled	Zoom
C3	On	Decoupled	Zoom
C4	On	Static	Zoom
C5	On	Coupled	Cut
C6	On	Coupled	Fade

Condition C1 represents the baseline typical 360° viewer experience, where there is no avatar, they are able to freely control the view using their mouse (click and drag), and the transitions are zoomed, occurring using a mouse click on the transition points. After a transition, the view does not rotate back towards the avatar (as there isn't one) and continues facing in the same direction. This was chosen as it is similar to common panorama viewers which offer some notion of movement (such as Google Street View [2]). This makes a widely adopted method as the baseline to evaluate whether Avatar360 can enhance the sense of movement beyond conventional methods.

The following five conditions (C2 through C6) represent our proposed avatar-assisted navigation approach with varying view control and transition variables. C3 and C4 represent variations of view control compared with C2, while C5 and C6 represent variations of view transitions compared with C2. Note that for C3, the view control is decoupled similar to C1, but unlike C1, view transitions are triggered when an avatar walks onto the mat as opposed to a mouse click.



(a) (b)

Figure 5: (a) The layout of the user study task and (b) two sample views from each room, where the avatar is completing the task.

4.2 Task Design

The participants engaged in a task-oriented finding game which encouraged them to navigate the environment. The environment is captured by four spatially connected panoramas in an indoor setting with variations in the size of the rooms to encourage both 360° viewing and diversity in navigation ('Large room', 'Small room', 'Small hallway', and 'Large hallway'). We are not particularly interested in how well participants completed the task, but instead in garnering insights into the participant's experience during the task. As such, the task itself is a simple game where six letters ('B', 'E', 'Y', 'O', 'N', 'D') are scattered throughout the environment in predetermined locations. The participants navigated the environment with their avatar, collecting all the letters in any order. Once all the letters are collected, the task is complete, before reloading the task again under another condition.

A map of the task, including the letter locations and transition points, is shown in Figure 5. The 'Large room' encouraged 360° viewing, the 'Large hallway' and 'Small hallway' encouraged 180° viewing, and the 'Small room' encouraged 45° viewing. The participant starts in 'Large room' and could freely navigate back and forth between rooms to find the letters.

4.3 Measures

We measure: 1) *Sense of movement*, 2) *Disorientation*, and 3) *Spatial presence*, with a total of eight questions asked for each of the six conditions. We also measure 4) *Preference*, which was asked at the end of the study after the participant had experienced all the conditions. Finally, we logged 5) *Task completion time*.

For the sense of movement, we used a modified version of an Avatar Embodiment questionnaire [34]. For disorientation, we used three questions based on disorientation questions used in 360° film [23]. To measure the spatial presence, we used a subset of the *iGroup Presence Questionnaire* (IPQ [44]). We used four questions relating to sub-scales general presence (GP), spatial presence (SP), and realism (RL). Since the virtual environment in our study involved real-world imagery, we paraphrased the IPQ questions from 'virtual environment' to 'captured environment' to avoid confusion. These questions are all answered on a 7-point Likert scale, where higher is better, shown in Table 2.

To measure preference, we asked the participants which they preferred for each independent variable (avatar, view control, view transition) with a ranked choice, as shown in Table 3. This is asked at the end of the study once the participant has experienced all the conditions. To support the participants' memory, we showed a short video of each effect while they answered the questions.

Figure 6: Participants in the user study controlling an avatar using a keyboard and mouse. Screen configurations include desktop (top left), laptop (bottom left), and wall-mounted TV (right).

Table 2: 7-point Likert scale questionnaire between conditions.

Sense of Movement	
1	It seemed as if I was moving in the captured environment
Disorientation	
2	I had no trouble orienting myself in the scene
3	I felt I had control over where I was looking
4	I had no problem finding what I wanted to look at
Spatial Presence	
5	In the captured environment, I had a sense of "being there"
6	Somehow I felt that the captured environment surrounded me
7	I had a sense of acting in the captured environment, rather than operating something from outside
8	How real did the captured environment seem to you?

Table 3: Ranked preference questionnaire at the end of the study.

	Preference Questionnaire Item
1	Do you prefer seeing the avatar or not seeing the avatar?
2	Rank which view control style you prefer most
3	Rank which view transition style you prefer most

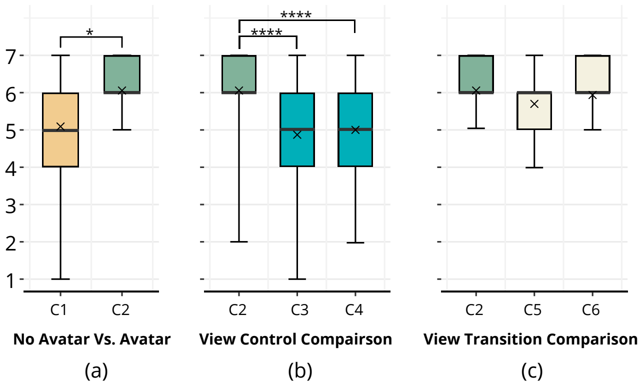


Figure 7: Results of Sense of Movement under each condition. = C1, = C2, = C3/C4, = C5/C6

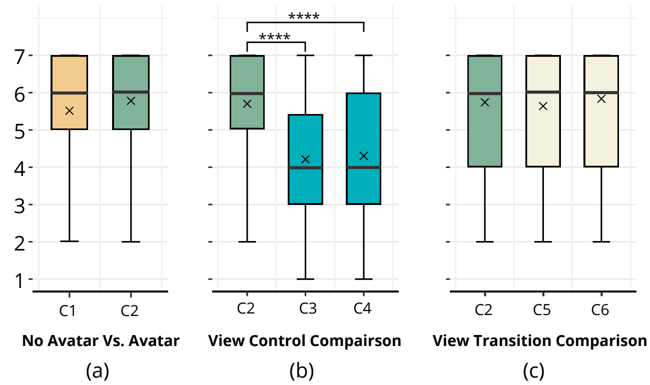


Figure 8: Results of Disorientation under each condition.

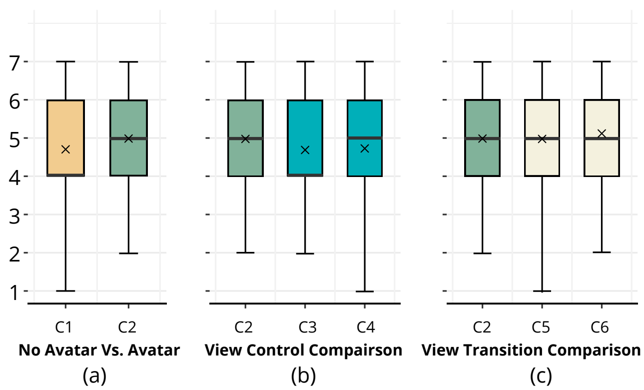


Figure 9: Results of IPQ under each condition.

4.4 Procedure

Before beginning the study, participants read an information sheet on the study, signed an ethics form, and filled in a demographic questionnaire (asking for age group, ethnicity, gender, video game experience, and normal or corrected-to-normal vision). They were then given the task description before selecting their avatar. The participants briefly familiarized themselves with both avatar and camera controls. Then, the study begins, where the participants attempt the task six times, once for each of the six conditions, with the order randomized. The participants are given an intermediate questionnaire to fill in after completing the task for a given condition. After completing all six conditions, they answered a post-experiment questionnaire regarding their preferences for each of the three independent variables. We also provided an open feedback form at the end of the study.

4.5 Participants

A total of 20 participants were recruited for this study, with ages ranging from 18 to 49 years (M = 28.28, SD = 8.06). Of the participants, 5 identified as female, and 15 identified as male. Most participants reported to have spent a high amount of time on digital media (M = 7.58, SD = 5.39 hours per week). Participants were recruited on campus, and all of them reported having normal or corrected-to-normal vision.

4.6 Hardware Setup

We conducted the study on three different computer setups: 1) a desktop monitor with a 24-inch display, 2) a wall-mounted 64-inch display with a 2-metre viewing distance, and 3) a laptop with a 13-inch display. The Avatar360 system was implemented in Unity (version 2021.3.1f1) running at beyond 120 frames per second on all hardware setups. Participants used a keyboard and mouse for the task (Figure 6).

5 Results

We conducted a post hoc test on the six conditions from the user study. To show whether Avatar360 can elicit the sensation of 6-DoF navigation, we compare the baseline C1 against our Avatar360 method C2. To evaluate view control, we compare between C2 (coupled), C3 (decoupled), and C4 (static). To evaluate view transitions, we compare between C2 (zoom), C5 (cut), and C6 (fade). We present the measures (sense of movement, disorientation, presence, preference, and task completion time) based on the different conditions.

5.1 Avatar-assisted navigation

We visualize the results between the baseline (C1) and our method (C2) for sense of movement in Figure 7a, disorientation in Figure 8a, and presence in Figure 9a. The mean, standard deviation, and one-way ANOVA results are shown in Table 4. The results show that C2 has an increased sense of movement with statistical significance, while both C1 and C2 had no statistically significant difference in disorientation and presence, both with positive scores. For preference, C2 was strongly preferred over C1 (shown in Figure 10).

Table 4: Results for avatar-assisted navigation (*p values in bold indicate statistical significance*).

	C1		C2		F	p
	μ	σ	μ	σ		
Movement	5.10	1.448	6.05	0.826	5.288	**0.028**
Disorientation	5.61	1.278	5.74	1.467	0.282	0.596
Presence	4.63	1.477	4.98	1.376	1.758	0.186

5.2 View Control

We visualize the view control results (C2 - coupled, C3 - decoupled, C4 - static) for sense of movement in Figure 7b, disorientation in Figure 8b, and presence in Figure 9b. The mean, standard deviation, and ANOVA results are shown in Table 5. The results show that a coupled view has a higher score across all three measures, with statistical significance for sense of movement and disorientation. For preference, participants unanimously preferred the coupled view over decoupled and static, while decoupled was slightly preferred over static.

Table 5: Results for view control (*p values in bold indicate statistical significance*).

	C2		C3		C4		F	*p*
	μ	σ	μ	σ	μ	σ		
Movement	6.05	0.826	4.85	1.352	5.00	1.337	8.87	**<0.001**
Disorientation	5.74	1.467	4.20	1.733	4.30	1.720	19.78	**<0.001**
Presence	4.98	1.376	4.61	1.201	4.67	1.385	2.57	0.078

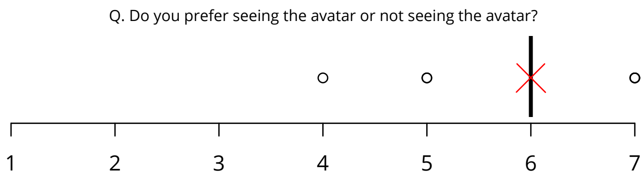


Figure 10: Results of avatar-assisted navigation preference (1: Strongly prefer not seeing an avatar, 4: No preference, 7: Strongly prefer seeing an avatar).

5.3 View transition

We visualize the view transition results for sense of movement (C2 - zoom, C5 - cut, C6 - fade) in Figure 7c, disorientation in Figure 8c, and presence in Figure 9c. The corresponding mean, standard deviation, and ANOVA results can be found in Table 6. The results showed no statistically significant difference between each method, though the sense of movement standard deviation for zoom is much narrower than cut and fade, where zoom had a slightly higher mean. The preferences showed that fade was slightly preferred over cut, and cut was slightly preferred over zoom.

Table 6: Results for view transition (*p values in bold indicate statistical significance*).

	C2		C5		C6		F	*p*
	μ	σ	μ	σ	μ	σ		
Movement	6.05	0.826	5.66	1.110	5.95	1.132	0.520	0.597
Disorientation	5.74	1.467	5.66	1.436	5.84	1.327	0.353	0.702
Presence	4.98	1.376	5.01	1.366	5.14	1.458	1.164	0.313

The tasks were completed in approximately 58.40 seconds on average (SD = 36.50) with no statistical difference between conditions, except C3, which approximately took twice as long (M = 97.40 seconds, SD = 51.63) with statistical significance. We also found no difference in scores regarding the hardware setup (desktop, laptop, TV). For statistical analysis of the data, we initially performed the Shapiro-Wilk test to assess whether the data follows a normal distribution. For non-normally distributed data, we applied the Aligned-rank transform. Holm-Bonferroni correction was used for all post hoc tests.

6 Discussion

Through our user study, we assessed the effectiveness of avatar-assisted navigation using our Avatar360 system, comparing it with a traditional 360° panorama viewer, with a particular emphasis on heightened movement perception. Furthermore, we examined the implications of various view and transition strategies within the Avatar360 framework.

6.1 Impact of avatar-assisted navigation

The results support our hypothesis that avatar-assisted navigation in 360° panoramas improves participants' sense of movement with statistical significance (C2) over the baseline without an avatar (C1) (Figure 7a). This was further supported in the participants' preference selection, where using an avatar was preferred unanimously among all participants (M = 6, SD = 0.72), with the large majority indicating a strong preference as shown in Figure 10. We also found that adding an avatar did not cause disorientation and was similar to the baseline (Figure 8a). This indicates that introducing an avatar did not negatively impact the experience, maintaining low disorientation.

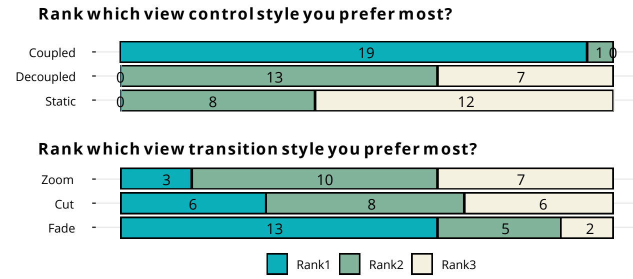


Figure 11: Ranking results for three camera view control methods (*coupled*, *decoupled*, *static*) and three camera transition methods (*zoom*, *cut*, *fade*). Rank 1 means most preferred.

In prior research [12], it was reported that the exocentric perspective tends to have lower immersion than the egocentric perspective (including non-VR-based applications). Interestingly, our IPQ (presence) results still managed to show no difference between our exocentric method with an avatar compared to the egocentric baseline without the avatar, both maintaining positive scores (Figure 9a).

6.2 Impact of view control and transition effects

Our second hypothesis was that some view controls and transition styles would be prioritized over others. We found that the coupled view control (C2) had much higher scores than decoupled (C3) and static view (C4) methods across sense of movement (Figure 7b) and disorientation (Figure 8b) with statistical significance, in addition to user preference (Figure 11, top). We also found that the task completion time was much higher for decoupled view control. This aligns with prior research [24], which found that an automatic camera is more effective than manual adjustments. No clear statistical difference was found regarding presence (Figure 9b), though the static view (C4) tended to skew negatively.

All three view transition styles were positive across measures (Figures 7c, 8c, 9c), with the zoom transition (C2) scoring slightly higher in sense of movement and with a much narrow standard deviation, while fade scored slightly higher in presence, though there was no statistical significance found. User preference was also mixed across styles, with most preferring fade (Figure 11, bottom). It was noted that participants often did not realize there was a difference in transitions during the study until it was shown to them during the preference questionnaire. The results suggest that all transition styles can be considered viable options across different applications.

6.3 Applications

Avatar360 offers a unique way to navigate 360° panoramas using avatars, making it suitable for a wide range of current and emerging applications involving the viewing of 360° images and videos. This technology can enhance existing 360° images, such as street views [2] and panoramic indoor views for accommodation and open homes [49]. Avatar navigation adds a new dimension to user presence by allowing them to explore these environments through an embodied avatar experience.

Furthermore, Avatar360 can be leveraged to create immersive viewing experiences for live events [39], mixed reality concerts [53], sports events, and exhibitions. Future work could involve applying these approaches to various applications, conducting case studies, and evaluating their impact.

6.4 Limitations and Future Work

Our Avatar360 framework is the first to seamlessly integrate avatar navigation into 360° panoramas. In this paper, we introduce this novel concept and thoroughly evaluate it, considering key aspects such as navigation, view control, and view transitions. Our evaluation primarily focuses on indoor environments with varying sizes (small and large) and shapes (rooms and hallways). In the future, we plan to extend this framework to outdoor environments, incorporating more complex features such as obstacles, intricate terrains, varying levels with stairs, and more. The view control and view transition styles selected in our study could be further expanded, such as more creative transition effects or visually guided transitions [47]. We used a single avatar in this research, but multiple avatar interaction would be an interesting next step to consider [54]. In addition, we did not study the avatar's appearance, which could be an interesting direction. We focused on using a desktop setup that is common for panoramic viewing, however, another common use case is using a head-mounted display. Future work could explore alternate hardware modalities to interact with the panoramas, which could show variations in measures, such as presence. In our study, we experimented with four spatially connected panoramas. Future work could explore transitions between spatially disconnected panoramas, as well as panoramic video and live-streaming video. The colliders we made for the environments were set up manually, but it can be improved by utilizing a depth-capturing setup.

7 Conclusion

We introduced Avatar360, a novel approach that emulates 6-DoF navigation in 360° panoramas using 3D avatars. This exocentric perspective-driven method offers a 6-DoF navigational experience within 360° panoramas without the need for specialized hardware or sophisticated algorithms.

We developed a working prototype of the Avatar360 system and evaluated its impact through a user study. Our results indicate that avatar-assisted navigation effectively elicits 6-DoF perception within 360° panoramas, with users showing a strong preference for this method over conventional 360° viewing techniques. Moreover, our user study investigated best practices for our innovative setup in terms of view control (coupled, decoupled, static) and transitions (cut, fade, zoom) between spatially connected panoramas. The results highlight that a view synchronized with the avatar's movements outperformed both decoupled and static view methods. Additionally, all view transitions received positive feedback, suggesting their applicability across diverse applications.

Acknowledgments

This research was supported by the Entrepreneurial University programme from the Tertiary Education Commission of New Zealand, and partially supported by the Brain Pool Program funded by the Ministry of Science and ICT through the National Research Foundation of Korea (RS-2023-00223105).

References

[1] G. Albanis, N. Zioulis, P. Drakoulis, V. Gkitsas, V. Sterzentsenko, F. Alvarez, D. Zarpalas, and P. Daras. Pano3d: A holistic benchmark and a solid baseline for 360deg depth estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3727–3737, 2021.

[2] D. Anguelov, C. Dulong, D. Filip, C. Frueh, S. Lafon, R. Lyon, A. Ogale, L. Vincent, and J. Weaver. Google street view: Capturing the world at street level. *Computer*, 43(6):32–38, 2010.

[3] T. Asai, M. Kanbara, and N. Yokoya. 3d modeling of outdoor environments by integrating omnidirectional range and color images. In *Fifth International Conference on 3-D Digital Imaging and Modeling (3DIM'05)*, pp. 447–454. IEEE, 2005.

[4] T. Bertel, M. Yuan, R. Lindroos, and C. Richardt. Omniphotos: Casual 360° vr photography. *ACM Trans. Graph.*, 39(6), nov 2020. doi: 10.1145/3414685.3417770

[5] F. Bonato, A. Bubka, S. Palmisano, D. Phillip, and G. Moreno. Vection change exacerbates simulator sickness in virtual environments. *Presence: Teleoperators and Virtual Environments*, 17(3):283–292, 2008.

[6] R. Cao, J. Walsh, A. Cunningham, C. Reichherze, S. Dey, and B. Thomas. A preliminary exploration of montage transitions in cinematic virtual reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 65–70. IEEE, 2019.

[7] R. Chen, F.-L. Zhang, S. Finnie, A. Chalmers, and T. Rhee. Casual 6-dof: free-viewpoint panorama using a handheld 360 camera. *IEEE Transactions on Visualization and Computer Graphics*, 2022.

[8] S. E. Chen. Quicktime vr: An image-based approach to virtual environment navigation. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pp. 29–38, 1995.

[9] S. E. Chen and L. Williams. View interpolation for image synthesis. In *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '93, p. 279–288. Association for Computing Machinery, New York, NY, USA, 1993. doi: 10.1145/166117.166153

[10] C. Choi, S. M. Kim, and Y. M. Kim. Balanced spherical grid for egocentric view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16590–16599, 2023.

[11] T. L. da Silveira and C. R. Jung. Dense 3d scene reconstruction from multiple spherical images for 3-dof+ vr applications. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 9–18. IEEE, 2019.

[12] A. Denisova and P. Cairns. First person vs. third person perspective in digital games: do player preferences affect immersion? In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pp. 145–148, 2015.

[13] R. Fribourg, F. Argelaguet, A. Lécuyer, and L. Hoyet. Avatar and sense of embodiment: Studying the relative preference between appearance, control and point of view. *IEEE transactions on visualization and computer graphics*, 26(5):2062–2072, 2020.

[14] G. Gorisse, O. Christmann, E. A. Amato, and S. Richir. First-and third-person perspectives in immersive virtual environments: presence and performance analysis of embodied users. *Frontiers in Robotics and AI*, 4:33, 2017.

[15] C. R. Gunadi, H. Shimizu, K. Kodama, and K. Aizawa. Construction of large-scale virtual environment by fusing range data, texture images, and airborne altimetry data. In *3D Data Processing Visualization and Transmission, International Symposium on*, pp. 772–772. IEEE Computer Society, 2002.

[16] H.-C. Huang and Y.-P. Hung. Panoramic stereo imaging system with automatic disparity warping and seaming. *Graphical Models and Image Processing*, 60(3):196–208, 1998.

[17] H.-C. Huang, Y.-P. Hung, S.-W. Shih, et al. Panoramic stereo imaging with complete-focus views for virtual reality. In *Proceeding of Second Workshop on Real-Time and Media Systems, RAMS*, vol. 96, pp. 405–410, 1996.

[18] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, et al. Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pp. 559–568, 2011.

[19] H.-D. Kang and K.-H. Jo. 3d reconstruction of planar objects using the properties of plane and vanishing points from a single panoramic image. In *The 10th World Multi-Conference on Systemics, Cybernetics and Informatics*, vol. 2, pp. 342–346, 2006.

[20] B. Keshavarz, B. E. Riecke, L. J. Hettinger, and J. L. Campos. Vection and visually induced motion sickness: how are they related? *Frontiers in psychology*, 6:472, 2015.

[21] H. Kim and A. Hilton. 3d scene reconstruction from multiple spherical stereo pairs. *International journal of computer vision*, 104:94–116, 2013.

[22] H. Kim and A. Hilton. Block world reconstruction from spherical stereo image pairs. *Computer Vision and Image Understanding*, 139:104–121, 2015.

[23] T. Kjær, C. B. Lillelund, M. Moth-Poulsen, N. C. Nilsson, R. Nordahl, and S. Serafin. Can you cut it? an exploration of the effects of editing in cinematic virtual reality. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, pp. 1–4, 2017.

[24] T. Kosch, R. Boldt, M. Hoppe, P. Knierim, and M. Funk. Exploring the optimal point of view in third person out-of-body experiences. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '16. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2910674.2910720

[25] A. Lippman. Movie-maps: An application of the optical videodisc to computer graphics. *Acm Siggraph Computer Graphics*, 14(3):32–42, 1980.

[26] C. Marañes, D. Gutierrez, and A. Serrano. Exploring the impact of 360 movie cuts in users' attention. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 73–82. IEEE, 2020.

[27] L. McMillan and G. Bishop. Plenoptic modeling: An image-based rendering system. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '95, p. 39–46. Association for Computing Machinery, New York, NY, USA, 1995. doi: 10.1145/218380.218398

[28] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.

[29] G. Miller, E. Hoffert, S. E. Chen, E. Patterson, D. Blackketter, S. Rubin, S. A. Applin, D. Yim, and J. Hanan. The virtual museum: Interactive 3d navigation of a multimedia database. *The Journal of visualization and computer animation*, 3(3):183–197, 1992.

[30] K. R. Moghadam and E. D. Ragan. Towards understanding scene transition techniques in immersive 360 movies and cinematic experiences. In *2017 IEEE Virtual Reality (VR)*, pp. 375–376. IEEE, 2017.

[31] U. Neumann, T. Pintaric, and A. Rizzo. Immersive panoramic video. In *Proceedings of the eighth ACM international conference on Multimedia*, pp. 493–494, 2000.

[32] Y. Onoe, K. Yamazawa, H. Takemura, and N. Yokoya. Telepresence by real-time view-dependent image generation from omnidirectional video streams. *Computer Vision and Image Understanding*, 71(2):154–165, 1998.

[33] R. S. Overbeck, D. Erickson, D. Evangelakos, M. Pharr, and P. Debevec. A system for acquiring, processing, and rendering panoramic light field stills for virtual reality. *ACM Transactions on Graphics (TOG)*, 37(6):1–15, 2018.

[34] T. C. Peck and M. Gonzalez-Franco. Avatar embodiment. a standardized questionnaire. *Frontiers in Virtual Reality*, 1:575943, 2021.

[35] S. Peleg and M. Ben-Ezra. Stereo panorama with a single camera. In *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, vol. 1, pp. 395–401. IEEE, 1999.

[36] S. Peleg, M. Ben-Ezra, and Y. Pritch. Omnistereo: Panoramic stereo imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(3):279–290, 2001.

[37] A. P. Pozo, M. Toksvig, T. F. Schrager, J. Hsu, U. Mathur, A. Sorkine-Hornung, R. Szeliski, and B. Cabral. An integrated 6dof video camera and system design. *ACM Trans. Graph.*, 38(6), nov 2019. doi: 10.1145/3355089.3356555

[38] 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*, 26(6):2273–2287, 2018.

[39] T. Rhee, A. Chalmers, F. Zaman, A. Stangnes, and V. Roberts. Real-time stage modelling and visual effects for live performances. In *ACM SIGGRAPH 2023 Real-Time Live!*, SIGGRAPH '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3588430.3597245

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

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

[42] C. Richardt, J. Tompkin, and G. Wetzstein. Capture, reconstruction, and representation of the visual real world for virtual reality. In *Real VR–Immersive Digital Reality: How to Import the Real World into Head-Mounted Immersive Displays*, pp. 3–32. Springer, 2020.

[43] G. D. Ripley. Dvi—a digital multimedia technology. *Communications of the ACM*, 32(7):811–822, 1989.

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

[45] H. Segerman. Elevr,"spherical video editing effects with mobius transformations", 2016.

[46] R. H. So, W. Lo, and A. T. Ho. Effects of navigation speed on motion sickness caused by an immersive virtual environment. *Human factors*, 43(3):452–461, 2001.

[47] M. Speicher, C. Rosenberg, D. Degraen, F. Daiber, and A. Krúger. Exploring visual guidance in 360-degree videos. In *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video*, pp. 1–12, 2019.

[48] P. Sturm. A method for 3d reconstruction of piecewise planar objects from single panoramic images. In *Proceedings IEEE Workshop on Omnidirectional Vision (Cat. No. PR00704)*, pp. 119–126. IEEE, 2000.

[49] M. Z. Sulaiman, M. N. A. Aziz, M. H. A. Bakar, N. A. Halili, and M. A. Azuddin. Matterport: Virtual tour as a new marketing approach in real estate business during pandemic covid-19. In *Proceedings of the International Conference of Innovation in Media and Visual Design (IMDES 2020)*, pp. 221–226. Atlantis Press, 2020. doi: 10.2991/assehr.k.201202.079

[50] W. Sun, L. Xu, O. C. Au, S. H. Chui, and C. W. Kwok. An overview of free view-point depth-image-based rendering (dibr). In *APSIPA Annual Summit and Conference*, pp. 1023–1030, 2010.

[51] R. Szeliski and H.-Y. Shum. Creating full view panoramic image mosaics and environment maps. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '97, p. 251–258. ACM Press/Addison-Wesley Publishing Co., USA, 1997. doi: 10.1145/258734.258861

[52] M. Tancik, V. Casser, X. Yan, S. Pradhan, B. Mildenhall, P. P. Srinivasan, J. T. Barron, and H. Kretzschmar. Block-nerf: Scalable large scene neural view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8248–8258, 2022.

[53] J. Young, S. Thompson, H. Downer, B. Allen, N. Pantidi, L. Stoecklein, and T. Rhee. Telefest: Augmented virtual teleportation for live concerts. In *ACM International Conference on Interactive Media Experiences*, pp. 69–78, 2022.

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

[55] J. Zhang, C. Zhu, L. Zheng, and K. Xu. Rosefusion: random optimization for online dense reconstruction under fast camera motion. *ACM Transactions on Graphics (TOG)*, 40(4):1–17, 2021.

[56] H. Zhao and R. Shibasaki. Reconstruction textured urban 3d model by fusing ground-based laser range image and ccd image. In *MVA*, pp. 232–237, 1998.

[57] P. Zhao, Q. Hu, Z. Tang, and M. Ai. A smooth transition algorithm for adjacent panoramic viewpoints using matched delaunay triangular patches. *ISPRS International Journal of Geo-Information*, 9(10):596, 2020.