



 Latest updates: <https://dl.acm.org/doi/10.1145/3635142>

RESEARCH-ARTICLE

# A Virtual Reality Scene Taxonomy: Identifying and Designing Accessible Scene-Viewing Techniques

**RACHEL L. FRANZ**, University of Washington, Seattle, WA, United States



HCI, Accessibility, Aging

**SASA JUNUZOVIC**, Microsoft Research, Redmond, WA, United States

**MARTEZ MOTT**, Microsoft Research, Redmond, WA, United States

**Open Access Support** provided by:

**Microsoft Research**

**University of Washington**



PDF Download  
3635142.pdf  
27 February 2026  
Total Citations: 13  
Total Downloads:  
4544

**Published:** 05 February 2024  
**Online AM:** 13 December 2023  
**Accepted:** 16 April 2023  
**Revised:** 10 October 2022  
**Received:** 22 December 2020

[Citation in BibTeX format](#)

# A Virtual Reality Scene Taxonomy: Identifying and Designing Accessible Scene-Viewing Techniques

RACHEL L. FRANZ, The Information School, University of Washington, Seattle, WA, USA  
SASA JUNUZOVIC and MARTEZ MOTT, Microsoft Research, Redmond, WA, USA

Virtual environments (VEs) afford similar interactions to those in physical environments: individuals can navigate and manipulate objects. Yet, a prerequisite for these interactions is being able to view the environment. Despite the existence of numerous scene-viewing techniques (i.e., interaction techniques that facilitate the visual perception of virtual scenes), there is no guidance to help designers choose which techniques to implement. We propose a scene taxonomy based on the visual structure and task within a VE by drawing on literature from cognitive psychology and computer vision, as well as virtual reality (VR) applications. We demonstrate how the taxonomy can be used by applying it to an accessibility problem, namely limited head mobility. We used the taxonomy to classify existing scene-viewing techniques and generate three new techniques that do not require head movement. In our evaluation of the techniques with 16 participants, we discovered that participants identified tradeoffs in design considerations such as accessibility, realism, and spatial awareness, that would influence whether they would use the new techniques. Our results demonstrate the potential of the scene taxonomy to help designers reason about the relationships between VR interactions, tasks, and environments.

CCS Concepts: • **Human-centered computing** → **Virtual reality**; **Accessibility systems and tools**;

Additional Key Words and Phrases: Virtual reality, virtual scene, virtual environment, taxonomy, scene-viewing technique, accessibility, situational impairment

## ACM Reference format:

Rachel L. Franz, Sasa Junuzovic, and Martez Mott. 2024. A Virtual Reality Scene Taxonomy: Identifying and Designing Accessible Scene-Viewing Techniques. *ACM Trans. Comput.-Hum. Interact.* 31, 2, Article 23 (February 2024), 44 pages.

<https://doi.org/10.1145/3635142>

## 1 INTRODUCTION

Scenes in **virtual reality (VR)** environments are designed to resemble physical environments, and, as a result, considerable effort goes into making them believable. The richness of scenes and their lack of physical constraints has resulted in the creation of various scene-viewing techniques (i.e., interaction techniques that facilitate the visual perception of a virtual scene) over the past few decades [27, 80]. These techniques have ranged from camera control techniques to techniques that modify the visual properties of a scene.

Authors' addresses: R. L. Franz, The Information School, 1851 NE Grant Ln, Seattle, WA 98105; e-mail: franzrac@uw.edu; S. Junuzovic and M. Mott, Microsoft Research, 14820 NE 36th St, Redmond, WA 98052; e-mails: {sasa.junuzovic, martezmott}@microsoft.com.



This work is licensed under a Creative Commons Attribution International 4.0 License.

© 2024 Copyright held by the owner/author(s).

1073-0516/2024/02-ART23

<https://doi.org/10.1145/3635142>

However, despite the abundance of scene-viewing technique research, the question remains: *how would a designer choose a scene-viewing technique to implement?* and more generally, *how would a designer reason about scene-viewing techniques given a particular scene?* Despite the significant number of available scene-viewing techniques, there is no guidance for determining if a technique is suitable for viewing a scene, particularly when the structure of a scene affects how it can be viewed. For example, a scene that has high object density would lend itself to a variety of techniques, such as techniques that make occluding objects transparent, techniques that distort the camera ray to see around occluding objects, and techniques that enable a user to choose a path through the scene that gives them the best view, but a designer would have to sift through the research to identify and compare these techniques.

To facilitate the selection of appropriate scene-viewing techniques, we approached scenes as spaces with affordances that indicate how they should be viewed. Based on this approach, we devised a scene taxonomy derived from insights into cognitive psychology and computer vision on how people and computers define, view, and describe virtual and physical scenes. In addition, to ground the taxonomy in existing implementations of VR environments, we surveyed 29 commercial VR applications to identify common visual properties and tasks.

Choosing an appropriate scene-viewing technique could be particularly important when designers want to make VR applications accessible. Head tracking, the predominant scene-viewing technique used in VR, enables the user to control the virtual camera in a manner that simulates physical reality when a person turns their head to look around [128]. However, situational impairments and disability can negatively impact a person's ability to use head tracking. As a result, these individuals might be unable to view scenes as easily as people with unconstrained mobility, making their experiences less enjoyable or even inaccessible. There are many scene-viewing techniques that require little head movement and could be a solution to this problem, but again, because there are so many techniques, it is difficult to know which to use for a specific context. To address this problem, we applied the scene taxonomy to identify and classify scene-viewing techniques that could be used by people with disabilities or situational impairments that affect their head movement.

In addition to identifying existing scene-viewing techniques that are accessible, we showed how the taxonomy could be used as an ideation tool to design new scene-viewing techniques. We designed and evaluated three scene-viewing techniques with 16 participants experiencing limited head movement. Our findings suggested that although participants thought the new scene-viewing techniques were as easy or easier to use than the default technique implemented in many VR applications, they considered tradeoffs related to accessibility, usability, realism, spatial awareness, comfort, familiarity with the interaction, and task objective when determining their preferred technique. As a result of these tradeoffs, the technique they thought was easier to use was not always the one they preferred.

Our contributions are threefold: (1) A scene taxonomy, informed by cognitive psychology and computer vision literature that can be used to reason about relationships between scene-viewing techniques, scenes, and tasks in VR experiences (2) A demonstration of how the taxonomy can be employed to classify and generate scene-viewing techniques that do not require head movement, and (3) Empirical data demonstrating participants' perceptions of three new scene-viewing techniques while they experienced limited head mobility.

## 2 RELATED WORK

Taxonomies are a design tool used to organize and reason about large design spaces. Since accessibility in VR is still a relatively new research area, taxonomies could be a useful tool to investigate

how existing VR interaction design can be adapted or applied to an accessibility context. Enhancing the accessibility of existing interactions is not only beneficial to people with disabilities but also to people who are experiencing temporary impairments as a result of their environments.

## 2.1 Design Spaces in Human-Computer Interaction

Design spaces and taxonomies are useful tools within HCI for designing new interaction methods [127], categorizing existing and emerging technologies [79], and synthesizing knowledge across research areas [15]. Surale et al. [127] investigated how tablets can be used to perform complex tasks in VR by constructing a design space with dimensions that explored tradeoffs between mid-air gestures and tablet-centric input. Hirzle et al. [53] proposed a design space for gaze-based interactions with head-mounted displays. Mackinlay, Card, and Robertson [79] created design spaces to categorize and analyze the properties of existing and emerging input devices, which helped them identify points in the design space that warranted further investigation.

Many VR interaction techniques have been developed in the past 30 years and several researchers have organized this design space using taxonomies. These taxonomies have mostly focused on locomotion [152] and object manipulation techniques [61], and are organized by feature and functionality. Organizing existing techniques using a taxonomy can guide the researcher when applying them to a new use case.

## 2.2 Accessibility Research in Virtual Reality

Understanding how to build accessible VR systems and applications is a relatively new area of research, but there is growing interest within the accessible computing and HCI communities to solve VR accessibility challenges. Researchers have explored solutions that improve the accessibility of existing systems by proposing haptic devices that can simulate white cane use in **virtual environments (VEs)** [155], a toolkit with 14 enhancements that make VEs more accessible to people with low vision [156], a taxonomy of sounds to improve accessibility for people who are deaf or hard-of-hearing [59], and a framework for developing accessible **point of interest (POI)** techniques for people with limited mobility [33]. Researchers have also explored the accessibility of movement in VR through wheelchair games [37], a critical examination of the minority body [38], and a design space for translating one-handed to two-handed interaction [148]. Communities of practice, such as XR Access,<sup>1</sup> are also working diligently toward identifying and addressing accessibility barriers endemic to existing VR systems.

Since VR has yet to reach mainstream adoption, there is an opportunity to incorporate accessibility as a core consideration in the design of VR applications [84]. Researchers can leverage the work that has already been done in VR and apply it to accessibility contexts to facilitate the integration of accessibility considerations.

## 2.3 Situational Impairments

Accessible techniques for VR not only benefit people with disabilities but also people experiencing situational impairments. Situational impairments are functional limitations caused by the environment (low lighting, loud noises, etc.) or an individual's mental, emotional, or physical state (divided attention, fear, inebriation, etc.) [116, 144]. Since situations like these can hinder technology use [116, 144], researchers have developed techniques to eliminate or reduce their impact. Most work in VR that addresses physical situational impairments has focused on locomotion for users who are seated [35, 62, 67, 149] or have limited tracking space [57, 58, 123, 131]. Williamson et al. [142]

<sup>1</sup><https://xraccess.org/>

also investigated the effect of social situational impairments caused by using VR on a plane, finding that reduced movement would make VR use more acceptable in this context.

## 2.4 Summary

Taxonomies have been used in HCI as design tools to help researchers organize large design spaces and design novel interactions. Taxonomies can be useful for helping a researcher reason about how existing VR techniques can be applied to new contexts, such as accessibility. Accessibility in VR is underexplored even though accessible solutions can also benefit anybody experiencing situational impairments.

## 3 DEVISING THE SCENE TAXONOMY

Our goal was to devise a scene taxonomy to help designers identify and design scene-viewing techniques that are appropriate based on the visual structure of and task within a VR scene. To devise the scene taxonomy, we reviewed literature in cognitive psychology and computer vision to understand how people and computers view scenes. First, we present the theoretical underpinning of the taxonomy, which is the concept of affordances. Then we introduce the taxonomy and discuss the literature that the taxonomy is grounded in.

### 3.1 Theory: Affordances

Real-life locations afford different ways of viewing. Humans design houses and buildings based on achieving particular views. For example, tourist destinations are engineered to enable the best lookouts and photo opportunities. Scenes in VEs also afford a variety of viewing opportunities. We use the concept of affordances to identify how VEs support different views and scene-viewing techniques.

Gibson, a psychologist, proposed that we perceive affordances directly and that everything in our world is perceived as an affordance [39]. He argued that instead of perceiving qualities or properties of scenes and the objects in them, we immediately perceive the relationship between the observer and the observed [39]. In other words, humans immediately perceive how to interact with an object. For example, the ground is a firm, flat surface that can bear human body weight, so it affords walking. Humans do not need to classify objects based on their characteristics to perceive how they could be interacted with; they do not need to know that the ground can be classified as a hard, flat surface in order to know that it is walkable [39].

Norman brought the concept of affordances to HCI and defined it in the context of technology design [90]. His definition extended Gibson's concept to differentiate between affordances and perceived affordances. He argued that everything affords some interaction. A perceived affordance signals a particular interaction with a particular outcome [89, 91]. So, a user perceives the meaning of a button, which is that it can be pushed to initiate some action. If a perceived affordance is designed well, somebody with little to no experience with the object or interface would know how to interact with it [90]. In Norman's words, "When you first see something you have never seen before, how do you know what to do? The answer, I decided, was that the required information was in the world: the appearance of the device could provide the critical clues required for its proper operation" [90]. If a scene in VR is perceived as a space with affordances that suggest how it can be viewed, how would we describe these affordances? We developed a scene taxonomy that breaks down a scene according to its visual properties to address this question.

### 3.2 Literature Review Approach

In this section, we discuss how we derived the taxonomy based on literature in two disciplines as well as a review of VR applications.

Using the lens of affordances, scenes afford specific ways of being viewed. In other words, the optimal technique to view a scene might depend on its components and structure. Therefore, we reviewed literature in cognitive psychology and computer vision and popular VR applications to identify ways of describing the structure and main components of a scene. We approached the review wanting to answer two research questions: (1) What is the definition of a scene? And (2) How are scenes described?

The objective of our literature review was to synthesize research across cognitive psychology and computer vision and ground a taxonomy in work that is relevant to scene-viewing techniques. We identified overarching patterns with regard to how scenes are defined and described. The breadth of the review, which spanned two disciplines, and the objective to demonstrate an understanding of relevant work, made a traditional literature review appropriate [136].

**Review of cognitive psychology and computer vision literature:** We started our literature review with Tversky and Hemenway’s “Categories of Environmental Scenes” (1983), which was the earliest scene taxonomy we could find that was relevant to our research questions. We then searched for more recent scene taxonomies and work related to scene properties in the 442 works that cited this article which were published between 1984–2020. We found that cognitive psychology and computer vision articles cited this taxonomy and so we reviewed work in both fields.

**Review of VR applications:** In July 2019 and August 2020, we reviewed 29 of the most popular VR applications on the HTC Vive App Store<sup>2</sup> and on the Oculus Rift App Store.<sup>3</sup> We examined the most popular applications so we could explore the types of VEs many people are commonly exposed to. If the app was free, we used it ourselves; otherwise, we viewed YouTube videos of other people using the app. Viewing applications on YouTube<sup>4</sup> was sometimes preferable to experiencing the app first-hand because parts of the app could only be reached with hours of use. While reviewing VR applications, we focused our attention on the VE’s visual properties and the tasks users were performing in them.

Based on our review of literature in cognitive psychology, computer vision, and the VR applications we examined, we identified properties that describe the visual structure of a scene and main task.

### 3.3 The Scene Taxonomy

This taxonomy aims to categorize scene-viewing techniques based on affordances of a virtual scene, namely, visual properties and tasks. Appendix A.1 demonstrates how the taxonomy was used to organize scene-viewing techniques. Prior research findings support this approach: researchers found that some scene-viewing techniques are more effective than others depending on the environment structure [100, 141]. Also, according to research in cognitive psychology, people look at scenes in different ways according to the task they are given [150]. These findings suggest that scene-viewing techniques could be useful for particular environments and tasks. As such, the scene taxonomy has two dimensions: *visual properties* and *task types*. We present an overview of the taxonomy and discuss the research it is grounded in later.

**3.3.1 Dimension and Property Definitions.** We take the approach of breaking down a scene according to its visual properties and tasks. This approach gives designers and researchers a standard language for comparing scenes, which we discuss more in Section 3.4. Comparing scenes would be challenging if we took other approaches such as describing a scene with a category (e.g., beach),

<sup>2</sup><https://www.viveport.com/>

<sup>3</sup><https://www.oculus.com/experiences/rift/>

<sup>4</sup><https://www.youtube.com/>

Table 1. The Visual Properties of the Scene Taxonomy

Visual properties	
Openness	Indoor
	Outdoor
	Abstract
Scale	Smaller
	Human scale
	Larger
Area	Small
	Medium
	Large
Object density	Low
	Moderate
	High
Object tracking	None
	One
	Multiple
Scene transitions	Infrequent
	Frequent
Contains social actors	Yes
	No

Fig. 1. Openness: (from left to right) outdoor,<sup>5</sup> abstract.<sup>6</sup>

by the names of objects in the scene (e.g., sofa) or by the physical interactions they afford (e.g., walkable) because of the numerous possible descriptors in each category. Also, similarities in the structure of the environment might not be captured with these approaches.

**3.3.2 Dimension 1: Visual Properties.** In this section, we present the visual properties of the scene taxonomy (Table 1).

**Openness:** The openness of the scene describes whether a scene is enclosed. *Indoor* scenes are enclosed, such as an office or a cave (Figure 3). *Outdoor* scenes are open scenes. Examples include a forest or an urban setting (Figure 1, left; Figure 2). Through our review of VR applications and scene-viewing techniques (Section 4.1) we found that many scenes in commercial applications feature a space that is neither indoor nor outdoor. Therefore, we added the value *Abstract*.

<sup>5</sup>Model courtesy of TriForge Assets, <https://assetstore.unity.com/packages/3d/environments/fantasy/fantasy-forest-environment-free-demo-35361>

<sup>6</sup>Model courtesy of Simply Rhino, <https://simplyrhino.co.uk/>; Image from: <https://sketchfab.com/blogs/community/preparing-and-uploading-3d-jewelry-designs-in-rhino/>



Fig. 2. Scale: (from left to right) smaller,<sup>7</sup> larger.<sup>8</sup>



Fig. 3. Area: (from left to right), large,<sup>9</sup> small.<sup>10</sup>

Abstract scenes act as a universal space in which 3D objects are designed or viewed (Figure 1, right) [16, 42].

**Scale:** The scale of the scene refers to how large or small objects are relative to the size of the user’s virtual body. If the scale is *human scale* the scene would be at scale relative to the human body. If the user is very large, the objects in the scene would appear small, and the scale of the scene would have the value *smaller* (Figure 2, left). Conversely if the user is very small, the scale would have the value *larger* (Figure 2, right).

**Area:** The area describes how much area the scene covers. A *large* scene covers a large area, such as a park (Figure 3, left). A *small* scene covers a small area, such as a bathroom (Figure 3, right). A *medium* scene is neither large nor small. A scene can be both large and indoor, for example, a warehouse.

**Object Density:** Object density describes the number of objects in the scene. A forest would be a scene with *high* object density (Figure 1, left), while an empty warehouse would be a scene with *low* object density (Figure 3, left). A scene with *moderate* object density has neither a particularly low nor high number of objects.

**Object Tracking:** Object tracking describes the quantity of moving objects in the scene. In many VR applications we surveyed, the user must track moving objects (e.g., any app in the offense/defense task category). We add object tracking as a property and categorize moving objects in terms of quantity: *multiple*, *one*, or *none*.

<sup>7</sup>Model courtesy of 255 pixel studios, <https://assetstore.unity.com/packages/3d/environments/urban/city-package-107224>

<sup>8</sup>Model courtesy of 255 pixel studios, <https://assetstore.unity.com/packages/3d/environments/urban/city-package-107224>

<sup>9</sup>Model courtesy of Nova Shade, <https://assetstore.unity.com/packages/3d/environments/fps-warehouse-74390>

<sup>10</sup>Model courtesy of Jake Sullivan, <https://assetstore.unity.com/packages/3d/props/interior/apartment-bathroom-101320>



Fig. 4. Scene transitions: infrequent.<sup>11</sup>

Table 2. The Task Type Dimension of the Scene Taxonomy Breaks Down the Scene Based on the Task the User is Performing in it

Tasks
Movement
Socializing and Collaboration
Exploration
Navigation
Offense/Defense
Creativity
Observation
Productivity

**Scene Transitions:** Through our review of VR applications (Section 3.2), we found that many scenes in commercial applications feature environments where scenes change frequently. Therefore, we added the scene transitions property to refer to how often the users' scenes change. For example, a scene with *frequent* scene transitions is a large building with multiple hallways and rooms. Users experience a scene transition whenever they enter a new room. An environment with *infrequent* scene transitions is an underwater landscape where there are no obvious transitions (Figure 4).

**Contains Social Actors:** This term describes whether the scene contains avatars or characters the user can interact with. Social actors can be avatars controlled by other users or by the app.

3.3.3 *Dimension 2: Task Types.* Since the way people describe a scene is affected by their task [150], we added task type as a second dimension (Table 2) to the scene taxonomy.

**Movement:** The app's primary goal is to encourage users to move their arms, heads, or bodies. For example, the app Dance Central<sup>12</sup> encourages users to try different dance moves.

**Socializing and Collaboration:** The user's primary goal is to socialize with other users in a virtual space. These VR applications allow users to converse, attend lectures, or play activities together.

**Exploration:** The primary goal is to actively interact with the scene to discover it without having a specific destination. An example is Ocean Rift,<sup>13</sup> in which users explore an underwater sea environment and observe sea animals (Figure 4).

<sup>11</sup>Image courtesy of Llŷr ap Cenydd, Ocean Rift, <https://www.meta.com/experiences/2134272053250863/>

<sup>12</sup><https://www.dancecentral.com/>

<sup>13</sup><https://www.meta.com/experiences/2134272053250863/>

**Navigation:** The primary goal is to travel to a target. An example of navigation in a VR app is Dreadhalls,<sup>14</sup> where the objective is to find the exit of a large building with numerous rooms and hallways.

**Offense/Defense:** These applications' primary goal is to defend against and/or eliminate enemies. An example is Superhot VR,<sup>15</sup> where users shoot or punch moving enemies to eliminate them.

**Creativity:** The primary purpose of these applications is to enable users to create 3D models, draw, or paint. An example is Kingspray Graffiti,<sup>16</sup> where the primary purpose is to create murals on virtual walls.

**Observation:** The primary purpose is to passively view content without interacting with it (e.g., an immersive movie).

**Productivity:** The objective is to accomplish tasks users would perform on their personal computers. An example is Virtual Desktop,<sup>17</sup> which replaces the 2D monitor with an immersive 3D version of users' desktops.

### 3.4 Scene Understanding

How we define and describe the environments we occupy is not straightforward, even though we experience scenes every moment of our lives. In this next section, we demonstrate how the scene taxonomy is based on literature in two disciplines: cognitive psychology and computer vision, to investigate how each field defines and describes a scene.

*3.4.1 Scene Definitions.* Scenes have similar definitions across the two disciplines, which are: object-based, affordance-based, and property-based definitions.

**Object-based definitions:** In object-based definitions of scenes, the scene is the background in which objects are located. In other words, "Scenes are a perceptual, spatial generalization of objects; they are the setting or context for objects, the background where objects are figural." [135] Scenes also contain information about how objects relate to each other and the background, both spatially and semantically [153].

An object-based definition of a scene also exists in computer vision literature. Researchers define a scene as a 3D space with 3D objects that have a particular spatial layout and label [19, 46, 56]. That is to say, "Virtual scenes are digital 3D representations of natural scenes and contain a set of objects, such as items and avatars that move and interact in 3D space" [96].

**Affordance-based Definitions:** Scenes can be defined in an affordance-oriented manner as a space that affords human movement, poses, and actions. Xiao et al. [145] write that a scene is "a place in which a human can act within or a place to which a human being could navigate".

**Property-based Definitions:** A scene can also be defined based on conceptual and perceptual properties. One definition of a scene, a scene gist, can be thought of as a rough mental sketch of a scene when a person is quickly shown a scene image [92]. It often contains information about the scene's basic category (e.g., beach), the names and locations of objects, and lower-level features such as color and texture.

Drawing from the above definitions, we define a VR scene as a 3D space with 3D objects that a user can interact with [96, 135]. We add to this definition that a scene is the part of a VE that is immediately visible to the user if they were to view it from any camera angle at their current location. A VE can contain multiple scenes; for example, a VE can contain a house and a forest.

<sup>14</sup><http://www.dreadhalls.com/>

<sup>15</sup><https://superhotgame.com/superhot-vr>

<sup>16</sup><http://infectiousape.com/>

<sup>17</sup><https://www.vrdesktop.net/>

The user can be inside the house, which is one scene, or outside in the forest, which is a different scene.

**3.4.2 Scene Descriptions.** Here we discuss how scenes are described in cognitive psychology and computer vision.

**Cognitive Psychology:** How people describe scenes reveals information about how scenes are viewed and conceptualized [74]. An individual can look at the same scene in different ways depending on the task they are given [150]. As a result, a scene can have various descriptions depending on the individual and task. Although viewing patterns can differ significantly, there is evidence to suggest that the human brain processes visual information in a consistent way [74].

A theory is that two different levels of processing take place during scene perception, they are referred to as perceptual and conceptual processes. Perceptual processes produce perceptual features, which contain purely visual information about a scene such as edges, reflection, texture, and so on. [36, 74]. Meanwhile, conceptual processes produce conceptual features, which contain semantic information, such as objects' names and functions [95]. Evidence suggests that scene recognition may not be solely dependent on the knowledge of objects in a scene. Instead, individuals can quickly process a scene based on visual features, suggesting that these features alone lead to the identification of the scene category and function [93]. A theory is that perceptual features are conceptualized based on experience and the probability that they are related to the shape and size of common objects in scenes [36].

**Computer Vision:** The computer vision research we surveyed focuses on automatically detecting, localizing, recognizing, classifying, and understanding scenes. Scene understanding is a computational approach to the process of perceiving and analyzing a scene. Researchers have used similar mechanisms to those in human scene perception to understand a scene. For example, Oliva and Torralba [93] showed that perceptual properties can provide a sufficient description of a scene's spatial structure. They call this description the "Spatial Envelope" of a scene and proposed five quantifiable properties: *naturalness*, *openness*, *roughness*, *expansion*, and *ruggedness* [93]. They found that by using these features as input, computer vision algorithms could categorize scenes by their perceptual properties rather than knowledge about objects in the scene [93].

Patterson et al. [97] have also found that adding attributes that people commonly use to describe scenes can improve automatic scene classification. When asked to describe a scene, participants used five attributes: materials (e.g., wood), surface properties (e.g., rusty), affordances (e.g., dining), spatial envelope properties (e.g., openness) [93], and objects (e.g., chair). Some scene attributes are related to their functions. For example, a narrow corridor affords walking through, not sitting, nor lying down. Materials can also suggest related human behaviors; a large body of water might be used for swimming. Finally, embedded objects afford behaviors; chairs and tables can be used for eating. The function of a scene is often present in people's scene descriptions, suggesting that scene function is important to an environment's cognitive schema (a mental framework for organizing knowledge about the world) [151].

Another approach to describing a scene is to detect and identify objects in a scene. Satkin et al. [113] used a database of 3D models of indoor spaces and matched objects in these models to the geometry of objects in a scene image. They used four visual attributes to describe both images and 3D models automatically. Once a 3D model was generated from the 2D image, it was matched with the most similar 3D model in a database of models of indoor spaces [114]. Ismail et al. [56] were able to achieve scene understanding without using a database of models. Their method deconstructed scenes into two parts: an estimation of objects' spatial layout and a model of the spatial relationships between objects.

Functional aspects of environments can be used as descriptors to achieve automatic scene understanding [54, 73, 75, 93, 140]. In a notable example, Gupta et al. [46] broadened the concept of scene understanding by estimating the 3D geometry in the space and the potential for human poses, which they call the “workspace” of a scene. They argue that scene understanding means acquiring a human-centric understanding of the space by prioritizing geometry that is important to humans. For example, estimates of the floor would be more important than the estimates of the ceiling because humans normally interact with the floor. They achieved scene understanding by estimating the scene’s geometry, modeling common human poses, and using these two sources of information to predict poses supported by the environment. Essentially, Gupta et al.’s [46] approach describes scenes by the human postures and activities available in them: the scene is sit-able, walk-able, and so on.

### 3.5 Scene Properties

Here, we discuss how the scene taxonomy’s visual properties are connected to the research discussed above. The task properties were derived from our review of VR applications.

#### 3.5.1 Dimension 1: Visual Properties.

**Openness:** We borrow Oliva and Torralba’s [94] spatial envelope term and concept, which refers to how enclosed a space is. However, we do not define openness by the amount of horizon line visible because it can be hard for humans to quantify. Our definition is closer to the distinction between “indoor” and “outdoor” observed in psychology experiments. Tversky and Hemenway [135] devised a taxonomy of environment categories to reflect cognitive schema. Their taxonomy was organized into three levels. The top, most generic level had only “indoor” and “outdoor” as categories. This dichotomous categorization has also been observed by other researchers in participants’ scene descriptions during psychology experiments [109, 110, 135].

Xiao et al. [145] proposed a taxonomy of scenes which was used to organize a large database of scene images for scene categorization tasks in computer vision research. Like Tversky and Hemenway’s [135] taxonomy, their scene categories were structured in a three-level taxonomy, with the highest-level categories being “indoor”, “outdoor natural”, and “outdoor man-made”.

**Area:** Participants used the amount of space covered in their descriptions of scenes in a study investigating how people categorize scenes [110]. Additionally, the area of scenes communicates the potential for certain human activities [46, 145], which is relevant to affordance-based scene descriptions.

**Scale:** Scene understanding research did not surface this property, likely because scenes used in the studies are realistic and human scale. Scale was informed by the VR app review. Some scenes in VR applications varied in scale, such as Google Earth, which had a smaller scale.

**Object Density:** Object density is related to the spatial envelope property *roughness*, which describes the number of planes in a scene [93, 94]. *Roughness* is related to the size and shapes of objects in a scene. A scene with high roughness, such as a forest, will have many small planes and contours, whereas an empty room will have low roughness because it consists of few large planes (i.e., walls). There will be more planes in a scene with a greater number of objects, resulting in a high degree of roughness.

**Object Tracking:** We include this property based on participants’ descriptions of scenes in cognitive psychology research. Rummukainen and Mendonça [109] found that motion in a scene was one of the first things participants described when asked to recall a scene’s properties. To further highlight the relevance of motion is the existence of dedicated neural processes for perceiving it [51].

**Social Actors:** We included this property based on research in cognitive psychology. One of the most salient features in participants’ scene descriptions was the presence of animate objects

or objects with faces. In addition, research has confirmed the existence of a dedicated region of the brain for processing face-like arrangements [41]. This property is binary (e.g., *yes*, *no*) because cognitive psychology research suggests that the mere presence of an animate object is enough to influence participants' scene descriptions [17, 31, 151]. Findings consistently show that individuals tend to fixate on faces in a scene [17, 31, 151], suggesting that faces are an important component of scene descriptions.

**3.5.2 Dimension 2: Tasks.** Our review of popular VR applications (see Appendix A.2) and app categories in mainstream VR app stores found that the most common task types were *movement*, *socializing*, *exploration*, *navigation*, *offense/defense*, *creativity*, *observation*, and *productivity*. We further refined the task types and definitions based on our review of scene-viewing techniques in the research literature, which we discuss below in Section 4.

### 3.6 Summary

Scenes are defined in various ways within cognitive psychology and computer vision; however, some definitions are similar across these fields. One definition that spans the three fields is that *a scene is an immersive, three-dimensional space that contains objects and affords various interactions*. Approaches to describing scenes are also similar across fields. A few ways of describing a scene are categorizing it, identifying the objects within it, or listing perceptual and conceptual features that describe its visual structure. In the next section, we demonstrate how the taxonomy can be used to address an accessibility problem and organize literature on scene-viewing techniques that require little to no head movement.

## 4 APPLICATION OF THE SCENE TAXONOMY

We applied the scene taxonomy to an accessibility problem in VR: the most pervasive scene-viewing technique in VR is head tracking, which requires the user to move their head and body to view the scene, yet head tracking might not be accessible to people with limited movement due to disability or situational impairment. The most common alternative to head tracking implemented in VR applications is camera panning, which we call *panning*, where the user manipulates the first-person camera to pan right or left using a controller button, often the thumbstick. Like head tracking, panning might also be inaccessible to people with motor or situational impairments because it requires the user to continuously press a button to change the view. This interaction can be tiring if it is used frequently, and it is inefficient for fast-paced tasks. Yet head tracking and panning are usually the only two scene-viewing techniques available in commercial VR applications.

As VR research has matured, many interaction techniques have been developed for viewing VEs. Designers could leverage the large number of techniques to provide users with more options and enable them to choose a technique that best suits their context, preferences, and abilities. We reviewed this literature to identify scene-viewing techniques that can be used as alternatives to head tracking or panning. The techniques we discuss below require little to no head or trunk movement from a viewer, which by design could make them accessible to users with limited head and trunk movement as a result of a disability or situational impairment.

Using the scene taxonomy to organize techniques narrows the scope of suitable scene-viewing techniques to those that are most appropriate for the user's VE and task. Although organizing scene-viewing techniques by scene properties does not guarantee that they are accessible for any type of impairment, it enables designers to choose a technique from a smaller subset of scene-viewing techniques, which facilitates the selection process. First, we summarize techniques in this space, then we show how the taxonomy can be used to organize the techniques.

#### 4.1 Review of Scene-Viewing Techniques

We reviewed literature on scene-viewing techniques designed for 3D virtual and physical environments to gain an overview of the design space see (Appendix A.1). We used Elmqvist and Tsigas' [27] and Cockburn, Karlson, and Bederson's [22] taxonomies as starting points. Our review includes the work cited in these surveys and expands upon them by including more recent work. We searched the works that cited the individual articles in these surveys to find more recent examples of scene-viewing techniques. We also reviewed examples of scene-viewing techniques designed for augmented reality and VR by searching the work that cites seminal VR articles by Stoakley et al., 1995 [124] and Ware and Osborne, 1990 [141].

We define scene-viewing techniques as techniques that facilitate the perception, discovery, and understanding of a 3D VE on any platform (desktop, CAVE, AR, etc.). The environments we discuss are all 3D, but how they are experienced can vary. These environments might have been presented on immersive VR platforms, mixed reality platforms, touchscreens, projected displays, or desktops. Some of these VEs do not resemble physical environments and are primarily used for making 3D models. Input methods also vary; users might employ touchscreen gestures, 3D movements, on-screen UIs, head movement, mouse selection, bi-manual mouse selection, and more.

*4.1.1 Camera Control.* Camera control techniques manipulate the orientation and position of a virtual camera. They can be grouped based on the level of user control they afford. They include automatic techniques, techniques that allow partial user control, and techniques that are exclusively controlled by a user.

*Automatic camera control* techniques aim to support user awareness of a VE without requiring the user to manipulate the virtual camera. These techniques find viewpoints in the environment which provide the most important information for understanding a scene. Automatic camera control is achieved through planned camera paths [1, 21, 24, 29, 121] or individually generated viewpoints [25, 85] in a VE.

Camera control techniques that allow partial user control to combine automatic and user-controlled approaches, which we refer to as *guided camera control* techniques. For example, users can define viewing constraints for generating viewpoints and animations [7, 16, 40]. Pavel et al. [98] developed a camera control technique that supports head tracking when a user is viewing a 360° movie using an HMD. When invoked, the technique repositions the scene such that the focus is directly in front of the user. Guided camera control techniques are also a feature of selection-based locomotion techniques, in which the camera orients and travels to a target selected by the user [10, 11, 152].

The last category contains techniques that enable direct control of the virtual camera's orientation, position, and movement. We refer to this category as *direct camera control*. Researchers examined how existing mental models can be leveraged for direct camera control techniques. For example, Ware and Osborne [141] evaluated three metaphors: (1) eyeball in hand, in which users directly manipulate a viewpoint (2) scene in hand, in which users manipulate the scene to achieve a view, and (3) flying vehicle control, in which users view the environment through a virtual vehicle they are controlling. Like scene in hand, Hinckley et al. [52] used physical props resembling the virtual object to control the viewing angle and examine it from multiple perspectives. Another direct camera control technique that uses metaphor is the **world in miniature (WIM)** technique [124]. With WIM, a user interacts with a miniature model of their environment. The individual can change their view by manipulating a miniature virtual camera in the model of the scene. Most VR locomotion techniques use direct camera control including walking in place [62, 63, 118], steering-based [72], WIM-based [108, 137, 143], and selection-based techniques [35].

**4.1.2 Scene Modification.** *Scene modification* techniques introduce objects or alter object properties to expose, highlight, or reveal semantically meaningful information in a scene. Researchers have investigated placing text labels in the environment to provide information about scene elements [103–105]. Lange et al. [70] introduced swarms of small objects into VEs to subtly capture and guide the user’s gaze to essential parts of a scene. Similarly, Sukan et al. [126] introduced a technique that added a cone-like object to the environment, which guided a user’s head position to achieve the desired view.

Altering object properties is another common scene modification approach. For example, Pierce and Pausch [100] enlarged landmarks to help individuals orient themselves in large VEs. Other techniques that alter parts of the scene include exploding object views [122], distorting terrain to make routes visible [129], deforming layers in volumetric objects to expose interior parts [81], changing the transparency of objects, and cutting away parts of objects to reveal occluded components [30].

**4.1.3 Social Cue Augmentation.** Social VR experiences are growing in popularity [130], and with this trend comes the challenge of perceiving and understanding non-verbal communication from other avatars. Tanenbaum et al. [130] identified four categories of non-verbal communication design in commercial social VR applications: (1) Movement and proxemics, (2) facial control, (3) gesture and posture, and (4) virtual environment-specific non-verbal communication. The authors identified ways in which avatars are designed to communicate with facial expressions, posture, and poses to make social cues more understandable and appropriate for other users. Mayer et al. [80] designed a *social cue augmentation* technique in VR by making deictic gestures, which are pointing gestures that communicate what an individual is referring to, understandable from an observer’s point of view. Specifically, they rotated the avatar’s pointing arm to make gestures visible to an observer.

Another challenge researchers have addressed is collaborating in VEs [20, 32, 69, 102, 154]. Giving users an accurate representation of their bodies relative to the environment and other users’ bodies is essential to experience VEs fully. Chenechal et al. [20] developed a system that enables users to leverage the advantages of being at different scales. When two users manipulated an object, the large-scale user performed more coarse-grained manipulations (e.g., placing the object in an environment), while the smaller user performed manipulations that required more precision (e.g., adjusting the object’s rotation).

**4.1.4 Multiple Views.** Additional views are typically overlaid on a primary view of a VE in *multiple view* techniques. For example, camera views overlaid on the primary camera view enables exploration of a virtual world from multiple viewpoints [5, 125, 138]. An alternative paradigm to overlay views, called magic lenses, presents another view of objects. These lenses exist between the pointer and interface and present changes to the properties of the interface beneath. Originally developed for 2D desktop applications [8], magic lenses have been applied to virtual objects to reveal occluded interior components, such as the skeleton of a hand [64–66, 77, 78, 139].

**4.1.5 Amplified Head Rotation.** *Amplified head rotation* techniques map a small head-movement in physical space to a larger movement in virtual space. In Sargunam et al.’s [111] amplified head rotation technique, a 360° view of the environment could be acquired with a fraction of the physical head movement. Unlike previous approaches [71, 86], they amplified head rotation by a dynamic factor. An issue with amplified head rotation techniques when a user is seated is that the user might have to hold an uncomfortable head position (e.g., looking over the shoulder) to achieve a particular view. To address this issue, Sargunam et al. [111] designed a guided head rotation technique, where the view shifted slightly each time the user teleported in the virtual world until

the user's head was facing forward. Amplified head rotations are feasible for performing 3D search tasks while maintaining spatial awareness as long as the amplification factor is not large [106].

**4.1.6 Field of View Extension.** *Field of view (FOV) extension* techniques alter the physical hardware of the **head-mounted display (HMD)** to simulate peripheral vision [45]. Xiao and Benko [146] used arrays of LEDs inside an HMD to replicate the VE colors outside of the user's FOV. Instead of LEDs, Rakkolainen et al. [107] used smartphones on either side of the viewer's face to show more of the VE. In addition to adding devices and lights to HMDs, camera hardware modifications, such as the use of Fresnel lenses and 360° cameras, have been explored as FOV extension techniques [2, 147].

**4.1.7 Multiscale Techniques.** *Multiscale* techniques enable a user to view an environment at different scales. These techniques are specifically made for VEs that offer a range of detail. Challenges involving multiscale techniques are camera speed and control during navigation and zooming [3, 34, 68, 82, 101, 134]. Without adjusting camera properties during zooming and navigation, the user could experience double vision and decreased depth perception in VR [3]. For example, GIANT [3], a technique for viewing environments with stereoscopic immersive displays, adjusts the scale factor and speed of the camera based on the user's perceived navigation speed.

**4.1.8 Cue-based POI Techniques.** *Cue-based POI* techniques provide cues about the location and direction of out-of-view POIs. Initially developed for desktop and mobile platforms [6, 47], cue-based techniques have also been successfully implemented in CAVE systems, AR, and VR [43, 44, 99]. For example, classic techniques such as Halo [6] and Wedge [47] have been used to support discovery in augmented and VEs [43, 44]. A novel cue-based technique, Outside-In [76], uses picture-in-picture displays to show out-of-view POIs in 360° videos. The rotations and positions of the superimposed displays provide distance and direction cues for the POIs. Nearmi identifies the basic components of cue-based POI techniques in VR and suggests how these components can be used to make the techniques accessible to people with limited head or trunk movement [33].

**4.1.9 Projection Distortion.** *Projection distortion* techniques integrate multiple camera views to support discovery and artistic expression in VEs [26]. For example, Singh [117] proposed a new interactive camera model that allows users to create nonlinear perspectives of virtual objects as a form of artistic exploration. Researchers have also investigated projection distortion as a means of overcoming occlusions [18, 23]. For example, Cui et al. [23] proposed a technique that bends the camera ray around occluding objects, so that target objects are in the line of sight. Although projection distortion techniques allow users to see more of a target and overcome occlusions, the resulting view could be confusing because it deviates from how we usually perceive 3D objects.

**4.1.10 Conclusion.** The scene-viewing techniques discussed above were developed for a variety of platforms, environments, and input methods. Scene-viewing techniques address problems that range from controlling a camera view, circumventing occluding objects, and enhancing awareness of out-of-view POIs. What these techniques have in common is that they address the overarching challenge of perceiving, discovering, and understanding a VE and require little head or trunk movement.

## 4.2 Classification of Scene-Viewing Techniques and Scenes

We classified the scene-viewing techniques reviewed above based on the types of environments they were designed for and evaluated in (see Appendix A.1). We categorized techniques by the visual structure of VEs used in the study apparatus they were evaluated in by inspecting images in the article and supplemental videos.

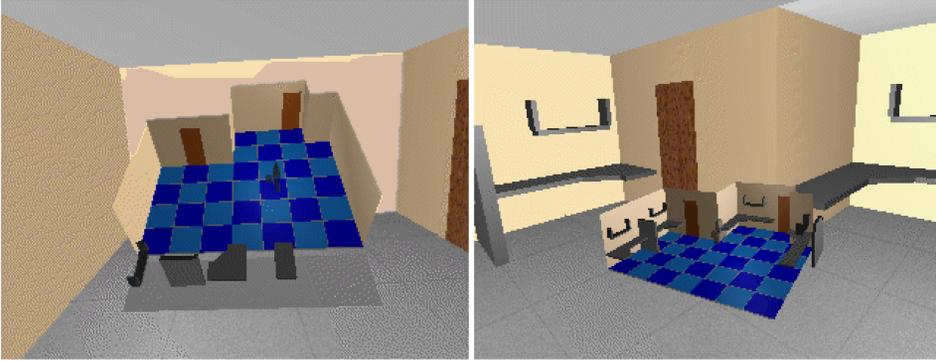


Fig. 5. Figures from [124]. Left: A miniature model of the VE. Right: the miniature model of the VE with the human-scale VE behind it.

Our purpose was to use the scene taxonomy as a designer might when they are examining the design space of scene-viewing techniques to identify some to implement. Classifying scene-viewing techniques can help the designer choose techniques based on the VE they are designing. Classifying techniques can also reveal patterns in the design space such as popular VE types and tasks. Finally, classification can reveal VEs and task types for which scene-viewing techniques do not yet exist.

We classified the VE that the **Worlds in Miniature (WIM)** technique [124] was evaluated in to demonstrate that scene-viewing techniques are appropriate for specific scene properties (Figure 5). WIM was tested in an *indoor* space because the VE was the inside of a building. We classified the scale as *human-* and *smaller-scale* because the user was inside the building while manipulating a smaller model of the building. The space covered a *small to medium area* because the user was navigating a room in a building. It had *low object density* since it was composed of walls, doors, and some shelves. Also, because the environment was a single room, it had *infrequent* scene changes. Finally, there were *no moving objects* or *social actors* because they were not mentioned in the article nor evident in screenshots of the VE. WIM was evaluated for exploration and navigation tasks [124]. The user manipulated the first-person camera by maneuvering a camera proxy in a miniature 3D model of the enclosed space (see Appendix A.1 for classification).

The metaphor of a miniature model was appropriate for the visual structure it was designed for; its boundaries were clearly defined because the VE was enclosed, making manipulating the camera like manually manipulating an object. If the environment was open, large, and had high object density, the miniature model metaphor might not be as effective. Guided camera control techniques might be more effective for these environments because the user can choose from various paths that present the most important parts of a scene [29].

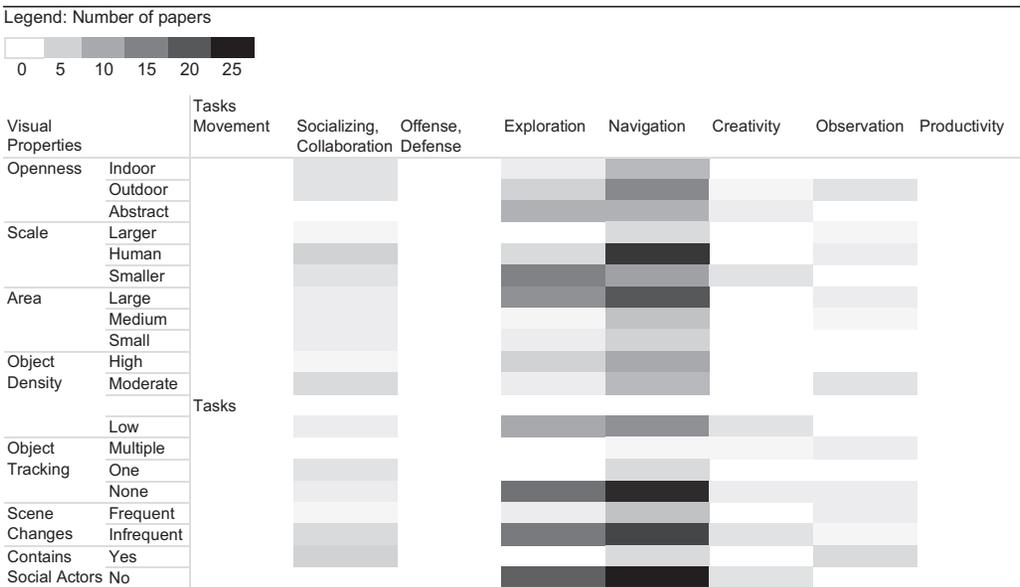
WIM could be a suitable technique for a VR app such as *The Room: A Dark Matter* (Figure 6) in which the user searches different rooms for clues to unlock the next level. The Room would also be classified as an *indoor, human-scale*, covering a *small to medium area* environment with *low to moderate object density, infrequent scene changes*, without *moving objects* nor *social actors*. The user could use WIM to position the camera proxy in the miniature model to wherever they wanted to be in the room.

The above example demonstrates how designers can use the scene taxonomy to identify appropriate scene-viewing techniques for a particular VE. Designers would first classify their VE using the taxonomy and then identify a scene-viewing technique that has been evaluated in a similar environment.



Fig. 6. A scene from the VR app *The Room: A Dark Matter*.<sup>18</sup> This scene would be categorized with the scene taxonomy in the same manner as WIM.

Table 3. This Heatmap Shows an Overview of the Scene-Viewing Techniques in Research and is a Summary of Appendix A.1



It reveals popular tasks and visual properties as well as gaps in research. We categorized the VEs and tasks that were used to evaluate the scene-viewing technique in the article and then counted the articles in each category to produce the heatmap. The darker the cell, the more articles on scene-viewing techniques in that category.

4.2.1 *Common Scene Properties and Tasks in Research Virtual Environments.* In our classification of viewing techniques from the related work section (see Appendix A.1), we found that the most common scene properties were *outdoor*, *human scale*, *large spaces*, *infrequent scene changes*, *no social actors*, and *no moving objects to track* (Table 3). The most common task was *navigation*. Other tasks, such as *offense/defense* or *socializing*, can be considered to have navigation or exploration as a base task. For example, social applications require the user to navigate to an avatar to interact. Because exploration and navigation are the building blocks for other tasks, it is reasonable for researchers to focus on viewing techniques for navigation tasks.

### 4.3 Using the Scene Taxonomy as an Ideation Tool

The empty cells in Appendix A.1 reveal underexplored areas in the scene-viewing technique design space. In the classification we performed, visual properties that were underexplored were *one or multiple objects to track*, *frequent scene changes* and the *presence of social actors*. Designers and researchers could design novel techniques for these properties by considering the usability or accessibility issues that panning would introduce in these environments. Designers could also identify a commercial app that has a VE with a particular property and then experience panning without moving their heads or bodies. In this way, the designer could understand what the user experience is when they are limited to a single scene-viewing technique. For example, the designer might find that panning would require multiple button-presses to orient the camera to discover a moving object. A designer or researcher might also classify their VE and identify scene-viewing techniques that were designed for the same VE types and tasks. Existing techniques built for a similar VE could provide inspiration for a novel technique.

Finally, researchers have demonstrated that the scene-viewing techniques we classified are successful if they are used in VEs like the VE used for the study apparatus. However, scene-viewing techniques might also be useful in different types of environments from the ones they were tested in. The researcher or designer could create a new VE with properties that the technique has not been tested with. If the technique would not work in an environment with different properties, the designer could adapt the technique so that it could be used in this new environment, which could result in a novel technique.

Up to this point we have introduced the scene taxonomy, reviewed the research it is grounded in, and discussed how the taxonomy can be used to classify scene-viewing techniques that require little to no head movement. Next, we evaluated the taxonomy and demonstrated its utility as an ideation tool by generating accessible techniques. We wanted to investigate if the scene taxonomy was useful for generating scene-viewing techniques that leverage scene properties. In the following section we demonstrate how we identified gaps in the taxonomy and designed scene-viewing techniques that address the gaps.

## 5 USING THE TAXONOMY TO INFORM THE DESIGN OF SCENE-VIEWING TECHNIQUES

We designed three scene-viewing techniques to demonstrate how the scene taxonomy can be used as an ideation tool. These three scene-viewing techniques can be used as alternatives to or extensions of panning when the user has limited head or trunk movement. A VE might not be accessible for people who cannot easily rotate their heads or bodies to change the camera orientation because most VR experiences rely on head tracking as the predominant scene-viewing technique. The goal of the techniques we developed was to facilitate the perception, discovery, and understanding of a scene without requiring body movement or repetitive controller manipulation.

### 5.1 Implementation Details

All techniques were implemented in Unity 2019.1.7. We used the Oculus Rift S headset and controllers, which was connected to a Lenovo ThinkStation P330.

**5.1.1 Thumbstick Panning, Teleport, and Pointing.** In addition to the three scene-viewing techniques, we implemented basic controls (panning, Teleport, pointing) that could be used in parallel with our scene-viewing techniques. In our implementation of panning, when the user pushed the

<sup>18</sup>Image courtesy of Fireproof Studios, <https://www.thevgrid.com/the-room-vr-a-dark-matter/>

left controller's thumbstick left or right, the camera would rotate  $30^\circ$  to the left or right on the  $y$ -axis. We chose a thirty-degree rotation angle based on prior work [112].

We used the VRTK 3.3 straight pointer, which rendered as a ray emitting from the user's left controller. To activate the pointer, the user touched the top of the left thumbstick. To select with the pointer, the user pulled the left trigger. We used the VRTK 3.3 Bezier point and Teleport technique [12], which rendered as a dotted curve emitting from the user's right controller. We used Teleport as the locomotion technique because it is implemented in many VR applications [35]. To activate the teleporter, the user touched the right thumbstick with his or her thumb. To teleport to a location, the user pressed the right trigger.

## 5.2 Design Considerations

The main objective when designing scene-viewing techniques was to make the techniques accessible for people with limited head or trunk movement. As such, we designed techniques that required no head movement and little body movement (e.g., arms, hands) and fewer interactions with the controller than panning. In addition to accessibility, we considered factors that are relevant to VR. VR interactions for locomotion and object manipulation are often evaluated based on presence, simulator sickness, and spatial awareness [152]. We identified additional design considerations based on early VR research on evaluating interaction techniques [9], and a textbook on best practices for designing VR experiences [60] to guide our design process.

**5.2.1 Accessibility.** The main accessibility issue we focused on was reducing the effort needed for a user to view a scene, assuming head or trunk rotation was not an option for the user. Panning requires the user to push their thumbstick either continuously or repeatedly to orient the camera. The camera then either moves continuously or at discrete angles until the desired orientation is achieved. If a user is in a VE that requires them to be aware of what is around and behind them, it could be tiring to repetitively pan the camera back and forth. This interaction might be challenging for people with conditions that limit their head movement and even more challenging for people who also have poor hand strength or coordination.

**5.2.2 Usability.** People who use their controllers to orient their cameras are also likely to need to use their controllers for locomotion and object manipulation. As a result, many different controls are mapped to one controller, which can be confusing and difficult to remember. The user must be able to easily recall and form a mental model of controls with low cognitive effort for the technique to have good usability [87, 88].

With current implementations of panning, the user can only orient the camera laterally, (i.e., rotate on the  $y$ -axis). Some VEs might have POIs above or below the user. However, the most common implementation of panning in commercial applications does not enable users to orient their camera towards these objects. The lack of freedom when controlling the camera angle with panning could be detrimental to its usability in environments where POIs surround the user.

**5.2.3 Realism.** Many VR experiences aim to make users feel as if they are physically present in the environment. There is evidence to suggest that a relationship between the realism of a VE and presence, the feeling of being physically present in a VE, exists [55]. As a result, many VR interactions aim to mimic real-world interactions.

**5.2.4 Spatial Awareness.** Some tasks in VR require the user to orient themselves relative to other objects in the VE. Therefore, a user must have a sense of where they are located. Users could feel disoriented after camera transitions, depending on how much of the scene is shown during the transition. Enabling the user to receive enough information to remain oriented is a part of this design consideration.

**5.2.5 User Comfort.** A main design problem in VR is that interactions could inadvertently cause simulator sickness in individuals. Simulator sickness occurs when there is a discrepancy between movement in the virtual and physical environment [83]. For example, continuous camera movement is usually avoided because there is a significant disconnect between perceived virtual movement and the user's physical movement. Simulator sickness is important to consider, especially for scene-viewing techniques that require little movement from the user.

**5.2.6 User Familiarity with Interaction.** Users might not be familiar with interaction techniques in VR because it is still emerging as a consumer technology. To help scaffold the learning process, designers can incorporate interactions that are used in more familiar technologies. For example, a user would likely be familiar with a point and click interaction from desktop computing, or a tap from touchscreen interaction. Using familiar interactions could also reduce the cognitive demand imposed by the controller mappings because there would be fewer novel interactions to learn [88].

**5.2.7 Task Objective.** Some VEs are rich in detail and the main task is to observe or explore the environment. In this case, an efficient scene-viewing technique might detract from the experience of the VE or make the task too easy. Therefore, understanding the task objective, and whether efficiency is a priority, can inform the design of the scene-viewing technique.

**5.2.8 Summary.** It would be ideal to optimize all design considerations but designing the perfect scene-viewing technique might not be possible. Designing a single "perfect" technique is probably not desirable because motor impairments manifest in diverse ways and having different options would likely improve the accessibility of a VE [133]. In addition, accessibility is only one consideration of many when designing a scene-viewing technique. An individual might want to choose a technique that has tradeoffs in other design considerations. For example, if a user is not sensitive to simulator sickness, they could choose a technique that is less comfortable but more realistic. Giving users the freedom to choose techniques based on their individual needs and preferences would enable people to take advantage of the diverse benefits that VR affords. We identify these tradeoffs when we designed the scene-viewing techniques.

### 5.3 Technique 1: Object of Interest

There were few techniques designed to track multiple objects in a VE (Table 3, Appendix A.1), revealing a gap in the design space. This gap surfaces design question: How would a user control the camera view while simultaneously being aware of multiple moving objects? We designed *Object of Interest* for *outdoor, large spaces with moderate to high object density, and multiple objects to track*. Our primary focus was to design a technique that would enable users to be aware of and track multiple moving objects in a scene while *exploring* or *navigating*. Objects in this environment might appear below or above the user. Unlike panning, Object of Interest enables users to orient the camera vertically to look above or below themselves.

Object of Interest was designed so that icons are visible in front of the user (see Figure 7). An icon indicates that an object of interest is outside the user's FOV and appears to the left or the right side of the user, indicating which side the object is on. When the user selects the icon, their camera view cuts to a view that is centered on the object of interest. Icons are visible so long as the user has them toggled on. The user can choose to toggle icons off.

We implemented Object of Interest in an underwater sea environment with sea animals swimming above, below, and around the user. There were four different icons for each sea creature: orca, turtle, dolphin, and octopus. The user could toggle the icons on or off with the left controller

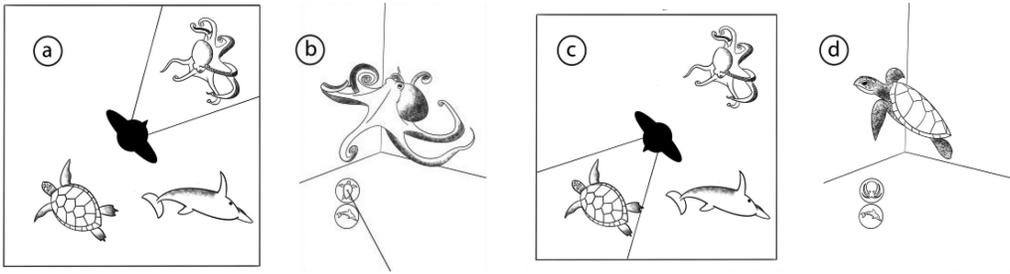


Fig. 7. Object of Interest: (a) The user is facing the octopus in the scene (birds-eye view), (b) She selects the turtle icon with her laser pointer, (c) Her camera rotates to face the turtle in the scene (birds-eye view), (d) She is facing the turtle in the scene.

grip button. When the user selected the orca button, for example, the user's view automatically shifted to an orca in the scene.

When designing the Object of Interest technique, our focus was on making a technique that was more accessible, usable, and comfortable than panning. The camera automatically orients to the user's desired object after the corresponding icon is selected, which demonstrates improved accessibility (5.2.1), because the user does not have to push the thumbstick multiple times to orient the camera to the moving object. We used a UI with icons floating in the virtual world so that functionality was offloaded from the controller. Icons provided affordances that facilitated recognition over recall, which is a usability heuristic [88]. The camera instantly oriented to the object of interest in order to prevent simulator sickness (5.2.5 user comfort).

On the other hand, the nature of the instant camera transition and the presence of a UI decreases realism because there is no real-world equivalent (5.2.3 realism). The immediate camera transition also prevents the user from understanding where the object is relative to other objects in the scene, which could diminish their spatial awareness (5.2.4). The selection mechanism, ray-casting, is unique to VR but the underlying principle of pointing and clicking might be familiar enough to help users learn how to use the technique (5.2.6 user familiarity with interaction). The immediate camera transition enables a more efficient and usable way of viewing objects compared to panning; however, efficiency might not be desirable when exploring an environment because the user would miss contextual information when the camera orients to the object (5.2.7 task objective).

#### 5.4 Technique 2: Proxemics Snapping

Despite the growing interest in social VR applications, few scene-viewing techniques assist social interactions (Table 3, Appendix A.1). The structure of the space changes when social actors are introduced because the space contains zones of interpersonal distance [48]. In the physical world, individuals can easily approach and orient themselves relative to other people while being aware of personal space. However, managing personal space is more complex in VR when using controller-based techniques such as panning and Teleport.

To use panning and Teleport, the user would have to aim their teleporter to an appropriate location in front of another user, which could take several attempts. Next, they would have to fine-tune the camera's orientation using panning, which could take time and be socially awkward. We designed *Proxemics Snapping* for indoor spaces with social actors where the main task is to socialize. This technique was designed to help users search for other avatars in an environment and achieve socially acceptable distances with little effort.

When the user invokes Proxemics Snapping, the camera switches to a third person, zoomed-out view of the environment (see Figure 8). Because the technique is designed for indoor spaces, the

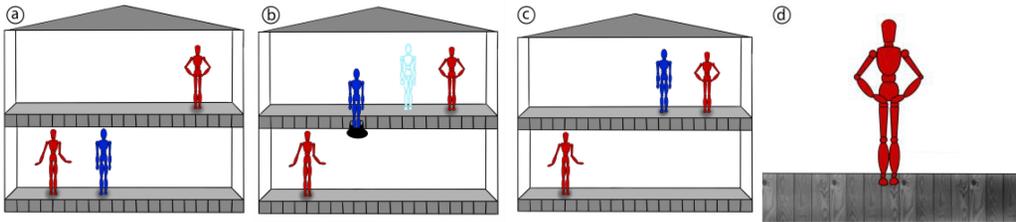


Fig. 8. Proxemics Snapping: (a) The user’s view is zoomed out so that the environment is in the frame, (b) He picks up and moves his dark blue avatar, (c) He places his avatar in the light blue snap drop zone where his avatar orients to a socially appropriate distance and orientation relative to another avatar, (d) He deactivates Proxemics Snapping and returns to a first-person view facing the red avatar.

building’s front wall is toggled off, allowing the user to see into it. The user can see their own avatar as well as the other avatars in the scene. The user can then pick up their avatar and place it in one of several pre-designated locations, or “snap drop zones” near other avatars. When users place their avatar in one of these zones, the avatar snaps to a specified orientation to face the avatar they want to interact with. When users deactivate Proxemics Snapping and resume their first-person view, they will be in front of the other avatar at a socially acceptable distance and orientation.

In our implementation of this technique, the environment was a two-story building with a staircase in the center. Four avatars representing other users were placed around the building. Four snap drop zones were placed in front of the avatars. Snap drop zones were preconfigured distances and orientations relative to the other avatar represented as a light blue shadow of the user’s avatar. When the Proxemics Snapping technique was invoked by pressing the “A” button on the right controller, the user’s controllers scaled up by a factor of 40 and the front wall of the structure was toggled off, revealing the inside of the house. The user could then pick up her avatar by reaching into the house and pressing the right controller grip button to pick up the avatar. When the user hovered their avatar over a snap drop zone in front of another avatar, it would appear as a blue, transparent copy of the user’s avatar. The user could then release the right grip to release the avatar into the snap drop zone. Upon releasing the avatar, the user would return to first-person view.

With Proxemics Snapping, the user can see all avatars in the environment and automatically be positioned in a socially acceptable way without performing small adjustments, which would require more controller use (5.2.1 accessibility). Since the main task objective is to interact with another avatar, the ability to execute the interaction efficiently is appropriate for this task (5.2.7 task objective). The outline of the avatar serves as an affordance indicating where the user can put their avatar. This feature satisfies the recognition over recall heuristic (5.2.2 usability). The interaction is like putting a doll in a dollhouse so is more realistic than using a user interface or teleporting to perform the interaction (5.2.3 realism). While viewing their avatar from a third-person point of view, the user can see where their avatar is relative to the rest of the environment, which could improve spatial awareness (5.2.4). And when the user performs the technique, there is no continuous camera movement that would cause simulator sickness, because the user immediately switches between stationary views (5.2.5 user comfort).

However, the interaction requires the user to reach into the environment and pick up their avatar, which could be inaccessible for people who have a limited range of motion, arm strength, or arm stability (5.2.1 accessibility). In order to invoke Proxemics Snapping, the user must remember to press a particular button on the controller, which violates the principle of recognition over recall (5.2.2 usability). Also, the interaction is unlike any other, so users could be unfamiliar with it and must learn how to use it (5.2.6 user familiarity with interaction).

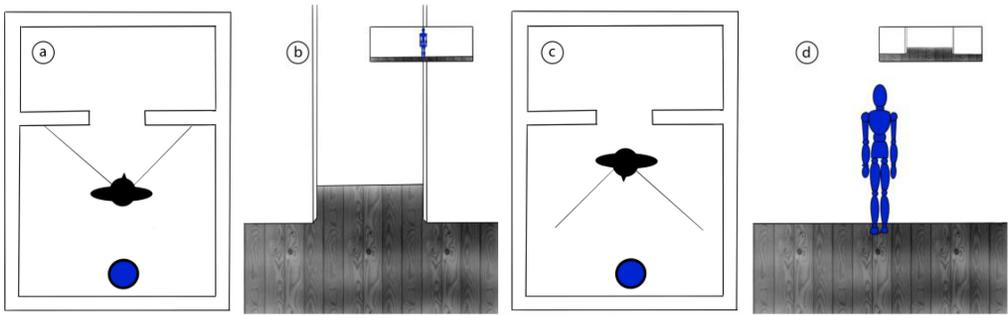


Fig. 9. Rearview Mirror: (a) The user is facing an empty room (birds-eye view), (b) She sees an avatar in her rearview mirror, (c) She presses a controller button to flip 180° (birds-eye view), (d) She faces the blue avatar.

### 5.5 Technique 3: Rearview Mirror

The third gap we identified was the need for viewing techniques for *offense/defense* tasks with *frequent scene changes* (Table 3, Appendix A.1). We designed *Rearview Mirror* for *indoor* spaces, with *one or multiple moving objects* to track, *frequent scene transitions*, and *no social actors*. The objective of this technique was to give users greater awareness of their environment and the ability to change their views quickly.

For this technique, a rearview mirror appears in front of the user and displays what is behind their avatar (see Figure 9). The user can also rotate 180° to face the opposite direction by pressing a controller button.

In our implementation, we positioned a second camera in front of the user and pointed it in the opposite direction the user was facing. To create the rearview mirror, the camera viewpoint was displayed on a small rectangle in front of the viewer and was always visible. When the user pressed the “A” button on the right controller, they rotated 180°.

Accessibility was the primary consideration when designing the Rearview Mirror technique. The user could see behind themselves without any interaction by simply looking in the rearview mirror. Also, the user could turn to face the opposite direction with a single button press instead of multiple thumbstick presses (5.2.1 accessibility). Users would be familiar with looking in a rearview mirror if they knew how to drive (5.2.6 user familiarity with the interaction, 5.2.3 realism). Users might be able to maintain spatial awareness, even when the camera flipped to the opposite direction, because the rearview mirror would remain in the same position and always show the environment behind the user (5.2.4 spatial awareness). Users would not experience discomfort because there is no continuous camera movement (5.2.5 user comfort). Because the task is time sensitive, efficiency is important. Therefore, this technique is appropriate for the task because it reduces the time a user would need to look behind themselves and react to whatever is there (5.2.7 task objective).

## 6 USER STUDY

In the user study, we limited participants’ head movements to investigate effectiveness of the three scene-viewing techniques discussed above compared to panning. Our main goal was to evaluate the scene taxonomy and demonstrate its utility as an ideation tool by designing three scene-viewing techniques that were as easy or easier to use than panning. Designing new usable techniques would give the user more options for ways of viewing scenes in VR. We were also interested in which of the design considerations were important to users and how they perceived tradeoffs in design considerations. We designed the techniques to be first and foremost accessible by not



Fig. 10. Participants' head movements were limited with a chinrest during the study.

requiring head or trunk movement. We also wanted to ensure that we did not sacrifice usability while improving accessibility (e.g., by making the technique difficult to remember or inefficient). So, we were particularly interested in how users would rate the usability of the techniques.

### 6.1 Participants

Sixteen individuals (women:  $n = 5$ , men:  $n = 11$ ) with a mean age of 34.6 ( $SD = 10.6$ ) participated in the study. Participants rated their expertise with computer systems (computers, tablets, smartphones, etc.) on a scale from 1 (novice) to 5 (expert). The median expertise was 3 ( $IQR = 1$ ). Five participants had never owned or used VR while 11 had. Of the 11 participants who had used VR, most reported using it a few times a year ( $n = 6$ ), followed by almost never ( $n = 4$ ), and one person reported using it every few days. Participants who had used or owned VR rated their expertise with VR on a scale from 1 (novice) to 5 (expert). The median expertise was 4 ( $IQR = 1$ ). Only one participant reported experiencing motor limitations, which were poor coordination and rapid fatigue. One participant reported hearing loss.

### 6.2 Apparatus

Participants used the Oculus Rift S headset and controllers. They sat in an office chair and placed their chin on a chinrest, which was attached to a desk (Figure 10). Participants' head movements were limited by the chinrest while completing tasks.

### 6.3 Procedure

The participant first signed a consent form and filled out a demographic questionnaire. They then adjusted the chin rest to a comfortable height.

Next, the researcher explained how to operate the VR controllers to rotate the camera, teleport, and point. These basic controls were available to the participant throughout the study. The participant then completed a training session to practice using the basic controls in a simple environment containing a floor and a keyboard that was suspended in the air. The researcher asked the participant to pan their view left and right using the thumbstick, teleport around the scene, and enter characters into the keyboard with the pointer.

After the basic controls training, the participant practiced using the first technique in an empty scene. Once the participant felt comfortable using the controls, the researcher loaded the task environment. The participant completed the task with the new scene-viewing technique either enabled or disabled (see Table 4 for tasks). When the technique was disabled, the participant could only use the basic controls to perform the task. When the technique was enabled, the participant could complete the task with the technique in addition to the basic controls.

After completing the task, the user took off the headset and completed a NASA-TLX questionnaire [50] to measure workload. We were interested in the physical and mental effort it took to

Table 4. Tasks Participants Completed with Panning and the New Scene-Viewing Techniques

Technique	Task
Object of Interest	“Look at one of each kind of sea animal (orca, turtle, dolphin, octopus)”.
Proxemics Snapping	“Stand in front of each avatar as if you were talking to him or her”.
Rearview Mirror	“Eliminate as many spheres as you can before 2 minutes is up”.

use the techniques to measure the extent that usability in the form of workload was compromised by improved accessibility. Participants then put the headset back on and placed their chins in the chinrest. Once participants were in position, they completed the same task again. If the technique was enabled the first-time participants completed the task, it was disabled the second time and vice versa.

After completing the task again (with the technique enabled or disabled), the user filled out a second NASA-TLX questionnaire. The researcher then interviewed the participant and elicited comparisons between panning and the new scene-viewing technique in terms of preference, ease of use, presence, and simulator sickness. We were mainly interested in the accessibility and usability of the technique, so we did not administer additional questionnaires, such as one for presence [119]. However, if participants were unsure of what we meant by “presence” we explained that it was the sense that they physically existed in the virtual space.

The researcher also asked the participants for their feedback on the scene-viewing technique and if the participants would use the technique in VR. We repeated this procedure for each of the three techniques.

#### 6.4 Design and Analysis

The study was a within-subjects design with one factor, *scene-viewing technique* and two levels: panning and new scene-viewing technique. We used a  $4 \times 4$  Latin square to counterbalance the techniques. We also counterbalanced the presentation order of the scene-viewing technique such that half of the participants experienced panning first and half experienced the new scene-viewing technique first.

We compared NASA-TLX scores for panning and the new scene-viewing technique for each task. We adapted the original 21-point scale to a 1 to 7 scale by eliminating the “high”, “medium”, and “low” increments for each point on the scale to improve readability [49]. Lower scores indicated lower workload. We then compared the raw (unweighted) NASA-TLX scores [49] for panning and new scene-viewing techniques using a Wilcoxon Signed-Rank test, since the data were not normally distributed. A Shapiro-Wilk test revealed that the distribution was significantly different from a normal distribution ( $W = .94, p < .05$ ).

Participants’ interview responses were audio recorded and transcribed. We then qualitatively analyzed participants’ responses using a thematic analysis by finding patterns across participant utterances for each question and deductively grouping utterances by theme (design considerations) [13]. We used qualitative data to understand how design considerations contributed participants’ perceptions and preferences for scene-viewing techniques. Qualitative data were also used to provide insight into NASA-TLX scores.

#### 6.5 Results

We summarize results from the user study below. Although participants were able to use panning in all conditions, we refer to the condition in which participants used the new scene-viewing technique with panning as the *[technique name]* condition. We refer to the condition in which participants used panning without the new scene-viewing technique as the *panning* condition.

Table 5. This Table Presents the Participant Counts of Their Preferred Technique and the Technique They Perceived as Easier

Technique	Object of Interest			Proxemics Snapping			Rearview Mirror		
	Preferred (count)	Easier (count)	NASA-TLX (Median, IQR)	Preferred	Easier	NASA-TLX	Preferred	Easier	NASA-TLX
Technique + panning	7	9	1.5, 1.0	8	12	1.0, 1.0	11	9	3.3, 1.8
Panning only	8	1	1.5, 1.3	7	1	2.0, 1.0	4	5	3.0, 1.8
No preference	1	6		1	3		1	2	

It also presents NASA-TLX median scores and IQR for the new technique and panning. higher NASA-TLX scores mean higher perceived workload. the data indicate that for some participants, their preferred technique was not the same as the technique they thought was easier.

### 6.5.1 Technique 1: Object of Interest.

**Workload:** We did not find a significant difference between NASA-TLX scores for panning vs. the Object of Interest technique ( $Z = .2$ , *n.s.*). Interview responses to the question, “which was easier to use for this task, the Object of Interest technique or panning?” do not reflect the questionnaire results: Even though participants rated both techniques the same in terms of workload, nine participants responded that the Object of Interest technique was easier, one responded that panning was easier, and six responded that they were the same (Table 5).

**Design Considerations:** Participants were divided almost equally on preference for panning vs. Object of Interest. Most participants liked that Object of Interest enabled them to identify and locate objects faster (5.2.7 task objective). For example, P06 explained that Object of Interest helped her find objects: “*Maybe I didn’t know where a particular animal was, then I could use the Object of Interest technique.*” (P06) (4.2.4 spatial awareness).

Some participants also pointed out that Object of Interest allowed them to locate moving objects more easily than panning: “*Since the object keeps moving, I think it’s easier to have [the Object of Interest technique] to easily locate it*” (P04) (5.2.2 usability). Some participants found that it was challenging to use panning when they were trying to track a moving object. For example, P11 said: “*[The Object of Interest technique] would make it easy to find things in a super crowded place, especially if you have limited mobility. I didn’t have to move much or work very hard to find it. Just, oh, it’s there*” (P11) (5.2.1 accessibility).

Icons also provide information about the identity of objects in the scene. P12 explained: “*I liked that I didn’t have to guess about whether or not I was seeing something. You go and look and there’s something that looks like a sea turtle. Okay well that’s a sea turtle because you told me there’s a sea turtle.*” (P12). The icons enabled participants to recognize the object that they were looking at, so they knew what types of objects were in the environment (5.2.4 spatial awareness).

However, half of our participants reported that they disliked feeling disoriented after using Object of Interest. P05 explained, “*It was quicker to get there, but I wasn’t as oriented as a result. So, it got me there, but I was not 100% about my bearings and I had no idea where any of the other things were.*” (P05). The immediate camera transition to the Object of Interest caused confusion about where in the environment they were currently looking (5.2.4 spatial awareness).

None of the participants felt simulator sickness using Object of Interest (5.2.5 user comfort). However, a majority reported feeling less present in the environment with Object of Interest compared to panning. P03 said, “*Just I found, I was focusing more on the buttons than the actual environment itself, I think that’s fun to have control like that, but for this scenario, I feel it takes away from the wonderment of just looking around and enjoying the environment*” (P03). P03 raised the point that a having a UI in the virtual space detracts from the realism of the environment which reduced his sense of presence (5.2.3 realism).

**Usefulness:** Despite half of participants feeling disoriented after using the technique, most participants ( $n = 14$ ) said they would want to use Object of Interest for finding objects in large, crowded environments if efficiency was important for the task (5.2.7 task objective).

#### 6.5.2 Technique 2: Proxemics Snapping.

**Workload:** We did not find a significant difference between NASA-TLX scores for panning vs. Proxemics Snapping ( $Z = 1.1$ , *n.s.*). Yet, most participants responded that Proxemics Snapping was easier to use (Table 5).

**Design Considerations:** Although questionnaire and interview results indicated that most participants perceived Proxemics Snapping to be easier, participants' preferences were not always based on ease of use (5.2.2 usability). Participants were almost equally divided in terms of their preferences for Proxemics Snapping vs. panning. Six participants liked Proxemics Snapping because they did not have to correct their position and orientation to achieve an appropriate personal space relative to the other avatars (5.2.1 accessibility, 5.2.2 usability). For example, P05 said, "*You get a perspective on everything in the space and then you can decide and just go to where you want basically in one move, which I think was more efficient*" (P05).

When using panning, participants were often too close to the other avatars, too far, or at an angle that was not optimal for conversing. P12 explained that he settled for a suboptimal orientation relative to the other avatar when using panning because it was difficult to achieve the right positioning (5.2.2 accessibility): "*[Panning] did make it more likely that I would keep something I thought was slightly off. So, if I was a little bit too close, then it's a lot of work to turn around and try to teleport sideways and then turn back around so I'd probably just leave it*" (P12).

Although some participants reported that they did not dislike anything about Proxemics Snapping, a few reported that interacting with their avatar from a third-person view decreased their presence because it took them out of the first-person view (5.2.3 realism). P14 explained: "*You couldn't really explore with [Proxemics Snapping]. I just like being able to walk around the house, not really walk, but just teleport around and look at stuff from different angles*" (P14). P15 also thought that using the third-person perspective negatively impacted her sense of presence: "*With the [Proxemics Snapping technique], I just dropped the avatar and it kind of oriented itself, so it was a little impersonal. So it was like a dollhouse, it wasn't an experience*" (P15).

One participant reported feeling simulator sickness with Proxemics Snapping when she was in the third-person view, but this was not an issue for most participants (5.2.5 user comfort).

**Usefulness:** Despite some participants feeling less present when using Proxemics Snapping, 13 participants reported that they would want to use the Proxemics Snapping in VR. Participants would use it for locating and moving to other users' avatars in social environments and/or if the environment was large and complex (e.g., with multiple stories and staircases).

#### 6.5.3 Technique 3: Rearview Mirror.

**Workload:** We did not find a significant difference between NASA-TLX scores for panning vs. Rearview Mirror ( $Z = .4$ , *n.s.*). A majority responded that Rearview Mirror was easier to use than panning but their NASA-TLX scores were similar (Table 5).

**Design considerations:** Most participants preferred Rearview Mirror over panning because they liked being able to see behind themselves, which increased their awareness of the environment (5.2.4 spatial awareness). P14 suggested that Rearview Mirror compensated for not having peripheral vision in VR: "*[I like that there was] just more visibility because you don't have as much peripheral vision. So, it gives you more visibility behind you. I mean you still don't have the peripheral but more visibility*" (P14). Participants also appreciated the fact that Rearview Mirror required only a single button press to turn 180° as opposed to panning which required multiple (5.2.1 accessibility, 5.2.2 usability). P01 explained: "*[I liked] the manner or ability of switching sides very easily.*"

*So even if I didn't end up like using that to kill the [enemy] I ended up using it someplace else just to go around faster" (P01). P09 also mentioned the challenge of repetitively pushing the controller to use panning (5.2.1 accessibility). He said, "When I got up the stairs it'd flick around and then have a better view instead of constantly having to turn. You have to flick quite a few times" (P09).*

Although participants appreciated the accessibility of Rearview Mirror, some participants felt that it increased cognitive load (5.2.2 usability). P11 explained: *"I felt like there were too many inputs into my brain and too many outputs in my hand that I had to remember. It was the figuring out how to get from seeing behind me to actually utilizing that information that was hard" (P11).* The 180° turn feature violated the heuristic of recognition over recall, which could explain why participants struggled to use it (5.2.2 usability). Like P11, a few other participants reported forgetting or underutilizing the 180° turn feature despite liking that they could see more with Rearview Mirror (5.2.2 usability). P16 verbalized this issue by saying, *"I'm not really making use of [the turn feature] because... My impression is that there are too many options in my hands so that may make it little difficult for me" (P16).* Despite these comments, the NASA-TLX scores indicated that the workload was not significantly greater with Rearview Mirror compared to panning.

Only one person reported feeling simulator sickness with Rearview Mirror compared to two participants with panning. The fact that two people felt simulator sickness with panning might be because of how frequently they had to pan the camera to see what was behind them (5.2.5 user comfort). Most participants reported feeling present with Rearview Mirror and a few even reported feeling more present compared to panning. P05 explained that being able to see more of the environment increased his sense of presence (5.2.3 realism): *"I did feel present a little more, because I could see more, what was going on around me" (P05).* On the contrary, a couple of participants reported feeling less present with Rearview Mirror because they were less focused on the environment (5.2.4 spatial awareness).

**Usefulness:** Most participants responded that they would use Rearview Mirror in VR if it was available ( $n = 12$ ). Some participants said they would use it for a first-person shooter game to see what was behind them.

## 7 DISCUSSION

In this section, we reflect on the results and process of designing new scene-viewing techniques with the scene taxonomy. We also reflect on the opportunities and limitations that the taxonomy affords. Finally, we discuss how the application of the scene taxonomy could inform the design of scene-viewing techniques for individuals with motor impairments.

### 7.1 Tradeoffs to Consider

Most participants reported that the techniques we designed were easier to use than panning, even though there was not a significant difference in their NASA-TLX scores for the new techniques (Table 5). Every technique we designed centered around specific scene properties, which we identified as gaps in the scene-viewing technique design space. Our findings suggest that by leveraging these affordances in the scene, it is possible to generate different options for viewing a scene with little head or trunk movement. We demonstrated the scene taxonomy's usefulness for generating scene-viewing techniques that are just as easy if not easier to use than the default panning technique.

However, participants did not always prefer the technique they considered easier or more efficient. This finding reveals that it is important to consider multiple design considerations when building accessible techniques.

**Object of Interest:** Even though panning and Object of Interest were similar in terms of workload according to the NASA-TLX scores, most participants verbally reported that they thought

Object of Interest was easier to use than panning ( $n = 9$ ). Participants indicated tradeoffs in spatial awareness, usability, accessibility, and realism: Object of Interest helped users be aware of their surroundings (5.2.4 spatial awareness), helped them identify objects they were viewing (5.2.2 usability), and could switch their view to one of the objects with less controller interaction than panning (5.2.1 accessibility). However, the instant camera transition to the object was disorienting (5.2.4 spatial awareness) and the presence of the UI in the scene detracted from the realism of the environment (5.2.3 realism). These tradeoffs appeared to be based on individual preferences since participants were almost equally divided for their preferred technique (Object of Interest:  $n = 7$ , panning:  $n = 8$ ).

**Proxemics Snapping:** Like Object of Interest, there was not a significant difference in NASA-TLX scores, but most participants reported that Proxemics Snapping was easier to use ( $n = 12$ ). Participants thought Proxemics Snapping enabled them to achieve appropriate personal distances relative to other avatars with less controller manipulation than panning (5.2.1 accessibility, 5.2.2 usability). However, most participants felt less present with Proxemics Snapping when the camera switched from first-person to third-person view, revealing a potential tradeoff between usability, accessibility, and realism (5.2.3). Also, like Object of Interest, these tradeoffs appear to be based on individual preferences since participants were almost equally divided on their preferred technique (Proxemics Snapping:  $n = 8$ , panning:  $n = 7$ ).

**Rearview Mirror:** There was also no significant difference in NASA-TLX scores for panning vs. Rearview Mirror even though most participants reported that they thought Rearview Mirror was easier to use ( $n = 9$ ). Participants appreciated that the Rearview Mirror technique required less physical exertion (5.2.1 accessibility). Even though some participants felt slightly more mentally burdened when using the turn feature (5.2.2 usability), participants felt the same or a greater level of presence when using Rearview Mirror compared to panning (5.2.4 spatial awareness, 5.2.3 realism). Most participants preferred Rearview Mirror to panning (Rearview Mirror:  $n = 11$ , panning:  $n = 4$ ) possibly because not only did it enhance accessibility, but for some participants it also enhanced their sense of presence.

**Summary:** The results suggest that NASA-TLX might not have captured participants' concept of workload because their scores did not always reflect the technique they verbally reported to be easier to use. There might be different factors related to workload in VR that the NASA-TLX did not capture. Also, although participants agreed that the scene-viewing techniques that we designed were easier to use than panning, this was not reflected in their preferred technique. This finding demonstrates that accessibility and usability are not the only design considerations that determine a person's preference for a scene-viewing technique.

The results suggest that participants weighed multiple factors when choosing their preferred technique. In Object of Interest and Proxemics Snapping the weighting of factors appeared to be based on individual preference. However, most participants preferred Rearview Mirror over panning which could indicate that they weighed the tradeoff between accessibility, usability, and realism in a similar way. Future work could explore what appears to be a cost-benefit analysis for a preferred scene-viewing technique. There might be universally desirable or undesirable tradeoffs that could help designers predict if their technique will be preferred over other scene-viewing techniques.

Because tradeoffs exist for scene-viewing techniques and panning, designers should provide the option of using alternative scene-viewing techniques in addition to panning. Users could then choose a technique depending on their preferences for accessibility, usability, realism, spatial awareness, comfort, and familiarity with the interaction, as well as the goal of the task they are performing. Greater flexibility would provide users with more options for overcoming limited head movement caused by a disability or situational impairments.

## 7.2 Reflecting on the Scene Taxonomy: Opportunities and Limitations

The scene taxonomy aims to describe the context of a VR experience in a manner that is consistent with research on how people and computers conceptualize scenes. We could have designed a taxonomy based on users' motor abilities so that designers could select a scene-viewing technique based on the type of impairment an individual has. The problem with this approach is that it is difficult to predict the VR interactions people will struggle with based on their condition and current context. Two people with the same condition could have very different ways of moving. As a result, it would be necessary to study how people with various motor impairments use the scene-viewing techniques to understand how impairment type and severity relate to performance. Because of this issue, we organized scene-viewing techniques based on the types of environments they have been designed for and evaluated in, since the success of the scene-viewing techniques in particular environments is already known. Then, given a VE, a designer could narrow down the subset of scene-viewing techniques that are applicable for their application and evaluate this subset with people with motor impairments to understand their performance.

We demonstrated the use of the scene taxonomy to reason about an accessibility problem in VR, but the taxonomy could be applied to organize the scene-viewing technique design space for a range of purposes. For example, it could be used to organize the design space of movement-based scene-viewing techniques as well. Also, because we constructed the taxonomy in a way that is independent of the properties of any particular set of techniques, it can also be used for categorizing other types of VR interaction such as locomotion techniques, object manipulation techniques, input devices, and so on, to reveal patterns and gaps in these design spaces. In general terms, the taxonomy could enable designers to reason about the design space of various types of VR interaction relative to VR scenes and tasks.

As for a limitation of the scene taxonomy, we found that participants considered multiple trade-offs when deciding whether they would use the new techniques. Looking back, it would not have been possible to use the taxonomy to predict the particular tradeoffs associated with a technique. These tradeoffs would need to be surfaced empirically by evaluating techniques with users. Therefore, while the scene taxonomy can identify patterns in the scene-viewing technique design space, it cannot be used to predict how users will evaluate a given technique.

## 7.3 Accessibility for Individuals with Motor Impairments

The scene-viewing techniques we designed could be useful for individuals with limited head or body movement due to ALS, cerebral palsy, or paralysis. However, controller interactions would also need to be made accessible. For example, if an individual only has use of one hand, interactions for Teleport, object manipulation, panning, and scene viewing would need to be usable with one controller. Also, the accessibility and usefulness of some VR interaction techniques could depend on the input devices being used (multiple switches, game console, foot controller, eye trackers, etc.). Automatically mapping inputs to interaction techniques could be a promising area of future work. Ultimately, the scene taxonomy could contribute to a recommender system for suggesting and mapping VR interaction techniques based on scene properties, task types, hardware affordances, users' abilities, as well as preferences for the design considerations we identified.

## 7.4 Limitations

We classified scene-viewing techniques based on the environments in which they were evaluated. However, it is possible—and likely—that the techniques could be classified under different visual properties and tasks. Therefore, the taxonomy is only a starting point for classifying scene-viewing

techniques, and we acknowledge that any single technique could be classified under different properties based on empirical evidence.

Furthermore, having more participants might have revealed significant differences between NASA-TLX scores. We saw a larger difference in NASA-TLX scores for Proxemics Snapping and panning compared to Object of Interest and Rearview Mirror, indicating that more participants in this condition might have revealed a significant difference in scores.

## 8 CONCLUSION

We devised a scene taxonomy based on the visual properties and tasks associated with VEs from a review of cognitive psychology and computer vision research, as well as a survey of 29 popular VR applications. We then demonstrated how the taxonomy can be used to organize the large body of literature on scene-viewing techniques to address the problem of situational and permanent impairments that affect users' abilities to view scenes. We applied the taxonomy to identify accessible scene-viewing techniques that could be used for different environments and tasks. We also used the taxonomy to identify gaps in the design space that could suggest scene-viewing techniques for people with limited mobility. Based on the gaps we identified, we prototyped three scene-viewing techniques, which we evaluated with participants experiencing limited head movement. We found that most participants thought the new techniques were easier to use compared to panning, however, participants based their preferences on tradeoffs in accessibility usability, realism, spatial awareness, comfort, familiarity with the interaction, and the task objective. The taxonomy could potentially be used to reason about various problems where the scene and task affect how users interact with VR.

## A APPENDIX

### A.1 Classification of Scene-Viewing Techniques Using the Scene Taxonomy

Table 6. Scene-Viewing Technique Papers Classified Based on the Visual Properties and Tasks That Were Used to Evaluate the Scene-Viewing Technique

Visual Properties	Tasks	Socializing, Collaboration	Offense, Defense	Exploration	Navigation	Creativity	Observation	Productivity
Openness	Indoor	Social Cues [80,130], Multi-scale [102]		Automatic [24], Direct [124]	Automatic [1, 21], Scene Modification [126], Amplified [71, 106, 111], Multi-scale [3], Direct [141] (Eyeball in Hand, Flying vehicle control)			
	Outdoor	Social Cues [130], Multi-scale [20, 154]		Multi-view [77, 125], Projection Distortion [23], Guided [7], Multi-scale [69]	Direct [35], Scene Modification [70, 100, 129], Multi-view [5, 14, 138], Amplified [86], FOV Extension [2, 146], Multi-scale [3, 34, 134], Guided [29]	Guided [40]	FOV extension [147], Cue-based [76], Guided [98]	
	Abstract			Automatic [120, 121], Scene Modification [81, 103–105], Multi-view [139], Projection Distortion [18, 117]	Cue-based [43, 44], Multi-scale [4, 68, 82], Projection Distortion [28], Direct [141] (Scene in Hand), Scene Modification [30, 122]	Direct [42], Guided [16]		
Scale	Larger	Multi-scale [20]			Multi-scale [3, 68, 4], Direct [141] (Scene in Hand)		Cue-based [76]	
	Human	Social Cues [80, 130], Multi-scale [20, 102, 154]		Automatic [24], Projection Distortion [23], Direct [124], Multi-scale [69]	Automatic [1, 21], Direct [35, 141] (Eyeball in hand, Flying vehicle control), Scene Modification [70, 126, 129], Multi-view [138], Amplified [71, 86, 106, 111], FOV Extension [2, 146], Multi-scale [3, 4, 34, 68, 82, 134], Guided [29], Cue-based [43, 44]		FOV extension [147], Guided [98]	
	Smaller	Multi-scale [20, 102, 154]		Automatic [120, 121], Scene Modification [81, 103–105], Multi-view [77, 125, 139], Guided [7], Projection Distortion [18, 23, 117], Direct [124], Multi-scale [69]	Scene Modification [30, 100, 122], Multi-view [5, 14], Amplified [71], Multi-scale [3, 34, 82, 134], Projection Distortion [28]	Direct [42], Guided [16, 40]		

(Continued)

Table 6. Continued

Visual Properties	Tasks	Socializing, Collaboration	Offense, Defense	Exploration	Navigation	Creativity	Observation	Productivity
Area	Large	Multi-scale [20, 154]		Automatic [120, 121], Scene Modification [81, 103-105], Multi-view [125, 139], Projection Distortion [18, 23, 117], Guided [7], Multi-scale [69]	Direct [35, 141] (Scene in Hand), Scene Modification [70, 100, 122, 129], Multi-view [5, 14, 138], Amplified [71, 86, 106], FOV Extension [146], Multi-scale [3, 4, 34, 68, 82, 134], Guided [29], Projection Distortion [28]	Direct [42], Guided [16, 40]	FOV extension [147], Cue-based [76]	
	Medium	Social Cues [130], Multi-scale [102]		Direct [124]	Automatic [1, 21], Amplified [71], FOV extension [2], Multi-scale [3], Cue-based [43, 44]		Guided [98]	
	Small	Social Cues [20, 80]		Automatic [24], Multi-view [77]	Scene Modification [126], Amplified [111], Multi-scale [3], Direct [141] (Eyeball in hand, Flying vehicle control)			
Object Density	High	Multi-scale [20]		Multi-view [125], Guided [7], Multi-scale [69], Projection Distortion [18, 23]	Scene Modification [70, 30], Multi-view [5, 14], Amplified [106], FOV Extension [146], Multi-scale [3, 134], Guided [29], Projection Distortion [28]			
	Moderate	Social Cues [130], Multi-scale [20, 102, 154]		Automatic [24], Direct [124]	Automatic [1, 21], Direct [35], FOV extension [2, 146], Multi-scale [34, 82], Multi-view [138]		FOV extension [147], Cue-based [76], Guided [98]	
	Low	Social Cues [20, 80]		Automatic [120, 121], Scene Modification [81, 103-105], Direct [141] (Flying vehicle control), Multi-view [77, 139], Projection Distortion [117]	Scene Modification [100, 126, 129], Amplified [71, 86, 111], Multi-scale [68, 4, 134], Cue-based [43, 44], Direct [141] (Scene in Hand)	Direct [42], Guided [16, 40]		

(Continued)

Table 6. Continued

Visual Properties	Tasks	Socializing, Collaboration	Offense, Defense	Exploration	Navigation	Creativity	Observation	Productivity
Object Tracking	Multiple				Multi-view [5]	Guided [40]	Cue-based [76], Guided [98]	
	One	Social Cues [80], Multi-scale [102, 154]			Automatic [21], Scene Modification [70,129], FOV extension [146]			
	None	Social Cues [130], Multi-scale [20]		Automatic [24, 120, 121] Scene Modification [81, 103–105], Multi-view [77, 125, 139], Projection Distortion [18, 23, 117], Guided [7], Multi-scale [69], Direct [122, 124]	Automatic [1], Multiview [14], Direct [35, 141] (Scene in Hand), Scene Modification [30, 100, 122, 126], Amplified [71, 86, 106, 111], FOV extension [2], Guided [29], Multi-scale [3, 4, 34, 68, 82, 134], Projection Distortion [28], Cue-based [43, 44]	Direct [42], Guided [16]	FOV extension [147], Multi-scale [82]	
Scene Changes	Frequent	Multi-scale [20]		Direct [124, 141] (Flying vehicle control)	Amplified [71, 106, 111], FOV extension [2], Multi-scale [3, 68, 134]		Cue-based [76], Guided [98]	
	Infrequent	Social Cues [80, 130], Multi-scale [102, 154]		Automatic [24, 120, 121], Scene Modification [81, 103–105, 122], Multi-view [77, 125, 139], Projection Distortion [18, 23, 117], Guided [7], Multi-scale [69]	Automatic [1, 21], Direct [35, 141] (Scene in hand), Scene Modification [30, 70, 100, 122, 126, 129], Multi-view [5, 86, 138], Multi-scale [34, 82, 4], Guided [29], Cue-based [43, 44], Projection Distortion [28], FOV extension [146]	Direct [42], Guided [16, 40]	FOV extension [147]	
Contains Social Actors	Yes	Social Cues [80, 130], Multi-scale [20, 102, 154]			Automatic [21], Multi-view [5, 23], FOV extension [2]			
	No			Automatic [24, 120, 121], Direct [122, 124, 141] (Flying vehicle control), Scene Modification [81, 103–105], Multi-view [77, 125, 139], Projection Distortion [18, 23, 28, 117], Guided [7], Multi-scale [69]	Automatic [1], Direct [35], Scene Modification [70, 100, 122, 126, 129], Amplified [71, 86, 106, 111], FOV Extension [146], Multi-scale [3, 4, 34, 82, 134], Guided [29], Cue-based [43, 44], Projection Distortion [28], Direct [141] (Scene in Hand), Scene Modification [30], Multi-view [138]	Direct [42], Guided [16, 40]	FOV extension [147], Multi-scale [82], Cue-based [76], Guided [98]	

Classification is based on our inspection of images in the article and supplemental videos. We group and label articles by the type of scene-viewing technique (e.g., “scene modification”).

## A.2 Classification of VR Applications Using the Scene Taxonomy

Table 7. Twenty-Nine of the Most Popular VR Applications in 2019 and 2020 Classified Based on the Main Visual Properties and Tasks of the VE

Visual Properties	Tasks							
	Movement	Socializing, Collaboration	Offense, Defense	Exploration	Navigation	Creativity	Observation	Productivity
Openness	Indoor	Beatsaber, Job Simulator, Dance Central, Vacation Simulator	VRChat, Altspace	Blade and Sorcery, Saints and Sinners, Robo Recall, Arizona Sunshine, Superhot VR, Pistol Whip, Gorn, Onward	Lone Echo, I Expect You to Die, The Room VR	Dreadhalls		
	Outdoor	The Climb		Ocean Rift, Minecraft, Google Earth	MazeVR	Kingspray Graffiti, Tiltbrush	Within	Virtual Desktop
Scale	Abstract					Quill		Supermedium
	Larger					Quill		Virtual Desktop
	Human	Beatsaber, The Climb, Job Simulator, Vacation Simulator, Dance Central	VRChat, Altspace	Blade and Sorcery, Saints and Sinners, Robo Recall, Arizona sunshine, Superhot VR, Onward, Pistol Whip, Gorn	Lone Echo, I Expect You to Die, Ocean Rift, Minecraft, The Room VR	MazeVR, Dreadhalls	Within	Supermedium
	Smaller			Google Earth				
Area	Large	Beatsaber, Job Simulator, Dance Central, The climb		Blade and Sorcery, Onward, Pistol Whip	Google Earth, Ocean Rift, Minecraft	Tiltbrush, Quill	Within	Virtual Desktop, Supermedium
	Medium	Vacation Simulator	VRChat, Altspace	Saints and Sinners, Robo Recall, Gorn, Superhot VR	I Expect You to Die	Kingspray graffiti		
	Small			Arizona Sunshine	Lone Echo, The Room VR	MazeVR, Dreadhalls		
Object Density	High			Saints and Sinners, Robo Recall, Onward, Pistol Whip	Lone Echo, Google Earth, Ocean Rift, Minecraft			
	Moderate	Vacation Simulator		Superhot VR	I Expect You to Die	Kingspray Graffiti		
	Low	Beatsaber, Job Simulator	VRChat, Altspace	Blade and Sorcery, Arizona Sunshine, Gorn	The Room VR	DreadHalls, MazeVR	Within	Virtual Desktop, Supermedium

(Continued)

Table 7. Continued

Visual Properties	Tasks	Socializing, Collaboration	Offense, Defense	Exploration	Navigation	Creativity	Observation	Productivity
Object Tracking	Movement Beatsaber		Blade and Sorcery, Superhot VR, Onward, Pistol Whip	Ocean Rift				
	One Job Simulator, Vacation Simulator, Dance Central		Arizona Sunshine	Lone Echo, Gorn				
	None The Climb	VRChat, Altspace	Saints and Sinners, Robo Recall	I Expect You to Die, Google Earth, Minecraft, The Room VR	Dreadhalls, MazeVR	Tiltbrush, Quill, Kingspray Graffiti	Within	Virtual Desktop, Supermedium
Scene Changes	Frequent		Saints and Sinners, Arizona Sunshine, Superhot VR	Lone Echo	Dreadhalls, MazeVR		Within	
	Infrequent		Blade and Sorcery, Onward, Pistol Whip, Corn	I Expect You to Die, Google Earth, Ocean Rift, Minecraft, The Room VR		Kingspray Graffiti, Tiltbrush, Quill		Virtual Desktop, Supermedium
Contains Social Actors	Yes Job Simulator, Vacation Simulator, Dance Central	VRChat, Altspace	Onward	Lone Echo		Kingspray Graffiti		
	No Beatsaber, The Climb		Blade and Sorcery, Saints and Sinners, Robo Recall, Arizona Sunshine, Superhot VR, Pistol Whip	I Expect to Die, Google Earth, Ocean Rift, Minecraft, The Room VR, Gorn	Dreadhalls, MazeVR	Tiltbrush, Quill	Within	Virtual Desktop, Supermedium

## ACKNOWLEDGMENTS

We would like to thank Eyal Ofek and Andy Wilson for their insight into the scene-viewing problem space. We would also like to thank Karly S. Franz for the illustrations.

## REFERENCES

- [1] C. Andujar, P. Vazquez, and M. Fairen. 2004. Way-Finder: Guided tours through complex walkthrough models. *Computer Graphics Forum* 23, 3 (2004), 499–508. DOI : <https://doi.org/10.1111/j.1467-8659.2004.00781.x>
- [2] Jérôme Ardouin, Anatole Lécuyer, Maud Marchal, Clément Riant, and Eric Marchand. 2012. FlyVIZ: A novel display device to provide humans with 360° vision by coupling catadioptric camera with HMD. In *Proceedings of the Virtual Reality Software and Technology*. 41–44. DOI : <https://doi.org/10.1145/2407336.2407344>
- [3] Ferran Argelaguet and Morgan Maignant. 2016. GiAnt: Stereoscopic-compliant multi-scale navigation in VEs. In *Proceedings of the Virtual Reality Software and Technology*. 269–277. DOI : <https://doi.org/10.1145/2993369.2993391>
- [4] Rob Aspin and Kien Hoang Le. 2007. Augmenting the CAVE: An initial study into close focused, inward looking, exploration in IPT systems. In *Proceedings of the Symposium on Distributed Simulation and Real-Time Applications Augmenting*. 217–224. DOI : <https://doi.org/10.1109/DS-RT.2007.9>
- [5] William H. Bares and James C. Lester. 1998. Intelligent multi-shot visualization interfaces for dynamic 3D worlds. In *Proceedings of the Intelligent User Interfaces*. 119–126. DOI : <https://doi.org/10.1145/291080.291101>
- [6] Patrick Baudisch and Ruth Rosenbaltz. 2003. Halo: A technique for visualizing off-screen locations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 481–488. DOI : <https://doi.org/10.1145/642611.642695>
- [7] Blaine Bell, Steven Feiner, and Tobias Höllerer. 2001. View management for virtual and augmented reality. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology*. 101–110. DOI : <https://doi.org/10.1145/502348.502363>
- [8] Eric A. Bier, Maureen C. Stone, Ken Pier, Ken Fishkin, Thomas Baudelf, Matt Conway, William Buxton, and Tony DeRose. 1994. Toolglass and magic lenses: The see-through interface. In *Proceedings of the Conference on Human Factors in Computing Systems*. 445–446. DOI : <https://doi.org/10.1145/166117.166126>
- [9] Doug A. Bowman and Larry F. Hodges. 1997. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics*. 35–38. DOI : <https://doi.org/10.1145/253284.253301>
- [10] Doug A. Bowman, Donald B. Johnson, and Larry F. Hodges. 1999. Testbed evaluation of virtual environment interaction techniques. In *Proceedings of the Virtual Reality Software and Technology*. 26–33. DOI : <https://doi.org/10.1145/323663.323667>
- [11] Doug A. Bowman, David Koller, and Larry F. Hodges. 1997. Travel in immersive virtual environments: An evaluation of viewpoint motion control techniques. In *Proceedings of the Virtual Reality Annual International Symposium*. 45–52.
- [12] Evren Bozgeyikli, Andrew Raji, Srinivas Katkooari, and Rajiv Dubey. 2016. Point and teleport locomotion technique for virtual reality. In *Proceedings of the Computer-Human Interaction in Play*. 205–216. DOI : <https://doi.org/10.1145/2967934.2968105>
- [13] Virginia Braun and Victoria Clarke. 2008. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2008), 77–101. DOI : <https://doi.org/10.1191/1478088706qp0630a>
- [14] Leonard D. Brown and Hong Hua. 2006. Magic lenses for augmented virtual environments. *IEEE Computer Graphics and Applications* 26, 4 (2006), 64–73. DOI : <https://doi.org/10.1109/MCG.2006.84>
- [15] Frederik Brudy, Christian Holz, Roman Rädle, Chi Jui Wu, Steven Houben, Clemens Nylandsted Klokose, and Nicolai Marquardt. 2019. Cross-device taxonomy: Survey, opportunities, and challenges of interactions spanning across multiple devices. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–28. DOI : <https://doi.org/10.1145/3290605.3300792>
- [16] Nicholas Burtynky, Azam Khan, George Fitzmaurice, Ravin Balakrishnan, and Gordon Kurtenbach. 2002. StyleCam: Interactive stylized 3D navigation using integrated spatial and temporal controls. In *Proceedings of the 15th Annual ACM Symposium on User Interface Software and Technology*. 101–110. DOI : <https://doi.org/10.1145/571985.572000>
- [17] Guy Thomas Buswell. 1935. *How People Look at Pictures: A Study of the Psychology and Perception in Art*. University of Chicago Press.
- [18] M. Sheelagh T. Carpendale, David J. Cowperthwaite, and F. David Fracchia. 1996. Distortion viewing techniques for 3-dimensional data. In *Proceedings of the IEEE Symposium on Information Visualization*. 46–54. DOI : <https://doi.org/10.1109/INFVIS.1996.559215>
- [19] Chad Carson, Serge Belongie, Hayit Greenspan, and Jitendra Malik. 2002. Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 8 (2002), 1026–1038. DOI : <https://doi.org/10.1109/TPAMI.2002.1023800>

- [20] Morgan Le Chenechal, Jeremy Lacoche, Jerome Royan, Thierry Duval, Valerie Gouranton, and Bruno Arnaldi. 2016. When the giant meets the ant: An asymmetric approach for collaborative and concurrent object manipulation in a multi-scale environment. In *Proceedings of the 2016 IEEE Third VR International Workshop on Collaborative Virtual Environments*. 18–22. DOI: <https://doi.org/10.1109/3DCVE.2016.7563562>.
- [21] Luca Chittaro, Roberto Ranon, and Lucio Ieronutti. 2003. Guiding visitors of Web3D worlds through automatically generated tours. In *Proceedings of the International Conference on 3D Web Technology*. 27–38. DOI: <https://doi.org/10.1145/636593.636598>
- [22] Andy Cockburn, Amy Karlson, and Benjamin B. Bederson. 2008. A review of overview+detail, zooming, and focus+context interfaces. *ACM Computing Surveys* 41, 1 (2008), 1–31. DOI: <https://doi.org/10.1145/1456650.1456652>
- [23] Jian Cui, Paul Rosen, Voicu Popescu, and Christoph Hoffmann. 2010. A curved ray camera for handling occlusions through continuous multiperspective visualization. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 1235–1242. DOI: <https://doi.org/10.1109/TVCG.2010.127>
- [24] Sylvian Desroche, Vincent Jolivet, and Dimitri Plemenos. 2007. Towards a plan-based automatic exploration of virtual worlds. In *Proceedings of the International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*. 25–32.
- [25] Steven Drucker. 1994. *Intelligent Camera Control for Graphical Environments*. Massachusetts Institute of Technology.
- [26] Niklas Elmqvist and Philippas Tsigas. 2007. View-projection animation for 3D occlusion management. *Computers and Graphics* 31, 6 (2007), 864–876. DOI: <https://doi.org/10.1016/j.cag.2007.09.006>
- [27] Niklas Elmqvist and Philippas Tsigas. 2008. A taxonomy of 3D occlusion management for visualization. *IEEE Transactions on Visualization and Computer Graphics* 14, 5 (2008), 1095–1109. DOI: <https://doi.org/10.1109/TVCG.2008.59>
- [28] Niklas Elmqvist and M. Eduard Tudoreanu. 2007. Occlusion management in immersive and desktop 3D virtual environments: Theory and evaluation. *The International Journal of Virtual Reality* 6, 1 (2007), 1–13.
- [29] Niklas Elmqvist, M. Eduard Tudoreanu, and Philippas Tsigas. 2007. Tour generation for exploration of 3D virtual environments. In *Proceedings of the 2007 ACM Symposium on Virtual Reality Software and Technology*. 207–210. DOI: <https://doi.org/10.1145/1315184.1315224>
- [30] Steven K. Feiner and Dorée Duncan Seligmann. 1992. Cutaways and ghosting: satisfying visibility constraints in dynamic 3D illustrations. *The Visual Computer* 8, 5–6 (1992), 292–302. DOI: <https://doi.org/10.1007/BF01897116>
- [31] Sue Fletcher-Watson, John M. Findlay, Susan R. Leekam, and Valerie Benson. 2008. Rapid detection of person information in a naturalistic scene. *Perception* 37, 4 (2008), 571–583. DOI: <https://doi.org/10.1068/p5705>
- [32] Cédric Fleury, Alain Chauffaut, Thierry Duval, Valérie Gouranton, and Bruno Arnaldi. 2010. A generic model for embedding users’ physical workspaces into multi-scale collaborative virtual environments. In *Proceedings of the 20th International Conference on Artificial Reality and Telexistence*. 1–9.
- [33] Rachel L. Franz, Sasa Junuzovic, and Martez Mott. 2021. Nearmi: A framework for designing point of interest techniques for VR users with limited mobility. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–14. DOI: <https://doi.org/10.1145/3441852.3471230>
- [34] Shinji Fukatsu, Yoshifumi Kitamura, Toshihiro Masaki, and Fumio Kishino. 1998. Intuitive control of “bird’s eye” overview images for navigation in an enormous virtual environment. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*. 67–76. DOI: <https://doi.org/10.1145/3441852.3471230>
- [35] Markus Funk, Florian Müller, Marco Fendrich, Megan Shene, Moritz Kolvenbach, Niclas Dobbertin, Sebastian Günther, and Max Mühlhäuser. 2019. Assessing the accuracy of point and teleport locomotion with orientation indication for virtual reality using curved trajectories. In *Proceedings of the Conference on Human Factors in Computing Systems*. 1–12. DOI: <https://doi.org/10.1145/3290605.3300377>
- [36] Wilson S. Geisler. 2008. Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology* 59, (2008), 167–192. DOI: <https://doi.org/10.1146/annurev.psych.58.110405.085632>
- [37] Kathrin Gerling, Liam Mason, and Patrick Dickinson. 2020. Virtual reality games for people using wheelchairs. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–11. DOI: <https://doi.org/10.1145/3313831.3376265>
- [38] Kathrin Gerling and Katta Spiel. 2021. A critical examination of virtual reality technology in the context of the minority body. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14. DOI: <https://doi.org/10.1145/3411764.3445196>
- [39] James J. Gibson. 1979. The Theory of Affordances. In *Proceedings of the Ecological Approach to Visual Perception*. Lawrence Erlbaum Associates, Inc., Publishers, 127–137.
- [40] Michael Gleicher and Andrew Witkin. 1992. Through-the-lens camera control. In *Proceedings of the ACM SIGGRAPH Computer Graphics*. 331–340.
- [41] Kalanit Grill-Spector, Nicholas Knouf, and Nancy Kanwisher. 2004. The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience* 7, 5 (2004), 555–562. DOI: <https://doi.org/10.1038/nn1224>

- [42] Tovi Grossman, Ravin Balakrishnan, Gordon Kurtenbach, George Fitzmaurice, Azam Khan, and Bill Buxton. 2002. Creating principal 3D curves with digital tape drawing. In *Proceedings of the Conference on Human Factors in Computing Systems*. 121–128. DOI: <https://doi.org/10.1145/503376.503398>
- [43] Uwe Gruenefeld, Abdallah El Ali, Susanne Boll, and Wilko Heuten. 2018. Beyond Halo and Wedge: Visualizing out-of-view objects on head-mounted virtual and augmented reality devices. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–11. DOI: <https://doi.org/10.1145/3229434.3229438>
- [44] Uwe Gruenefeld, Abdallah El Ali, Wilko Heuten, and Susanne Boll. 2017. Visualizing out-of-view objects in head-mounted augmented reality. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–7. DOI: <https://doi.org/10.1145/3098279.3122124>
- [45] Uwe Gruenefeld, Tim Claudius Stratmann, Abdallah El Ali, Susanne Boll, and Wilko Heuten. 2018. RadialLight: Exploring radial peripheral LEDs for directional cues in head-mounted displays. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–6. DOI: <https://doi.org/10.1145/3229434.3229437>
- [46] Abhinav Gupta, Scott Satkin, Alexei A. Efros, and Martial Hebert. 2011. From 3D scene geometry to human workspace. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 1961–1968. DOI: <https://doi.org/10.1109/CVPR.2011.5995448>
- [47] Sean Gustafson, Patrick Baudisch, Carl Gutwin, and Pourang Irani. 2008. Wedge: Clutter-free visualization of off-screen locations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 787–796. DOI: <https://doi.org/10.1145/1357054.1357179>
- [48] Edward Twitchell Hall. 1966. *The Hidden Dimension*. Garden City, NY: Doubleday.
- [49] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. DOI: <https://doi.org/10.1177/154193120605000909>
- [50] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology* 52 (1988), 139–183. DOI: [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- [51] R. H. Hess, C. L. Baker Jr., and J. Zihl. 1989. The “motion-blind” patient: low-level spatial and temporal filters. *The Journal of Neuroscience* 9, 5 (1989), 1628–1640. DOI: <https://doi.org/10.1523/JNEUROSCI.09-05-01628.1989>
- [52] Ken Hinckley, Randy Pausch, John C. Goble, and Neal F. Kassell. 1994. Passive real-world interface props for neurosurgical visualization. In *Proceedings of the Conference on Human Factors in Computing Systems*. 452–458. DOI: <https://doi.org/10.1145/191666.191821>
- [53] Teresa Hirzle, Jan Gugenheimer, Florian Geiselhart, Andreas Bulling, and Enrico Rukzio. 2019. A design space for gaze interaction on head-mounted displays. In *Proceedings of the 2019 SIGCHI Conference on Human Factors in Computing Systems*. 1–12. DOI: <https://doi.org/10.1145/3290605.3300855>
- [54] Derek Hoiem, Alexei A. Efros, and Martial Hebert. 2005. Geometric context from a single image. In *Proceedings of the IEEE International Conference on Computer Vision*. 654–661. DOI: <https://doi.org/10.1109/ICCV.2005.107>
- [55] Jonatan S. Hvass, Oliver Larsen, Kasper B. Vendelbo, Niels C. Nilsson, Rolf Nordahl, and Stefania Serafin. 2017. The effect of geometric realism on presence in a virtual reality game. In *Proceedings of the IEEE Virtual Reality*. 339–340. DOI: <https://doi.org/10.1109/VR.2017.7892315>
- [56] A. S. Ismail, M. M. Seifelmasr, and Hongxing Guo. 2018. Understanding indoor scene: Spatial layout estimation, scene classification, and object detection. In *Proceedings of the International Conference on Multimedia Systems and Signal Processing*. 64–70. DOI: <https://doi.org/10.1145/3220162.3220182>
- [57] Hiroo Iwata, Hiroaki Yano, Hiroyuki Fukushima, and Haruo Noma. 2005. CirculaFloor. *IEEE Computer Graphics and Applications* 25, 1 (2005), 64–67. DOI: <https://doi.org/10.1109/MCG.2005.5>
- [58] Hiroo Iwata, Hiroaki Yano, and Hiroshi Tomioka. 2006. Powered Shoes. In *Proceedings of the ACM SIGGRAPH Emerging Technologies*. 28-es. DOI: <https://doi.org/10.1145/1179133.1179162>
- [59] Dhruv Jain, Sasa Junuzovic, Eyal Ofek, Mike Sinclair, John Porter, Chris Yoon, Swetha MacHanavajhala, and Meredith Ringel Morris. 2021. A taxonomy of sounds in virtual reality. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference*. 160–170. DOI: <https://doi.org/10.1145/3461778.3462106>
- [60] Jason Jerald. 2015. *The VR Book: Human-Centered Design for Virtual Reality*. ACM and Morgan and Claypool. DOI: <https://doi.org/10.1145/2792790>
- [61] Jason Jerald. 2018. A taxonomy of spatial interaction patterns and techniques. *IEEE Computer Graphics and Applications* 38, 1 (2018), 11–19. DOI: <https://doi.org/10.1109/MCG.2018.011461524>
- [62] Ji Sun Kim, Denis Gračanin, Krešimir Matković, and Francis Quek. 2008. Finger Walking in Place (FWIP): A traveling technique in virtual environments. In *Proceedings of the Smart Graphics*. A. Butz, B. Fisher, A. Krüger, P. Olivier, and M. Christie (Eds.), Springer, Berlin, 58–69. DOI: [https://doi.org/10.1007/978-3-540-85412-8\\_6](https://doi.org/10.1007/978-3-540-85412-8_6)
- [63] Ji Sun Kim, Denis Gračanin, Krešimir Matković, and Francis Quek. 2009. iPhone/iPod touch as input devices for navigation in immersive virtual environments. In *Proceedings of the IEEE Virtual Reality Conference*. 261–262. DOI: <https://doi.org/10.1109/VR.2009.4811045>

- [64] Ulrike Kister. 2018. *Interactive Visualization Lenses: Natural Magic Lens Interaction for Graph Visualization*. Dresden University of Technology, Dresden. Retrieved from <http://nbn-resolving.de/urn:nbn:de:bsz:14-qucosa-236782>
- [65] Ulrike Kister, Patrick Reipschläger, and Raimund Dachsel. 2014. Multi-Touch manipulation of magic lenses for information visualization. In *Proceedings of the International Conference on Interactive Tabletops and Surfaces*. 431–434. DOI : <https://doi.org/10.1145/2669485.2669528>
- [66] Ulrike Kister, Patrick Reipschläger, Fabrice Matulic, and Raimund Dachsel. 2015. BodyLenses: Embodied magic lenses and personal territories for wall displays. In *Proceedings of the 2015 International Conference on Interactive Tabletops and Surfaces*. 117–126. DOI : <https://doi.org/10.1145/2817721.2817726>
- [67] Alexandra Kitson, Abraham M. Hashemian, Ekaterina R. Stepanova, Ernst Kruijff, and Bernhard E. Riecke. 2017. Comparing leaning-based motion cueing interfaces for virtual reality locomotion. In *Proceedings of the IEEE Symposium on 3D User Interfaces*. 73–82. DOI : <https://doi.org/10.1109/3DUI.2017.7893320>
- [68] Regis Kopper, Tao Ni, Doug A. Bowman, and Marcio Pinho. 2006. Design and evaluation of navigation techniques for multiscale virtual environments. In *Proceedings of the IEEE Virtual Reality Conference*. 175–182. DOI : <https://doi.org/10.1109/VR.2006.47>
- [69] Eike Langbehn, Gerd Bruder, and Frank Steinicke. 2016. Scale matters! Analysis of dominant scale estimation in the presence of conflicting cues in multi-scale collaborative virtual environments. In *Proceedings of the IEEE Symposium on 3D User Interfaces*. 211–220. DOI : <https://doi.org/10.1109/3DUI.2016.7460054>
- [70] Daniel Lange, Tim Claudius Stratman, Uwe Gruenfeld, and Susanne Boll. 2020. HiveFive: Immersion preserving attention guidance in virtual reality. In *Proceedings of the 2020 SIGCHI Conference on Human Factors in Computing Systems*. 1–13. DOI : <https://doi.org/10.1145/3313831.3376803>
- [71] Joseph J. LaViola Jr., Daniel Acevedo Feliz, Daniel F. Keefe, and Robert C. Zeleznik. 2001. Hands-free multi-scale navigation in virtual environments. In *Proceedings of the 2001 Symposium on Interactive 3D Graphics*. 9–15. DOI : <https://doi.org/10.1145/364338.364339>
- [72] Joseph J. LaViola Jr., Ernst Kruijff, Ryan P. McMahan, Doug Bowman, and Ivan P. Poupyrev. 2017. *3D User Interfaces: Theory and Practice*. Addison-Wesley Professional.
- [73] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. 2006. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2169–2178. DOI : <https://doi.org/10.1109/CVPR.2006.68>
- [74] Fei-Fei Li, Asha Iyer, Christof Koch, and Pietro Perona. 2007. What do we perceive in a glance of a real-world scene? *Journal of Vision* 7, 1 (2007), 10. DOI : <https://doi.org/10.1167/7.1.10>
- [75] Fei-Fei Li and P. Perona. 2005. A Bayesian hierarchical model for learning natural scene categories. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 524–531. DOI : <https://doi.org/10.1109/CVPR.2005.16>
- [76] Yung Ta Lin, Yi Chi Liao, Shan Yuan Teng, Yi Ju Chung, Liwei Chan, and Bing Yu Chen. 2017. Outside-in: Visualizing out-of-sight regions-of-interest in a 360° video using spatial picture-in-picture previews. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. 255–265. DOI : <https://doi.org/10.1145/3126594.3126656>
- [77] Julian Looser, Mark Billinghurst, and Andy Cockburn. 2004. Through the looking glass: The use of lenses as an interface tool for augmented reality interfaces. In *Proceedings of the International Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia*. 204–211. DOI : <https://doi.org/10.1145/988834.988870>
- [78] Weizhou Luo, Eva Goebel, Patrick Reipschläger, Mats Ole Ellenberg, and Raimund Dachsel. 2021. Exploring and slicing volumetric medical data in augmented reality using a spatially-aware mobile device. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality Adjunct*. 1–6. DOI : <https://doi.org/10.1109/ISMAR-Adjunct54149.2021.00076>
- [79] Jock Mackinlay, Stuart K. Card, and George G. Robertson. 1990. A semantic analysis of the design space of input devices. *Human-Computer Interaction* 5, 2–3 (1990), 145–190. DOI : <https://doi.org/10.1080/07370024.1990.9667153>
- [80] Sven Mayer, Jens Reinhardt, Robin Schweigert, Brighten Jelke, Valentin Schwind, Katrin Wolf, and Niels Henze. 2020. Improving humans’ ability to interpret deictic gestures in virtual reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14. DOI : <https://doi.org/10.1145/3313831.3376340>
- [81] Michael J. McGuffin, Liviu Tancu, and Ravin Balakrishnan. 2003. Using deformations for browsing volumetric data. In *Proceedings of the IEEE Visualization*. 401–408. DOI : <https://doi.org/10.1109/VISUAL.2003.1250400>
- [82] Tim Menzner, Travis Gesslein, Alexander Otte, and Jens Grubert. 2020. Above surface interaction for multiscale navigation in mobile virtual reality. In *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces*. 372–381. DOI : <https://doi.org/10.48550/arXiv.2002.03037>
- [83] Jason D. Moss and Eric R. Muth. 2011. Characteristics of head-mounted displays and their effects on simulator sickness. *Human Factors* 53, 3 (2011), 308–319. DOI : <https://doi.org/10.1177/0018720811405196>
- [84] Martez Mott, Edward Cutrell, Mar Gonzalez Franco, Christian Holz, Eyal Ofek, Richard Stokley, and Meredith Ringel Morris. 2019. Accessible by design: An opportunity for virtual reality. In *Proceedings of the IEEE International*

- Symposium on Mixed and Augmented Reality Adjunct*. 451–454. DOI : <https://doi.org/10.1109/ISMAR-Adjunct.2019.00122>
- [85] Konrad Mühler, Mathias Neugebauer, Christian Tietjen, and Bernhard Preim. 2007. Viewpoint selection for intervention planning. In *Proceedings of the IEEE-VGTC Symposium on Visualization*. 1–8. DOI : <https://doi.org/10.2312/VisSym/EuroVis07/267-274>
- [86] Luan Le Ngoc and Roy S. Kalawsky. 2013. Evaluating usability of amplified head rotations on base-to-final turn for flight simulation training devices. In *Proceedings of the IEEE Virtual Reality*. 51–54. DOI : <https://doi.org/10.1109/VR.2013.6549359>
- [87] Jakob Nielsen. 1994. *Usability Engineering*. Morgan Kaufmann.
- [88] Jakob Nielsen. 1994. Heuristic evaluation. In *Proceedings of the Usability Inspection Methods*. John Wiley and Sons, New York, 25–64.
- [89] Donald A. Norman. 1999. Affordance, conventions, and design. *Interactions* 6, 3 (1999), 38–43. DOI : <https://doi.org/10.1145/301153.301168>
- [90] Donald A. Norman. 2013. *The Design of Everyday Things: Revised and Expanded Edition*. Basic Books.
- [91] Donald A. Norman. 2018. *Affordances and Design*. jnd.org. Retrieved from [https://jnd.org/affordances\\_and\\_design/](https://jnd.org/affordances_and_design/)
- [92] Aude Oliva. 2005. Gist of the scene. In *Proceedings of the Neurobiology of Attention*. Laurent Itti, Geraint Rees, and John K. Tsotsos (Eds.), Academic Press, 251–256. DOI : <https://doi.org/10.1016/B978-012375731-9/50045-8>
- [93] Aude Oliva and Antonio Torralba. 2001. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision* 42, 3 (2001), 145–175. DOI : <https://doi.org/10.1023/A:1011139631724>
- [94] Aude Oliva and Antonio Torralba. 2002. Scene-centered description from spatial envelope properties. In *Proceedings of the Biologically Motivated Computer Vision. BMCV 2002*. Heinrich H. Bülthoff, Christian Wallraven, Seong-Whan Lee, and Tomaso A. Poggio (Eds.), Lecture Notes in Computer Science, Springer, Berlin, 263–272. DOI : [https://doi.org/10.1007/3-540-36181-2\\_26](https://doi.org/10.1007/3-540-36181-2_26)
- [95] Aude Oliva and Antonio Torralba. 2006. Building the gist of a scene: The role of global image features in recognition. In *Proceedings of the Progress in Brain Research*. S. Martinez-Conde, S. L. Macknik, L. M. Martinez, J. -M. Alonso, and P. U. Tse (Eds.), Elsevier, 23–36. DOI : [https://doi.org/10.1016/S0079-6123\(06\)55002-2](https://doi.org/10.1016/S0079-6123(06)55002-2)
- [96] Oyewole Oyekoya, William Steptoe, and Anthony Steed. 2009. A saliency-based method of simulating visual attention in virtual scenes. In *Proceedings of the Virtual Reality Software and Technology*. 199–206. DOI : <https://doi.org/10.1145/1643928.1643973>
- [97] Genevieve Patterson, Chen Xu, Hang Su, and James Hays. 2014. The SUN attribute database: Beyond categories for deeper scene understanding. *International Journal of Computer Vision* 108, 1–2 (2014), 59–81. DOI : <https://doi.org/10.1007/s11263-013-0695-z>
- [98] Amy Pavel, Björn Hartmann, and Maneesh Agrawala. 2017. Shot orientation controls for interactive cinematography with 360° video. In *Proceedings of the User Interface Software and Technology*. 289–297. DOI : <https://doi.org/10.1145/3126594.3126636>
- [99] Julian Petford, Iain Carson, Miguel A. Nacenta, and Carl Gutwin. 2019. A comparison of guiding techniques for out-of-view objects in full-coverage displays. In *Proceedings of the 2019 SIGCHI Conference on Human Factors in Computing Systems*. 1–13. DOI : <https://doi.org/10.1145/3290605.3300288>
- [100] Jeffrey S. Pierce and Randy Pausch. 2004. Navigation with place representations and visible landmarks. In *Proceedings of the IEEE Virtual Reality*. 173–288. DOI : <https://doi.org/10.1109/VR.2004.1310071>
- [101] Thammathip Piumsomboon, Gun A. Lee, Barrett Ens, Bruce H. Thomas, and Mark Billinghurst. 2018. Superman vs. giant: A study on spatial perception for a multi-scale mixed reality flying telepresence interface. *IEEE Transactions on Visualization and Computer Graphics* 24, 11 (2018), 2974–2982. DOI : <https://doi.org/10.1109/TVCG.2018.2868594>
- [102] Thammathip Piumsomboon, Gun A. Lee, Andrew Irlitti, Barrett Ens, Bruce H. Thomas, and Mark Billinghurst. 2019. On the shoulder of the giant: A multi-scale mixed reality collaboration with 360 video sharing and tangible interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–17. DOI : <https://doi.org/10.1145/3290605.3300458>
- [103] Bernhard Preim, Rainer Michel, Knut Hartmann, and Thomas Strothotte. 1998. Figure captions in visual interfaces. In *Proceedings of the Conference on Graphics Interface*. 235–246. DOI : <https://doi.org/10.1145/948496.948528>
- [104] Bernhard Preim, Andreas Raab, and Thomas Strothotte. 1997. Coherent zooming of illustrations with 3D-graphics and text. In *Proceedings of the Working Conference on Advanced Visual Interfaces*. 105–113.
- [105] Bernhard Preim, Alf Ritter, Thomas Strothotte, Tilo Pohle, Lyn Bartram, and David A. Forsey. 1995. Consistency of rendered images and their textual labels. In *Proceedings of the CompuGraphics*. 201–210.
- [106] Eric D. Ragan, Siroberto Scerbo, Felipe Bacim, and Doug A. Bowman. 2017. Amplified head rotation in virtual reality and the effects on 3D search, training transfer, and spatial orientation. *IEEE Transactions on Visualization and Computer Graphics* 23, 8 (2017), 1880–1895. DOI : <https://doi.org/10.1109/TVCG.2016.2601607>

- [107] Ismo Rakkolainen, Roope Raisamo, Matthew Turk, and Tobias Höllerer. 2017. Field-of-view extension for VR viewers. In *Proceedings of the 21st International Academic Mindtrek Conference*. 1–4. DOI: <https://doi.org/10.1145/3131085.3131088>
- [108] Timo Ropinski, Frank Steinicke, and Klaus Hinrichs. 2005. A constrained road-based VR navigation technique for travelling in 3D city models. In *Proceedings of the 2005 International Conference on Augmented Tele-existence*. 228–235. DOI: <https://doi.org/10.1145/1152399.1152441>
- [109] Olli Rummukainen and Catarina Mendonça. 2016. Reproducing reality: Multimodal contributions in natural scene discrimination. *ACM Transactions on Applied Perception* 14, 1 (2016), 1–10. DOI: <https://doi.org/10.1145/2915917>
- [110] Olli Rummukainen, Jenni Radun, Toni Virtanen, and Ville Pulkki. 2014. Categorization of natural dynamic audiovisual scenes. *PLoS ONE* 9, 5 (2014), 1–14. DOI: <https://doi.org/10.1371/journal.pone.0095848>
- [111] Shyam Prathish Sargunam, Kasra Rahimi Moghadam, Mohamed Suhail, and Eric D. Ragan. 2017. Guided head rotation and amplified head rotation: Evaluating semi-natural travel and viewing techniques in virtual reality. In *Proceedings of the IEEE Virtual Reality*. 19–28. DOI: <https://doi.org/10.1109/VR.2017.7892227>
- [112] Shyam Prathish Sargunam and Eric D. Ragan. 2018. Evaluating joystick control for view rotation in virtual reality with continuous turning, discrete turning, and field-of-view reduction. In *Proceedings of the International Workshop on Interactive and Spatial Computing*. 74–79. DOI: <https://doi.org/10.1145/3191801.3191815>
- [113] Scott Satkin and Martial Hebert. 2013. 3DNN: Viewpoint invariant 3D geometry matching for scene understanding. In *Proceedings of the International Conference on Computer Vision*. 1873–1880. DOI: <https://doi.org/10.1109/ICCV.2013.235>
- [114] Scott Satkin, Jason Lin, and Martial Hebert. 2012. Data-driven scene understanding from 3D models. In *Proceedings of the British Machine Vision Conference*. 1–11.
- [115] Philippe G. Schyns and Aude Oliva. 1994. From blobs to boundary edges: Evidence for time-and spatial-scale-dependent scene recognition. *Psychological Science* 5, 4 (1994), 195–200. DOI: <https://doi.org/10.1111/j.1467-9280.1994.tb00500.x>
- [116] Andrew Sears, Min Lin, Julie Jacko, and Yan Xiao. 2003. When computers fade: Pervasive computing and situationally-induced impairments and disabilities. *HCI International* 2, 3 (2003), 1298–1302.
- [117] Karan Singh. 2002. A fresh perspective. In *Proceedings of the Graphics Interface*. 17–24. DOI: <https://doi.org/10.20380/GI2002.03>
- [118] Mel Slater, Anthony Steed, and Martin Usoh. 1995. The virtual treadmill: A naturalistic metaphor for navigation in immersive virtual environments. In *Proceedings of the Selected Papers of the Eurographics Workshops on Virtual Environments*. 135–148. DOI: [https://doi.org/10.1007/978-3-7091-9433-1\\_12](https://doi.org/10.1007/978-3-7091-9433-1_12)
- [119] Mel Slater, Martin Usoh, and Anthony Steed. 1994. Depth of presence in virtual environments. *Presence: Teleoperators and Virtual Environments* 3, 2 (1994), 130–144. DOI: <https://doi.org/10.1162/pres.1994.3.2.130>
- [120] Dmitry Sokolov and Dimitri Plemenos. 2007. High level methods for scene exploration. *Journal of Virtual Reality and Broadcasting* 3, 12 (2007), 1–12. DOI: <https://doi.org/10.20385/1860-2037/3.2006.12>
- [121] Dmitry Sokolov, Dimitri Plemenos, and Karim Tamine. 2006. Viewpoint quality and global scene exploration strategies. In *Proceedings of the 1st International Conference on Computer Graphics Theory and Applications*. 184–191. DOI: <https://doi.org/10.5220/0001350101840191>
- [122] Henry Sonnet, Sheelagh Carpendale, and Thomas Strothotte. 2004. Integrating expanding annotations with a 3D explosion probe. In *Proceedings of the Working Conference on Advanced Visual Interfaces*. 63–70. DOI: <https://doi.org/10.1145/989863.989871>
- [123] Frank Steinicke, Gerd Bruder, Jason Jerald, Harald Frenz, and Markus Lappe. 2010. Estimation of detection thresholds for redirected walking techniques. *IEEE Transactions on Visualization and Computer Graphics* 16, 1 (2010), 17–27. DOI: <https://doi.org/10.1109/TVCG.2009.62>
- [124] Richard Stoakley, Matthew J. Conway, and Randy Pausch. 1995. Virtual reality on a WIM: Interactive worlds in miniature. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 265–272. DOI: <https://doi.org/10.1145/223904.223938>
- [125] Stanislav L. Stoev and Dieter Schmalstieg. 2002. Application and taxonomy of through-the-lens techniques. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*. 57–64. DOI: <https://doi.org/10.1145/585740.585751>
- [126] Mengu Sukan, Carmine Elvezio, Ohan Oda, Steven Feiner, and Barbara Tversky. 2014. ParaFrustum: Visualization techniques for guiding a user to a constrained set of viewing positions and orientations. In *Proceedings of the 27th annual ACM Symposium on User Interface Software and Technology*. 331–340. DOI: <https://doi.org/10.1145/2642918.2647417>
- [127] Hemant Bhaskar Surale, Aakar Gupta, Mark Hancock, and Daniel Vogel. 2019. TabletInVR: Exploring the design space for using a multi-touch tablet in virtual reality. In *Proceedings of the 2019 SIGCHI Conference on Human Factors in Computing Systems*. 1–13. DOI: <https://doi.org/10.1145/3290605.3300243>

- [128] Ivan E. Sutherland. 1968. A head-mounted three dimensional display. In *Proceedings of the Seminal Graphics: Pioneering Efforts that Shaped the Field*. 757–764. DOI : <https://doi.org/10.1145/1476589.1476686>
- [129] Shigeo Takahashi, Kenichi Yoshida, Kenji Shimada, and Tomoyuki Nishita. 2006. Occlusion-free animation of driving routes for car navigation systems. *IEEE Transactions on Visualization and Computer Graphics* 12, 5 (2006), 1141–1148. DOI : <https://doi.org/10.1109/TVCG.2006.167>
- [130] Theresa Jean Tanenbaum, Nazely Hartoonian, and Jeffrey Bryan. 2020. "How do I make this thing smile?": An inventory of expressive nonverbal communication in commercial social virtual reality platforms. In *Proceedings of the 2020 SIGCHI Conference on Human Factors in Computing Systems*. 1–13. DOI : <https://doi.org/10.1145/3313831.3376606>
- [131] James N. Templeman, Patricia S. Denbrook, and Linda E. Sibert. 1999. Virtual locomotion: Walking in place through virtual environments. *Presence: Teleoperators and Virtual Environments* 8, 6 (1999), 598–617. DOI : <https://doi.org/10.1162/105474699566512>
- [132] Simon Thorpe, Denis Fize, and Catherine Marlot. 1996. Speed of processing in the human visual system. *Nature* 381 (1996), 520–522. DOI : <https://doi.org/10.1038/381520a0>
- [133] Shari Trewin. 2000. Configuration Agents, Control and Privacy. In *Proceedings of the Conference on Universal Usability*. 9–16. DOI : <https://doi.org/10.1145/355460.355466>
- [134] Daniel R. Trindade and Alberto B. Raposo. 2011. Improving 3D navigation in multiscale environments using cubemap-based techniques. In *Proceedings of the 2011 ACM Symposium on Applied Computing*. 1215–1221. DOI : <https://doi.org/10.1145/1982185.1982454>
- [135] Barbara Tversky and Kathleen Hemenway. 1983. Categories of Environmental Scenes. *Cognitive Psychology* 15, 1 (1983), 121–149. DOI : [https://doi.org/10.1016/0010-0285\(83\)90006-3](https://doi.org/10.1016/0010-0285(83)90006-3)
- [136] Jeff C. Valentine, Larry V. Hedges, and Harris M. Cooper. 2009. *The Handbook of Research Synthesis and Meta-Analysis*. Russell Sage Foundation.
- [137] Dimitar Valkov, Frank Steinicke, Gerd Bruder, and Klaus H. Hinrichs. 2010. Traveling in 3D virtual environments with foot gestures and a multi-touch enabled WIM. In *Proceedings of the Virtual Reality International Conference*. 171–180.
- [138] Eduardo Veas, Raphael Grasset, Ernst Kruijff, and Dieter Schmalstieg. 2012. Extended overview techniques for outdoor augmented reality. *IEEE Transactions on Visualization and Computer Graphics* 18, 4 (2012), 565–572. DOI : <https://doi.org/10.1109/TVCG.2012.44>
- [139] John Viega, Matthew J. Conway, George Williams, and Randy Pausch. 1996. 3D magic lenses. In *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology*. 51–58. DOI : <https://doi.org/10.1145/237091.237098>
- [140] Julia Vogel and Bernt Schiele. 2007. Semantic modeling of natural scenes for content-based image retrieval. *International Journal of Computer Vision* 72, 2 (2007), 133–157. DOI : <https://doi.org/10.1007/s11263-006-8614-1>
- [141] Colin Ware and Steven Osborne. 1990. Exploration and virtual camera control in virtual three-dimensional environments. In *Proceedings of the 1990 Symposium on Interactive 3D Graphics*. 175–183. DOI : <https://doi.org/10.1145/91394.91442>
- [142] Julie R. Williamson, Mark McGill, and Khari Outram. 2019. PlaneVR: Social acceptability of virtual reality for aeroplane passengers. In *Proceedings of the Conference on Human Factors in Computing Systems*. 1–14. DOI : <https://doi.org/10.1145/3290605.3300310>
- [143] Chadwick A. Wingrave, Yonca Haciahmetoglu, and Doug A. Bowman. 2006. Overcoming world in miniature limitations by a scaled and scrolling WIM. In *Proceedings of the IEEE Symposium on 3D User Interfaces*. 1–6. DOI : <https://doi.org/10.1109/VR.2006.106>
- [144] Jacob O. Wobbrock. 2019. Situationally aware mobile devices for overcoming situational impairments. In *Proceedings of the Symposium on Engineering Interactive Computing Systems*. 1–18. DOI : <https://doi.org/10.1145/3319499.3330292>
- [145] Jianxiong Xiao, James Hays, Krista A. Ehinger, Aude Oliva, and Antonio Torralba. 2010. SUN database: Large-scale scene recognition from abbey to zoo. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 3485–3492. DOI : <https://doi.org/10.1109/CVPR.2010.5539970>
- [146] Robert Xiao and Hrvoje Benko. 2016. Augmenting the field-of-view of head-mounted displays with sparse peripheral displays. In *Proceedings of the 2016 SIGCHI Conference on Human Factors in Computing Systems*. 1221–1232. DOI : <https://doi.org/10.1145/2858036.2858212>
- [147] Wataru Yamada and Hiroyuki Manabe. 2016. Expanding the field-of-view of head-mounted displays with peripheral blurred images. In *Adjunct Proceedings of the 29th Annual ACM Symposium on User Interface Software and Technology*. 141–142. DOI : <https://doi.org/10.1145/2984751.2985735>
- [148] Momona Yamagami, Sasa Junuzovic, Mar Gonzalez-Franco, Eyal Ofek, Edward Cutrell, John R. Porter, Andrew D. Wilson, and Martez E. Mott. 2022. Two-In-One: A design space for mapping unimanual input into bimanual

- interactions in VR for users with limited movement. *ACM Transactions on Accessible Computing* 15, 3 (2022), 1–25. DOI: <https://doi.org/10.1145/3510463>
- [149] Zhixin Yan, Robert W. Lindeman, and Arindam Dey. 2016. Let your fingers do the walking: A unified approach for efficient short-, medium-, and long-distance travel in VR. In *Proceedings of the IEEE Symposium on 3D User Interfaces*. 27–30. DOI: <https://doi.org/10.1109/3DUI.2016.7460027>
- [150] Alfred L. Yarbus. 1968. *Eye Movements and Vision*. Springer New York. DOI: <https://doi.org/10.1007/978-1-4899-5379-7>
- [151] Kiwon Yun, Yifan Peng, Dimitris Samaras, Gregory J. Zelinsky, and Tamara L. Berg. 2013. Exploring the role of gaze behavior and object detection in scene understanding. *Frontiers in Psychology* 4 (2013), 1–14. DOI: <https://doi.org/10.3389/fpsyg.2013.00917>
- [152] Majed Al Zayer, Paul MacNeilage, and Eelke Folmer. 2020. Virtual locomotion: A survey. *IEEE Transactions on Visualization and Computer Graphics* 26, 6 (2020), 2315–2334. DOI: [10.3389/fpsyg.2013.00954](https://doi.org/10.3389/fpsyg.2013.00954)
- [153] Gregory J. Zelinsky. 2013. Understanding scene understanding. *Frontiers in Psychology* 4 (2013), 1–3.
- [154] Xiaolong Zhang and George W. Furnas. 2005. MCVEs: Using cross-scale collaboration to support user interaction with multiscale structures. *Presence* 14, 1 (2005), 31–46. DOI: <https://doi.org/10.1162/1054746053890288>
- [155] Yuhang Zhao, Cynthia L. Bennett, Hrvoje Benko, Edward Cutrell, Christian Holz, Meredith Ringel Morris, and Mike Sinclair. 2018. Enabling people with visual impairments to navigate virtual reality with a haptic and auditory cane simulation. In *Proceedings of the 2018 SIGCHI Conference on Human Factors in Computing Systems*. 1–4. DOI: <https://doi.org/10.1145/3170427.3186485>
- [156] Yuhang Zhao, Meredith Ringel Morris, Edward Cutrell, Christian Holz, and Andrew D. Wilson. 2019. SeeingVR: A set of tools to make virtual reality more accessible to people with low vision. In *Proceedings of the 2019 SIGCHI Conference on Human Factors in Computing Systems*. 1–14. DOI: <https://doi.org/10.1145/3290605.3300341>

Received 22 December 2020; revised 10 October 2022; accepted 16 April 2023