Scaling Techniques for Exocentric Navigation Interfaces in Multiscale Virtual Environments

Jong-in Lee, Wolfgang Stuerzlinger

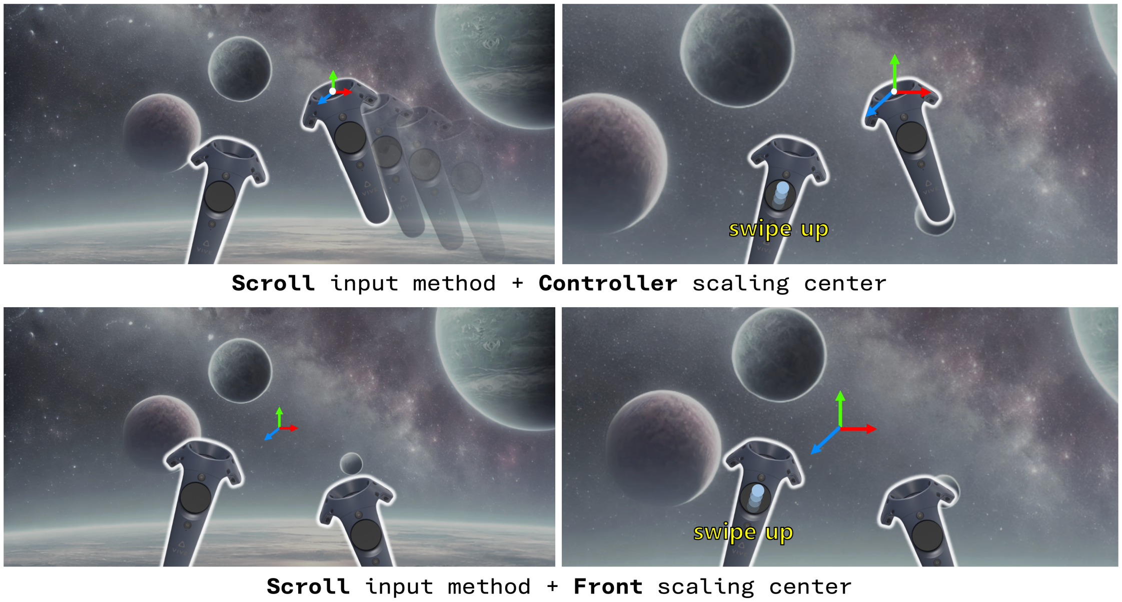

Fig. 1: Two novel scaling techniques for exocentric navigation in multiscale virtual environments, each using a different method to specify the scaling center (CONTROLLER, FRONT), all evaluated in our user study in comparison with the state-of-the-art two-handed pinch/grab-based technique (TWO-HANDED).

Abstract— Navigating multiscale virtual environments necessitates an interaction method to travel across different levels of scale (LoS). Prior research has studied various techniques that enable users to seamlessly adjust their scale to navigate between different LoS based on specific user contexts. We introduce a scroll-based scale control method optimized for exocentric navigation, targeted at scenarios where speed and accuracy in continuous scaling are crucial. We pinpoint the challenges of scale control in settings with multiple LoS and evaluate how distinct designs of scaling techniques influence navigation performance and usability. Through a user study, we investigated two pivotal elements of a scaling technique: the input method and the scaling center. Our findings indicate that our scroll-based input method significantly reduces task completion time and error rate and enhances efficiency compared to the most frequently used bi-manual method. Moreover, we found that the choice of scaling center affects the ease of use of the scaling method, especially when paired with specific input methods.

Index Terms—Scaling technique, Interaction technique, VR navigation

1 INTRODUCTION

Multiscale Virtual Environments (MVEs) are useful for a variety of real-world VR application scenarios, including medical, astrophysical, geoinformation, and entertainment applications [14, 19, 20, 40]. In an MVE, objects occur at different *Levels of Scale* (LoS), which necessitate users to frequently adapt their scale to be able to interact with these objects of various sizes, e.g., from mountains to molehills or even a larger range such as from an atom to a galaxy cluster [36, 37]. Researchers have explored different navigation interfaces to facilitate traveling across widely varying scales in MVEs [2, 10, 35, 68]. When designing such interfaces, designers can choose between two paradigms: ego-centric vs. exo-centric navigation. Ego-centric navigation provides a mechanism to move a user relative to their own viewpoint. The most prominent example is flying. This type of navigation is known to induce a high level of simulator sickness and cause frequent navigation errors such as over- or undershoots [62]. To address this issue, flying techniques that minimize the navigation error through automatic speed control [2] and target-based navigation techniques with optical flow cues to mitigate simulator sickness [35] have been presented. While ego-centric navigation allows users to explore the environment from a 1st-person perspective with a high level of immersion [18, 25], it is not ideal for some use cases. One example is a VR authoring tool [21, 50], in which precise control over a user's viewpoint and scale might be crucial, e.g., to see how accurately various objects line up. In this case, exo-centric navigation is more appropriate so that the user can precisely control the viewpoint in the environment from a 3rd-person perspective. One of the important design components for designing an exo-centric navigation interface is scale control.

- *Jong-in Lee is with the College of Performance, Visualization, and Fine Arts at Texas A&M University. E-mail: jongin@tamu.edu.*
- *Wolfgang Stuerzlinger is with the School of Interactive Arts and Technology at Simon Fraser University. E-mail: w.s@sfu.ca.*

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

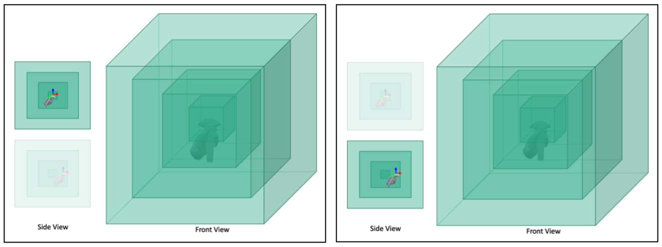

Fig. 2: A depth ambiguity occurs when an object, a controller in our case, is inside a structure hierarchically nested with multiple objects rendered transparently. Pointing the controller inside the innermost cube (left figure) or right behind the inner cube (right figure) both result in images that look very similar from the user's viewpoint, i.e., front view, if transparency is used.

For MVE navigation, previous work has presented different types of scaling techniques (see section 2), which either automatically or manually control the scale of the user with different gestures, input methods, 2D/3D widgets, and scaling centers. Satriadi et al. developed a multiscale navigation technique for navigating 2D maps quickly and precisely in an Augmented Reality (AR) environment [55]. However, their work was limited to navigating 2D space and navigation constrained to the surface of the model, which does not permit navigating freely in 3D space, e.g., between stars in a galaxy or cells at a microscopic level. Moreover, they used only a fixed scaling center on the surface and did not consider other choices for scaling centers. Work by Cho et al. [10] and Yang et al. [67] studied unconstrained 3D navigation in immersive environments, but their techniques were limited to the scenario of navigating within a single open space. Further, the 7-DOF navigation technique by Cho et al. does not permit a user to enter an object since the technique itself prevents such behavior by limiting the scaling center to the surface. Their surface-based approach does not allow users to locate the scaling center inside an object (such as the globe in their work) to zoom into a smaller LoS. In our work, we study scaling techniques to address the unique challenges of navigating in a more general type of multiscale environment, in which a small LoS can be nested within multiple larger LoS, i.e., we deal with MVEs that contain nested structures. Such MVEs appear in many different fields, including the human body in medicine [14], engineering [61], and games [13, 19]. To enable navigation across different LoS in such environments, entering and exiting objects (e.g., entering an organ in a human body, a tumor cell in an organ, etc.) should be easily possible, ideally with minimal discomfort.

The bi-manual pinch/grab-based scale control method [42], which is arguably the most commonly used scale control method in VR painting and authoring software [21, 45, 46, 50] and research [10, 42, 67], is an easy-to-learn, intuitive gesture-based interaction that allows users to control the scale of the given environment (or equivalently the user). Even though this method has been employed in many VR systems, it has inherent limitations.

Bi-manual *manipulation* is known to enhance the accuracy of interactive tasks [26]. However, such manipulation requires continued and extensive movement of both arms and hands, resulting in significant physical exertion. This limitation of bi-manual interaction makes it less suitable for scenarios requiring constant scale adjustments, as users must then potentially repetitively perform large movements throughout the interaction. Moreover, we believe that the bi-manual pinch/grab-based *navigation* method is not ideal for navigation in MVEs with nested structures. Specifically, the bi-manual method suffers from a challenge associated with specifying the scaling center, as the technique encourages erroneous behaviors that make it harder to get to a desired view, due to failures to set an appropriate scaling center for entering an object, resulting in its previously documented enlarge-and-push-away behavior [10]. Yet, entering an object via scaling is a common navigation task in MVEs with nested structures, where each LoS is nested within a larger LoS Figure 8. To enter an object by "scaling space", i.e., "expanding the environment", users need to be able to put the scaling center quite accurately at the intended target location, which is *inside* an object and thus typically invisible. Assume the user is standing in front of a surface/wall of an object they intend to enter. As the scaling center for bi-manual pinch/grab is typically the midpoint between the controllers, the user has to put both controllers *into* the wall to transition into the object. As the controllers then typically disappear into the wall, this is non-intuitive. Even when the user is large and plans to scale into a small object on a table, the fact that the scaling center is not visible in the object makes it challenging to aim accurately "into" the object, as documented in previous work [10]. In a pilot study for our work, we confirmed that users struggled much with such issues and – even after repeated explanations – could not reliably navigate within a nested MVE.

One often-proposed solution, the option of making it easier to view everything in the environment by making overlapping objects semitransparent, creates depth ambiguities, due to the multiple overlapping semi-transparent layers created by the nested object structure (see Figure 2), which makes it hard to point to a specific location. In such scenarios, the use of bi-manual pinch/grab techniques then again leads often to errors, notably inadvertent "zooming" into an incorrect LoS [10], which then requires corrective navigation actions. This not only hampers the efficiency of the process but also causes delays in navigating to the target LoS. Finally, the bi-manual pinch/grab method is not ideal for extended usage, as it requires fairly large movements with both arms and hands for pinching/grabbing space. To travel to a distant location at a LoS far from the current LoS, constant clutching is then necessary, which further increases physical fatigue.

To study the above issues, we conducted a comparative evaluation of different scaling techniques, including a novel scroll-based technique (Figure 1), examining design components of scaling techniques for MVEs. In our user study, we assessed these approaches in terms of their navigational efficiency and overall usability.

2 Related Work

In this section, we review previous work on bi-manual interaction techniques and navigation interfaces for MVEs.

2.1 Exocentric Navigation in Multiscale Virtual Environments

Since Sutherland presented the "Sketchpad" system [60], which allows direct manipulations of objects on the screen and controlling the scale of the content, pan-and-zoom navigation has made great advances. Perlin and Fox [52] presented the Pad interface that displays all interactable objects within a large 2D space while also enabling users to have a closer look, through portals and semantic zoom to view hierarchically embedded information in each object. Bederson and Hollan [7] developed Pad++ which uses a more intuitive zooming interface that lets users directly zoom in or out around the mouse cursor. As touchpad and touchscreen technology advanced, numerous types of direct pan-and-zoom techniques have been presented. Hinckley et al. [28] presented a two-handed pinch-like gesture and panning for navigation, which has become the de-facto standard 2D navigation technique in multi-touch interaction. Subsequently, numerous variations of such pan-and-zoom and zooming techniques have been presented [1, 4, 5, 8, 17, 29, 41, 48]. The concept of pan-and-zoom navigation was also explored on large displays [47, 57, 59].

Researchers also adapted these 2D pan-and-zoom techniques to exocentric 3D navigation through direct manipulation in MVEs. Arguably, Mapes and Moshell [42] were the first to introduce a bi-manual grab-and-scale gesture to "grab space" to move. Henry et al.'s exocentric navigation method used a wand to rotate the environment around a selected object and two hands to scale around a selected node [27]. Pai et al. presented PinchMove, an exocentric navigation technique that enables orientation control, similar to orbiting, with rotating gestures that use two-handed and movement control by grabbing a point in space and dragging it towards the user's viewpoint [51]. Yang et al. studied the presence of an overview and a bi-manual zooming (i.e. scaling) technique for understanding 3D scatterplot data in VR [67]. Austin et al. performed an elicitation study to identify which hand and foot gestures were perceived most useful for exocentric map navigation in AR [3].

The previous works above adapting pan-and-zoom techniques to exocentric navigation did not investigate their application in multi-scale environments containing spatial representations at different scales. Notable exceptions include Satriadi et al.'s research on freehand gesture-based technique for navigating planar geographical data in AR [55] and Cho et al.'s [10] evaluation of automatic scaling and translation techniques in comparison to one-handed manual scaling for navigating a 3D globe model. However, these studies constrained user navigation to the surface of a single model—either planar or spherical—limiting interaction to 2D spatial data mapped onto that surface. Our work explores scaling techniques for exocentric 3D navigation interfaces in more complex, nested MVEs where each LoS is nested within a larger LoS. Such a 3D environment requires unconstrained navigation independent of surfaces, enabling exploration of open spaces and entering or exiting objects to travel across LoS without the availability of traditional pan-and-zoom approaches typically used for planar map navigation.

2.2 Scale Control in Multiscale Virtual Environments

In 3D and AR/VR research, scale control has been approached in different ways. A prevalent approach involves predetermining the scale for distinct regions, subsequently allowing the system to automatically adjust the environmental scale as users traverse into such regions. Pierce et al. [53], for instance, devised a method to establish a hierarchy within an MVE, assigning a specific scale to each location, represented by a semantic symbol, and then facilitating seamless transitions of the user's position and scale between the current and selected locations. Extending a target-based navigation technique [31], Bacim et al. [6] proposed a similar concept, termed HiSMap (Hierarchically-Structured Map), comprising representations of all MVE regions with predetermined scales, and enabling automatic transitions from the current LoS to the target LoS. Conversely, Li et al. [38] introduced a power-law-based technique for a World-in-Miniature (WIM) within an astrophysical setting, mapping mouse/keyboard input exponentially. This empowered users to control a third-person perspective through the WIM and navigate freely at a designated power scale, while linearly scaling the physical dataset relative to the current scale. Similarly, Kouvil et al. [32] introduced a navigation technique employing labels for each LoS, enabling users to traverse different sizes of molecular 3D models by simply clicking on these labels. Additionally, several studies have focused on automatic scale adjustment between two view modes: a normal view at a human scale and an overview at a giant scale, contingent upon the user's chosen navigation mode [12, 22, 33, 69].

An alternative approach to scale control involves the automatic adjustment of scale based on the current context of the environment. McCrae [43] presented a technique to dynamically modify the near and far planes of the viewing frustum. This method computes the minimal distance through a cubemap, which then determines the viewer's scale and movement speed. Argelaguet and Maignant [2] considered not only the smallest distance in the z-buffer but also the optical flow of pixels within the user's field of view. Cho et al. [10] also used the minimal distance in the z-buffer to find the closest scene point in the stereo display and utilized the cyclopean scale to guarantee that the nearest point can be fused in stereo vision.

Another approach lets users manually control the scale of the environment through a scaling technique. Robinett and Holloway [54] suggested a manual scale control technique to let a user expand or shrink the virtual world around the user's hand with uniform scaling. Laviola et al [34] studied a foot-gesture-based scaling technique that controls the scale around the point where the user's head is projected onto a floor, while they are standing on a floor-map-like WIM in a CAVE environment. Wingrave et al. [65] utilized the wheel of a wireless mouse to control the scale of the WIM. Zhang [68] presented two scaling techniques for 3D environments, which scale the environment around two different scaling centers. Kiss et al. [30] studied four input methods, including one- and two-handed pinch, one-handed up-down movement with a physical button, and voice commands for controlling the zoom level for microscopic data in AR. Mezner et al. [44] presented a scaling technique that uses the height of the user's finger above the screen to control the zoom level of a map on a mobile device. Englmeier et al. [15] designed a position-based mapping for controlling the scale of spherical and planar WIMs.

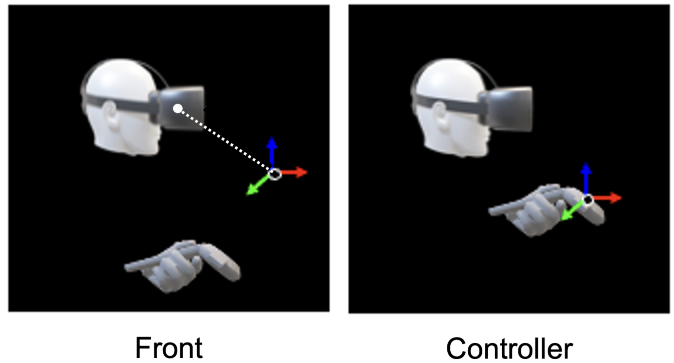

Fig. 3: The FRONT scaling center is located at a fixed position relative to the user's head, set at a fixed offset at initialization. The CONTROLLER scaling center coincides with the dominant hand controller or the midpoint between controllers for bi-manual interaction.

The work mentioned above explored different approaches for controlling scale, such as automatic scaling based on the selected LoS or the distance from the user's position to the surrounding objects and providing scaling techniques to manually control the scale of a given environment. Yet, the scaling technique was not the main focus of these works, and they typically also did not explore different variations within the design space of scaling techniques. One exception is Satriadi et al. [55], who explored the design space of pan-and-zoom techniques to navigate multiscale AR maps. Particularly for zooming (i.e., scaling) techniques, they studied not just different gestures for freehand scaling techniques but also explored different input mappings. However, their work was only limited to the use case of 2D navigation constrained to planar multiscale data and the scaling center always followed the input device at the user's hand. In our work, we investigate scaling techniques for 3D unconstrained navigation in nested MVEs with a LoS hierarchy and explore different design variations not just for input mechanisms but also for scaling centers within such environments. Other relevant studies include investigations into 7-DOF object manipulation techniques aimed at controlling the position, orientation, and scale of objects within immersive environments [11, 16]. Although these studies examined various methods for adjusting the scale of objects, they did not explore the application of these techniques for navigation within MVEs that feature hierarchical structures.

3 Bi-manual Scaling Techniques for Multiscale Navigation

Here, we investigate variations for two core components of bi-manual scaling techniques for multiscale navigation: 1) the input method for scale control and 2) the scaling center. The following subsections describe each component and the overall navigation interface used in our studies.

3.1 Input Methods for Scale Control

For scale control, we designed a novel SCROLL input method (See the right column in Figure 4) that inherits the benefits of both bi-manual interaction and position-based mapping. Building on the left-right scale differentiation theory of Guiard's kinematic chain model for bi-manual action [23], the SCROLL method separates a scale control operation into actions on two separate hands: the control to assign the scale level on the non-dominant hand, which requires relatively granular movement, and the control for placing the scaling center, which requires relatively finer movement, on the dominant hand (See below in Figure 6). Moreover, SCROLL employs a scroll gesture on

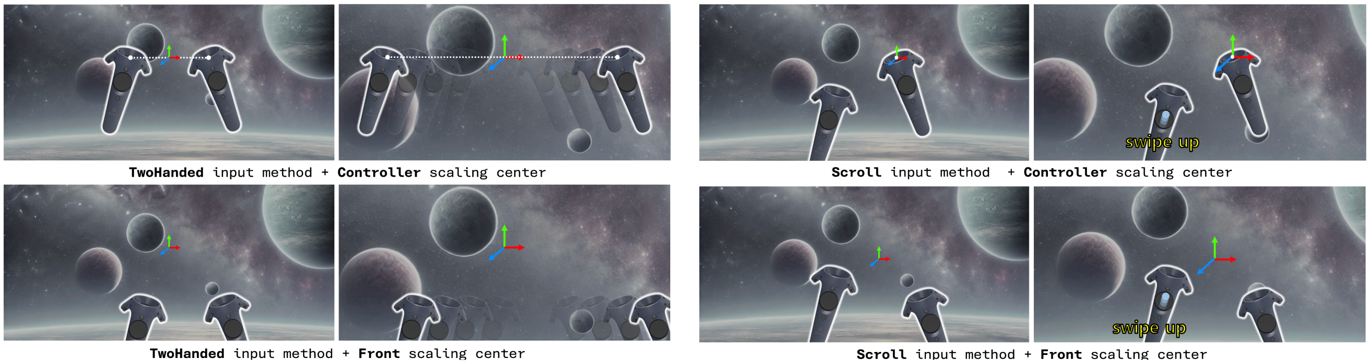

Fig. 4: The four variations of scaling techniques for exocentric navigation in multiscale virtual environments, resulting from the combination of two input methods for scale control (TWO-HANDED, SCROLL) and two methods to specify the scaling center (CONTROLLER, FRONT).

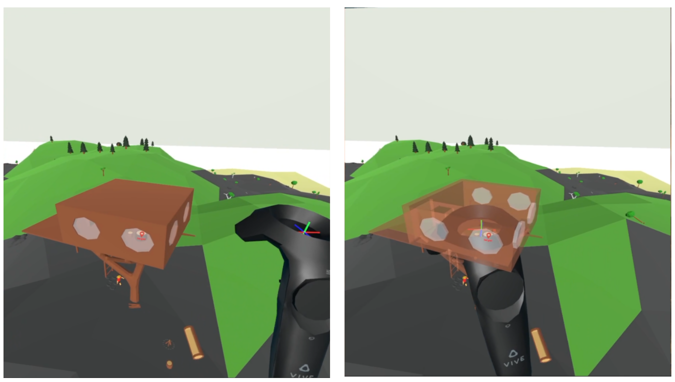

Fig. 5: The occlusion management technique shows a see-through version of the object pointed at and highlights the scaling center with a cross of axes. This technique was applied to all conditions in our user study.

the trackpad to control the scale (see Figure 6), a position-based input method known for its higher precision compared to rate-based input [49, 55]. Although Feng et al.'s findings demonstrate the advantages of rate-based input for scale control [16], we opted for position-based input for our trackpad implementation. Their study focused on gesture-based interactions that may not directly translate to trackpad input. The trackpad's key advantage lies in enabling precise value control through minimal finger movements [64]. This allows users to make both large-scale adjustments and fine-tuning with equal precision through simple swipe gestures, potentially diminishing the speed advantages typically associated with rate-based input. Our pilot study, which replicated the main experiment with fewer trials, supported this design decision. We initially implemented both input methods, with rate-based input controlled by the distance from the trackpad's center. Users consistently preferred position-based scrolling over rate-based input and did not use it, so we only provided position-based scrolling input. While rate-based input facilitated quick scale adjustments, position-based scrolling achieved comparable speeds for large-scale changes while maintaining superior precision.

The One-Handed with Two-Handed Scaling (OTS) technique from previous work [11] might appear similar to SCROLL method especially when it is combined with the CONTROLLER scaling center (see subsection 3.2 for more details about the scaling center design). Both techniques use one controller/hand to locate the scaling center. However, there are important differences between these techniques. First, OTS assigns scaling center control to the left controller, whereas SCROLL utilizes the one in the user's dominant hand. This distinction is significant, as precise placement of the scaling center is crucial for effective multiscale navigation [68]. By leveraging the dominant hand's superior motor control for finer movement, SCROLL enables more precise scaling center placement compared to OTS's fixed left-hand approach. Moreover, the techniques differ substantially in their physical demands. OTS requires extensive mid-air arm movements for scale control, which can lead to increased fatigue. In contrast, SCROLL employs small finger movements on a trackpad. This reduced physical demand is a key design consideration for SCROLL, directly addressing the limitations of mid-air gesture-based techniques of two-handed input exemplified by OTS.

As the baseline, we chose the TWO-HANDED pinch/grab method (See the left column of Figure 4), a natural, intuitive input method [3] that increases pointing accuracy when used with both hands [26] and which is the most prevalent one in many VR applications and research [10, 45, 46, 50, 67]. To mitigate the physical demand of large, repetitive arm movements, we designed a hybrid method inspired by Satriadi et al.' work [55], dividing each hand's input region into position-based and rate-based scaling zones (Figure 7). When users perform the pinch/grab action and spread their arms within the inner region, scaling the environment up (and themselves down), the environment enlarges proportionally to the distance the arms are spread (See above in Figure 6). When spreading their arms beyond the inner region, the system enlarges the environment with rate-based control, where scaling speed is proportional to distance, until the trigger button is released. The same principle applies to scale-down operations, with the outer region using the position-based scaling mode and vice versa.

$$S_{n+1}^{Two} = S_n^{Two} \cdot a^{l_n - l_{n-1}} \tag{1}$$

$$S_{n+1}^{Scroll} = S_n^{Scroll} \cdot a^{p_n - p_{n-1}} \tag{2}$$

The position-based scaling approach adjusts the scale factor (S_{n+1}) by multiplying the current scale (S_n) with an exponential term. This mechanism is implemented in two variants: For the TWO-HANDED method 1, the exponent ($l_n - l_{n-1}$) represents the change in distance between the controllers (in meters) from the previous to the current frame. For the SCROLL method 2, the exponent ($p_n - p_{n-1}$) represents the change in the thumb's vertical position on the trackpad between consecutive frames. For both methods, the base coefficient a was set to 5, heuristically determined through pilot studies. This creates an exponential relationship between the input movement (either controller separation or thumb position) and the resulting scale adjustment.

$$S_{n+1}^{Two} = S_n^{Two} \cdot (1 + b(l_n - l_0)) \tag{3}$$

The rate-based scaling updates the scale factor (S_{n+1}) using a linear relationship with input displacement. The new scale is computed by multiplying the current scale (S_n) with a term that varies between the two implementations: For the TWO-HANDED method (Equation 3), the term is $(1+b(l_n - l_0))$, where $l_n - l_0$ represents the displacement from the initial controller distance l_0 measured when scaling begins. The coefficient b was set to 0.1, as determined through pilot studies, which determines the sensitivity of the scaling response to the respective input displacement.

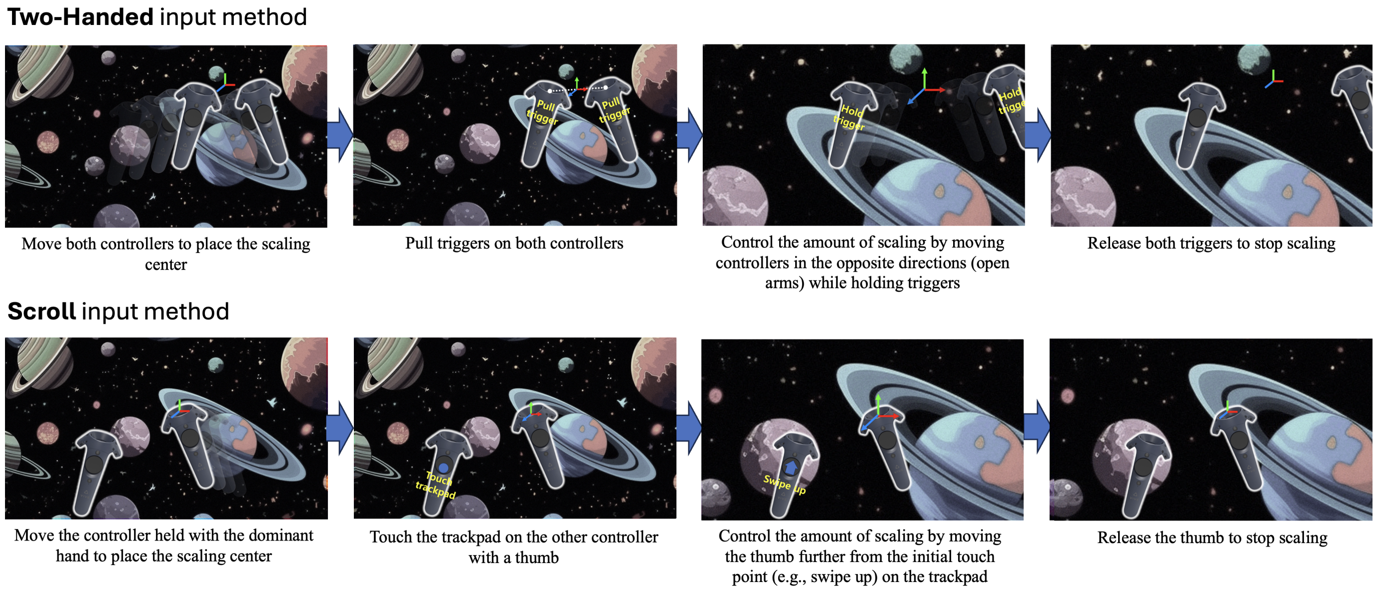

Fig. 6: Each step of the scale control actions for the TWO-HANDED and SCROLL input methods when using them with the CONTROLLER scaling center. When using them with the FRONT scaling center, users move the environment, using the click-and-drag locomotion method to place the scaling center at the desired position instead of moving the controller(s) in the first step. An indicator of the scaling center's position through a 3D crosshair with red, green, and blue axes remains constantly visible (for both input methods).

Through pilot studies, we determined coefficients a and b for our system to strike a balance: the scaling speed needs to be slow enough to allow users to reach their intended LoS without overshooting, yet fast enough so they do not perceive it as being overly slow.

3.2 Scaling Center

The second design component we investigate is the scaling center, the location around which the environment shrinks or expands when a user modifies the scale of the environment using a given input method. This important design component determines the efficiency and the difficulty of the navigation interface in MVEs [68]. We investigated two variations for this study, the CONTROLLER and FRONT scaling centers. Figure 3. The CONTROLLER scaling center is the most common design chosen in many VR systems, allowing users to freely choose where the environment shrinks or expands [10, 45, 46, 50, 67]. The CONTROLLER scaling center is either the position of the dominant hand controller (with SCROLL) or the middle position between two controllers (with TWO-HANDED), depending on the users' input method. The FRONT scaling center is anchored in front of the user's head, below eye level, which is set in an initialization stage, where the system uses the height of the user's head to record the offset. Specifically, the scaling center was placed at a point on a line starting midway between the eyes and rotated 20° below horizontal Figure 3(a)). We made this choice because we observed that participants naturally looked down when the virtual environments expanded or shrank relative to a co-located reference on the plane (i.e., a ground-plane visualization). Similar to previous work [37], we found 20° to be optimal for interacting with information displayed below eye-level [56]. Additionally, we adjusted the distance of the FRONT offset along the 20° line so that each participant could comfortably reach the scaling center with their controller, without fully extending their arms.

In a pilot study, we evaluated a scaling center that maintained a fixed offset relative to the user's current head position and orientation, even during navigation. We found this approach challenging as even minor movements of the user's head caused shifts in the scaling center's position, necessitating accurate head movements simultaneous with hand/finger movement for scale control, significantly complicating the accurate specification of a desired scaling center position. We also explored body-anchored scaling centers, such as at the user's head, feet, or torso. The scaling center at the user's head was discarded because we found that it causes severe motion sickness and diplopia when the objects in the environment converge towards the center between the two eyes while scaling down. The scaling centers at the user's feet and torso were also not chosen because looking at the environment converging or expanding around a position more or less directly below their head led to significant neck strain, even when only used for a brief duration. Another potential issue that contributed is also the weight of the VR headset, but we acknowledge that the newest headsets are becoming more lightweight.

The advantage of the FRONT scaling center lies in its independence from the user's gaze, requiring no visual tracking of the scaling center by the user. In our implementation, the scaling center is defined as a static offset relative to the current orientation of the head (and thus body) of the user during the initial loading phase of the application, and retains its position relative to the user throughout the interaction process, without altering its location in response to changes in the user's view direction. We made this design choice for our experiment to avoid potential issues with disorientation and/or cybersickness. Within the context of our experiment, which did not require the participant to substantially alter their orientation, this design for the interaction technique was sufficient and suitable to achieve all experimental tasks. Note that our user study primarily investigates locomotion—specifically the motor aspects of navigation—without delving into the cognitive elements involved in wayfinding, particularly spatial orientation, which caused substantial issues in a pilot study, where participants got (too) easily "lost in space" as they had to monitor not only spatial position and scale but also orientation. We thus decided to omit virtual rotation from our study to reduce the participant's cognitive load. Even without having to think about spatial orientation, participants nevertheless found the tasks in our study to be challenging.

3.3 Locomotion and Visualization Techniques for Multi-scale Virtual Environments

For moving in 3D, we implemented a well-known locomotion technique [42, 54], which allows a user to click/pick any 3D position with the controller on either hand, press the trigger button, and then pull the controller in the direction opposite to where they intend to move.

To mitigate physical exertion through repetitive clutching with click-and-drag movements, we apply rate-based locomotion control beyond a certain distance threshold, making this a hybrid locomotion technique,

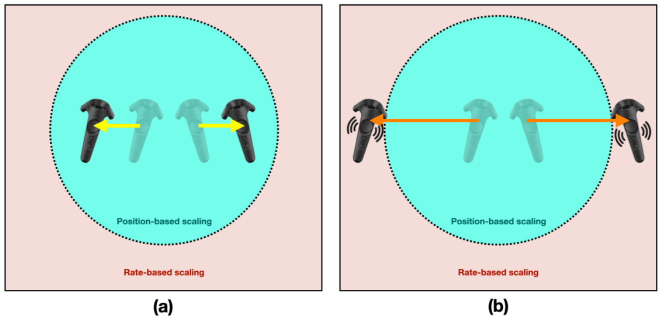

Fig. 7: Position-based scaling was used when the arms were within the inner region (blue) and rate-based scaling when outside (reddish). In the experiment, the inner and outer regions were not visualized to avoid a visual distraction. Instead, the two controllers vibrated (only) when they were in the outside region.

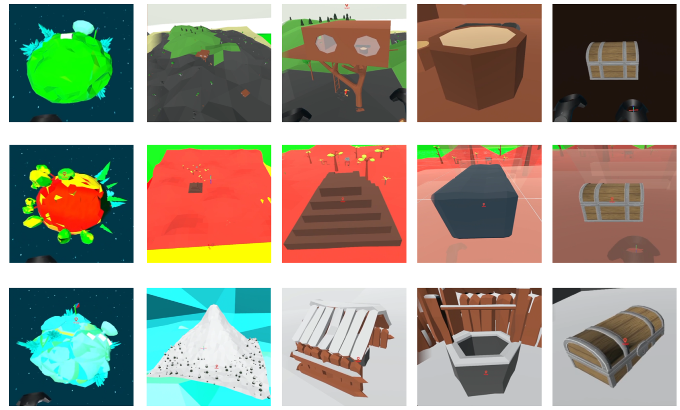

Fig. 8: The three variations of planets (Planet A, B, C shown in the 1st column) with the same configuration for their LoS hierarchy. Note that the lower LoS nested inside other objects are not always visible from the outside.

similar to Satriadi et al.' method [55]. Our hybrid locomotion technique thus enables users to switch seamlessly from position-based click-and-drag locomotion to rate-based flying locomotion by moving the controller more than a threshold distance.

Also, we used two visualization techniques to facilitate the user's spatial understanding of the nested MVE. To support users when they need to look inside an object, the first is an occlusion management technique, which then allows them to accurately put the scaling center at an appropriate position inside the object, which is necessary for entering an object (Figure 5). For this purpose, we applied an X-ray, i.e., translucent, visualization effect, which is known to be effective for understanding the spatial context *within an area occluded by a wall* [39]. We applied the X-ray effect in all conditions but limited it to a single object, namely the one inside (or behind) of which the user positioned the scale center. This ensured that the scaling center remained visible from the user's viewpoint, unobstructed by intervening objects. The scaling center was consistently visualized as the red-green-blue 3D crosshair, positioned at the midpoint between the two controllers, similar to the Spindle technique [11], or on the right controller, regardless if a button was pressed, as illustrated in Figure 6. We implemented this effect by changing the opacity of the texture of each object in Unity using object colliders. When the scaling center intersects with an object's collider, that object's texture opacity reduces to 30%; otherwise, it remains fully opaque. The second visualization technique showed the current ground plane as a grid (using the height of the user as a reference, modulated by the current scale) that illustrated how the environment expanded or shrank around the scaling center and helped users understand how objects came closer or moved further away. This feature enabled users to recognize when they committed navigation errors, such as enlarge-and-push-away errors [10] or when they estimated the visual depth incorrectly, which then enabled them to recover quickly from such errors through corrective actions to achieve the intended view.

The discussed locomotion and visualization techniques may influence travel performance when combined with different scaling techniques. While our study did not reveal clear behavioral differences for these combinations, further research is warranted to investigate the interaction between these techniques and the scaling techniques evaluated in this study.

3.4 Navigation Interface Design Decisions

Our user study intentionally separated scaling from locomotion operations based on pilot study findings. When testing the same tasks with fewer trials, we observed that simultaneous scaling and locomotion using the baseline TWO-HANDED technique frequently caused user confusion. Unlike 3D object manipulation tasks [11] where simultaneous scaling, translation, and rotation are beneficial, combined operations in our context where we evaluate technique for scene navigation led to unintended viewpoint changes. Users attempting to scale the space through mid-air arm movements often triggered unwanted translations, causing accidental entries or exits from objects at the current LoS. These errors significantly disrupted user orientation and required considerable time to correct. To ensure a fair comparison, we applied this separated operation design to both TWO-HANDED and FRONT techniques.

Our user study was designed to evaluate the travel performance with these scaling technique variations. As mentioned above, we did not incorporate rotation functionality as in Pai et al.'s work [51], since the ability to maintain a sense of direction while constantly changing the orientation of the environment varies largely between individuals, which is made worse by the constant scale changes. Note that users can still freely look around in any direction by naturally turning their head or body, we just did not implement a "virtual" rotation technique for locomotion that rotates the entire space through a single input, such as snap turns in VR games [58, 63].

Also (worse yet) changing orientation virtually without vestibular cues can cause disorientation [36, 37], which might confound results regarding travel performance. This is different from the user looking around and enabling them to turn their body. We just prevented the "virtual" rotation where the user changes the orientation of the space using buttons or any type of input on the controller while standing still and looking forward.

4 USER EVALUATION OF SCALING TECHNIQUES

We conducted a user study to test the following hypotheses:

H1: The SCROLL input method enables more accurate navigation of a nested MVE than TWO-HANDED.

H2: The SCROLL input method is more efficient than the TWO-HANDED input method for navigating a nested MVE.

H3: The FRONT scaling center enables more accurate navigation of a nested MVE than the CONTROLLER scaling center.

H4: The FRONT scaling center is easier to use than the CONTROLLER scaling center for navigating a nested MVE.

We based **H1** on previous work [10] that found that a two-handed scaling technique introduces erroneous behavior in multiscale navigation and our own pilot study for this work where we observed specifically that users were experiencing difficulties with placing the scaling center correctly with a two-handed technique. **H2** is based on Guiard's theories of two-handed interfaces [23], which outlined the advantages of assigning different roles to the two hands. More specifically, they suggest using the non-dominant hand for coarse and less frequent movement and the dominant one for fine and frequent ones. **H3** and **H4** are based on Zhang's findings [68] regarding the increased accuracy and lower difficulty of multiscale navigation when the scaling center is fixed instead of giving the user control of it. This study was reviewed and approved by the Simon Fraser University Research Ethics Board.

4.1 Experimental Design

We used a two-factor within-subject design. The first variable was the *Input* method for scale control: TWO-HANDED and SCROLL. The second was the *ScalingCenter* method: FRONT and CONTROLLER. Each participant experienced all four conditions, i.e., all combinations of two *Input* methods and two *ScalingCenters*.

4.2 Participants

We recruited 20 participants from the local university, 11 male and nine female, with an average age of 23.0 years(SD=4.65). Two were left-handed and eighteen right-handed. Sixteen participants had prior experience with VR technologies such as HTC Vive or Meta Quest, with an average of 26.6 hours (SD=29.5) of total VR usage. All participants had experience with first-person 3D games, averaging 80.25 hours (SD=30.6) of total gameplay. They were compensated with the equivalent of US$ 10.

4.3 Apparatus and Environment

We conducted our experiment with an HTC Vive Pro HMD, with 1080×1200 pixels for each eye, and Vive controllers for input. The experiment was performed in a space of 3 m $\times$ 3 m, spacious enough to let participants turn freely and perform controller-based navigation in any direction. To evaluate the scaling techniques, we designed an MVE with two planets, each of which contains multiple smaller nested LoS. One of the planets is the one that every task begins on; we call this the initial planet in the experiment. The initial planet at the 2st LoS contains four other LoS: the 3st LoS of a village, the 4nd level of a building in that village, the 5rd level of a piece of furniture in that building, and the 6th level of a chest (Figure 8) within the furniture. The target planet had a similar LoS hierarchy but featured different objects at each level, with the target located at the 6th LoS. We created three planet variations (Figure 8, each row showing different planets and their internal objects), with planet order randomized for each task. The frequency with which each version appeared as the initial and the target planet was the same, i.e., planets A, B, and C appeared twice each as the initial and the target planet.

4.4 Task

The main experimental task required participants to travel from the initial to the target's positions as quickly as possible. Starting from the 6th LoS inside a chest (Figure 8, last column), participants needed to reach the target's location in another chest on a distant target planet. This journey required first scaling the environment down to reach the 1st LoS, and then obtaining a global view where both initial and target planets appeared as miniatures within reach. Participants then positioned their scaling center near the target planet and scaled up, traversing from the 2nd LoS (planet surface, Figure 8, first column) to the 6th LoS where the target is. To ensure that participants could easily figure out where to travel, the direction to the target was always shown through a direction indicator and a widget (see Figure 5). We detected that participants had reached the target when their head entered the chest, while the chest was scaled up to be bigger than $1m^3$.

4.5 Procedure

At the start, participants filled out a consent form and a pre-questionnaire detailing their background. Following this, they were instructed to put on the HMD and grasp both controllers while positioning themselves at the center of the designated experimental area. While participants had the liberty to turn their heads, the experimenter recommended that they keep their position predominantly centered within the area and also their body orientation aligned with the starting direction. We did this to ensure that they navigated the environment in an exocentric fashion through manipulation, rather than adopting an egocentric approach by moving themselves.

The experiment entailed four distinct conditions, presented in a counterbalanced sequence to negate potential learning effects. The purpose of our experiment was to test how different *Input* methods and *ScalingCenter* result in different travel performance in unconstrained navigation tasks described in subsection 4.4. With each new condition, participants first received guidance on utilizing the respective *Input* method in conjunction with the specific *ScalingCenter*, followed by three tasks designated as practice. Upon completion of the practice tasks, participants proceeded to the three primary experimental tasks. For all tasks (practice and main), the planet that contains the target appeared in one of the six canonical directions. The order of target directions was randomly shuffled for each condition. After the completion of all tasks for a condition, participants were asked to fill out a NASA Task Load Index (NASA-TLX) questionnaire [24]. Upon completion of all conditions, participants were prompted to fill out a post-task questionnaire, using a 7-point Likert scale to assess each condition in terms of five categories with the following questions:

1) *Preference*: You liked the technique.

2) *Intuitiveness*: Do you agree that this technique is intuitive, i.e., easy to learn and understand how it works?

3) *Accuracy*: Do you agree that this technique helps you accurately perform navigation, i.e., helps you accurately get to the target without making erroneous movements?

4) *Efficiency*: Do you agree that this technique reduced the time to reach the target?

5) *Ease of Use*: Do you agree that this technique is easy to use?

On average, the study took about 80 minutes per participant.

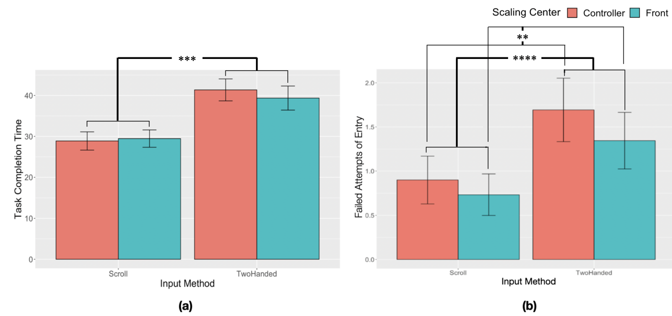

Fig. 9: (a) The SCROLL input method significantly reduced task completion time compared with the TWO-HANDED input method. Error bars show 95% confidence intervals. (b) Compared to TWO-HANDED, the SCROLL input method significantly reduced the number of failed attempts for entering a smaller LoS. Also, the FRONT input method significantly reduced the number of failed attempts for entering a smaller LoS compared to CONTROLLER. Error bars show 95% confidence intervals.

4.6 Results

We analyzed task completion time, error rate, task load, and subjective rating data with inferential analysis using repeated-measures ANOVA with $\alpha = 0.05$ in R. For task load and subjective ratings, we used ART [66] before ANOVA.

Task completion time: We analyzed completion times for each task type individually. The data was normally distributed. The two-way repeated measures ANOVA identified a significant effect of *Input* ($F(1,19) = 40.20, p < 0.001, \eta^2 = 0.14$) but no effect for the *ScalingCenter* $F(1,19) = 1.21, p = 0.29, \eta^2 = 0.01$) on the completion time. There was no interaction between *Input* and *ScalingCenter* $F(1,19) = 0.32, p = 0.58, \eta^2 = 0.002$). See Figure 9 (a).

Travel Distance: To analyze travel distance, we first computed the distance between the relative position of participants' heads from the origin of the environment model and the target position at every moment. Then, we summed up the travel distance for each condition and compared them between conditions. Two-way ANOVA did not identify a significant effect of *Input* nor *ScalingCenter* on travel distance.

Number of Failed Attempts of Entry: We also analyze the number of failed attempts for entering objects to evaluate the navigation accuracy of each condition for achieving the intended view at a target scale. This identified a significant effect of *Input* ($F(1,19) = 33.96, p < 0.0001, \eta^2 = 0.64$) and *ScalingCenter* ($F(1,19) = 5.93, p < 0.01, \eta^2 = 0.24$) but there was no interaction between

Input and *ScalingCenter* ($F(1, 19) = 1.75, p = 0.19, \eta^2 = 0.08$). See Figure 9 (b).

Task Load Index: A two-way ANOVA identified a significant effect of *Input* on the overall score, physical or mental demand, and effort. The TWO-HANDED method received higher scores in all the above categories than the SCROLL method. However, there was no significant effect of *Input* on temporal demand. The effects of *Input* on frustration and performance scores were marginal, but not significant. There was no effect of *ScalingCenter*: overall, physical or mental demand, temporal demand, effort, frustration, performance. There was no interaction between *Input* × *ScalingCenter* on the overall score or any of the TLX subcategories. See Table 1.

Subjective Ratings: In terms of user preference, we identified a significant effect of *Input*, with SCROLL being preferred, but not of *ScalingCenter*. There was no interaction between *Input* and *ScalingCenter*. For intuitiveness, we identified a significant effect of *ScalingCenter* with FRONT being perceived as more intuitive than CONTROLLER, but not for *Input*. There was no interaction *Input* × *ScalingCenter*. For ease of use, the analysis identified a significant effect of *Input* with SCROLL being scored higher but no effect of *ScalingCenter*. There was an interaction between *Input* and *ScalingCenter*. A post-hoc test with Tukey-HSD showed a significant difference between the SCROLL input with the CONTROLLER scaling center vs. the TWO-HANDED input with CONTROLLER ($p < 0.0001$), the SCROLL input with the CONTROLLER scaling center vs. the SCROLL input with the FRONT scaling center ($p < 0.01$), and the TWO-HANDED input with the CONTROLLER scaling center vs. the TWO-HANDED input with the FRONT scaling center ($p < 0.001$). For efficiency, the results showed a significant effect of *Input*, where the SCROLL method got a higher score than TWO-HANDED, but no effect of *ScalingCenter*. There was no interaction *Input* × *ScalingCenter*. For accuracy, there was no significant effect of *Input* nor *ScalingCenter*. See Table 1.

Post-task Interview Responses: Participants predominantly favored the SCROLL method over TWO-HANDED for its efficiency and ease of use. P2 and P19 found it quicker for task completion, while P10 highlighted the minimal effort required. P17 described it simply as "the most efficient method." Physical comfort was also a significant factor, with P9 appreciating the "freedom of movement" and P18 noting the "least physical load" due to minimal arm movement. Flexibility was another advantage, with P8 mentioning the ability to easily move to specific locations. Many participants found SCROLL intuitive, e.g., P12, while P22 and P21 related it to familiar daily scrolling activities on phones and computers. P20 added that it mimicked real-life actions. In contrast, participants had reservations about the TWO-HANDED method, alluding to physical demand and discomfort. P21 mentioned the awkwardness of "controlling with two hands" and the potential risk of hitting others. P2 and P9 found it the "most difficult to use" due to the increased physical effort. P12 noted the tiring nature of the scaling motion, and P18 mentioned the need to "hold my arms out for an elongated period." The lack of intuitiveness was also a concern, with P18 and P20 finding the arm movements and switching between hands challenging. P19 observed that TWO-HANDED required more effort and time. Despite these concerns, some participants appreciated aspects of the TWO-HANDED method. P16 noted the finer control it provided through body motions. P12 found the two-handed technique intuitive, resembling "real-life pulling," and P11 felt it was easier to calculate and control.

Participants' evaluations of the CONTROLLER and FRONT scaling centers were influenced by the *Input* method. The CONTROLLER scaling center was generally preferred with the SCROLL input method. P22 and P21 found it intuitive and natural, with P20, P2, and P12 noting its reduced effort and ease of use with one hand. P8, P9, and P10 acknowledged an initial learning curve but recognized its efficiency, with P9 describing it as a "fun mental challenge." P21 found CONTROLLER more accurate with SCROLL than with TWO-HANDED, while P19 and P22 highlighted its natural feel and simplicity. Some participants preferred the FRONT scaling center with SCROLL. P4 appreciated the division of interaction between hands, making it easy to learn. P1 found the fixed position more manageable, leading to fewer error corrections. P13 noted the ease of using SCROLL with minimal finger movement and appreciated the fixed position for target alignment. Two participants favored TWO-HANDED with FRONT and CONTROLLER scaling centers. P14 appreciated the alignment with their focal point, and P16 highlighted the flexibility and finer control. However, most participants found TWO-HANDED with CONTROLLER frustrating due to high fatigue, and uncomfortable hand and arm placement.

5 DISCUSSION

Here, we first discuss how our results relate to the hypotheses we set out above. Regarding navigation accuracy, the user study showed that the TWO-HANDED method exhibited a higher navigation error rate compared to SCROLL, as indicated by the higher number of failed entry attempts. The NASA-TLX performance score was also significantly lower (better) for SCROLL. Participants felt more confident and performed tasks more accurately with SCROLL, supporting **H1**. The SCROLL method significantly reduced task completion time, supporting **H2**. Overall, SCROLL received higher ratings and more positive feedback in terms of preference, ease of use, and efficiency. The advantages of SCROLL included quick and precise scale control with minimal effort and the ability to simultaneously control scale and movement, significantly reducing task completion time. This parallelism is not possible with the TWO-HANDED method, which requires both hands for scale control.

Although travel distance did not differ significantly between the FRONT and CONTROLLER scaling centers, FRONT significantly reduced failed entry attempts compared to CONTROLLER, partially supporting **H3**. This aligns with Zhang et al. [68], who found that specifying the scaling center increased navigation errors. Users showed the most accurate navigation performance when FRONT was paired with SCROLL, suggesting a synergistic effect. Further investigation is needed to fully understand the benefits of combining FRONT and SCROLL in multiscale navigation.

We found no effect of *ScalingCenter* on ease of use, rejecting **H4**. However, the *Input* method affected ease of use, with an interaction observed between *ScalingCenter* and *Input*. With SCROLL, ease of use was consistent between CONTROLLER and FRONT, as users could swiftly adjust the scale and the scaling center's position. Conversely, with TWO-HANDED, navigation was more challenging than with CONTROLLER due to difficulties in positioning both hands and the lack of a mechanism to refine the scaling center. The FRONT method, needing only one hand to manipulate the environment, offered greater precision and more natural hand placements, avoiding awkward positions.

In terms of task load, the TWO-HANDED input method scored higher in overall, mental, physical, and effort. By minimizing physical movement and level of effort by letting users easily place the scaling center at an intended position, the SCROLL method reduced the overall task load by reducing mental and physical challenges that the TWO-HANDED method poses.

In terms of subjective ratings, the SCROLL input method scored higher ratings in preference, ease of use, and efficiency. Overall, it was evaluated more highly for its efficacy and better UI design for navigating MVE with nested structures, over the TWO-HANDED method. Surprisingly, the TWO-HANDED *Input* method did not score higher in intuitiveness, despite its widespread use in immersive applications with exocentric navigation interfaces. This finding contrasts with previous research [3], where two-handed grabbing-based methods received the highest rating for intuitiveness. A potential explanation for this discrepancy may lie in users' familiarity with scroll/swipe-like gestures from their daily interactions with touchscreen devices such as smartphones and smartwatches. The prevalence of these interactions in daily life might have influenced users' perceptions of intuitiveness. This unexpected result warrants further investigation to better understand the factors influencing perceived intuitiveness in immersive interfaces.

The user's VR experience did not significantly affect performance, task load, or subjective ratings across input methods. However, three participants (20-21-year-old male gamers with more than 100 hours in first-person 3D games and more than 15 hours in VR) preferred the TWO-HANDED technique. We did not find any significant differences

Measurements		Descriptive Analysis				Inferential Analysis (ANOVA)		
		Input						
		TwoHanded		Scroll		*Input*	*ScalingCenter*	*Input X ScalingCenter*
		Controller	Front	Controller	Front			
NASA-TLX	Overall	*M* =35.90 , *SD* = 19.35	*M* = 33.84, *SD* = 17.45	*M* = 21.12, *SD* = 15.94	*M* = 25.00, *SD* = 15.17	**$F(1,19) = 8.28, p < 0.05, \eta^2 = 030$**	$F(1,19) = 0.13, p = 0.72, \eta^2 = 0.01$	$F(1,19) = 1.61, p = 0.22\ \eta^2 = 0.08$
	Mental	*M* = 32.80, *SD* = 22.02	*M* = 32.40, *SD* = 17.90	*M* = 21.40, *SD* = 14.80	*M* = 27.80, *SD* = 18.20	**$F(1,19) = 4.70, p < 0.05, \eta^2 = 0.20$**	$F(1,19) = 0.80, p = 0.38, \eta^2 = 0.04$	$F(1,19) = 1.13, p = 0.30, \eta^2 = 0.06$
	Physical	*M* = 42.55, *SD* = 23.17	*M* = 39.85, *SD* = 19.53	*M* = 19.85, *SD* = 16.93	*M* = 20.10, *SD* = 16.46	**$F(1,19) = 46.43, p < 0.001, \eta^2 = 0.71$**	$F(1,19) = 0.16, p = 0.70, \eta^2 = 0.01$	$F(1,19) = 0.41, p = 0.53\ \eta^2 = 0.02$
	Temporal	*M* =30.10 , *SD* =19.37	*M* =29.20 , *SD* = 22.65	*M* = 24.00, *SD* = 25.13	*M* = 25.15, *SD* = 23.14	$F(1,19) = 0.73, p = 040, \eta^2 = 0.04$	$F(1,19) = 0.00, p = 0.97, \eta^2 < 0.001$	$F(1,19) = 0.15, p = 0.70, \eta^2 = 0.01$
	Frustration	*M* = 23.50, *SD* =20.19	*M* = 22.90, *SD* = 19.64	*M* = 13.65, *SD* = 15.05	*M* = 17.25, *SD* = 15.37	$F(1,19) = 3.93, p = 0.06, \eta^2 = 0.17$	$F(1,19) = 0.31, p = 0.58, \eta^2 = 0.02$	$F(1,19) = 0.37, p = 0.55, \eta^2 = 0.02$
	Effort	*M* = 39.45, *SD* =25.50	*M* = 36.20, *SD* = 21.66	*M* = 23.95, *SD* = 18.08	*M* = 26.65, *SD* =18.61	**$F(1,19) = 5.04, p < 0.05, \eta^2 = 021$**	$F(1,19) = 0.01, p = 0.93, \eta^2 = 0.001$	$F(1,19) = 1.26, p = 0.28, \eta^2 = 0.06$
	Performance	*M* = 28.10, *SD* = 23.07	*M* = 25.20, *SD* = 19.45	*M* = 15.85, *SD* = 20.85	*M* = 18.50 , *SD* = 14.99	$F(1,19) = 3.59\ p = 0.07, \eta^2 = 0.16$	$F(1,19) = 0.00, p = 0.97, \eta^2 = 0.001$	$F(1,19) = 1.01, p = 0.33, \eta^2 = 0.05$
Subjective ratings	Preference	*M* = 3.05, *SD* = 1.93	*M* = 3.35, *SD* = 1.42	*M* = 5.10, *SD* = 1.21	*M* = 4.55, *SD* = 0.94	**$F(1,19) = 20.93, p < 0.001, \eta^2 = 0.52$**	$F(1,19) = 0.15, p = 0.70, \eta^2 = 0.01$	$F(1,19) = 1.75, p = 0.20, \eta^2 = 0.09$
	Ease of use	*M* = 3.25 , *SD* = 1.62	*M* = 4.20, *SD* = 1.20	*M* = 5.55, *SD* = 0.68	*M* = 5.20, *SD* = 0.89	**$F(1,19) = 25.61, p < 0.001, \eta^2 = 0.65$**	$F(1,19) = 1.93, p = 0.18, \eta^2 = 0.09$	**$F(1,19) = 6.97, p < 0.05\ \eta^2 = 0.27$**
	Accuracy	*M* = 3.70, *SD* = 1.56	*M* = 4.50, *SD* = 1.36	*M* = 4.60, *SD* = 1.54	*M* = 4.35, *SD* = 1.57	$F(1,19) = 0.67, p = 0.42, \eta^2 = 0.03$	$F(1,19) = 0.61, p = 0.44, \eta^2 = 0.03$	**$F(1,19) = 5.44, p < 0.05, \eta^2 = 0.21$**
	Efficiency	*M* = 3.20, *SD* = 1.94	*M* = 3.25, *SD* = 1.41	*M* = 5.35, *SD* = 0.99	*M* = 4.45, *SD* = 1.57	**$F(1,19) = 18.18, p < 0.001, \eta^2 = 0.49$**	$F(1,19) = 1.67, p = 0.21, \eta^2 = 0.08$	$F(1,19) = 3.27, p = 0.09, \eta^2 = 0.15$
	Intuitiveness	*M* = 3.75, *SD* = 1.65	*M* = 4.70, *SD* = 1.03	*M* = 4.80, *SD* = 1.36	*M* = 4.75, *SD* = 1.01	$F(1,19) = 2.28, p = 0.15, \eta^2 = 0.10$	**$F(1,19) = 4.68, p < 0.05, \eta^2 = 0.20$**	$F(1,19) = 3.33, p = 0.08, \eta^2 = 0.15$

Table 1: Descriptive and inferential analysis results of NASA-TLX scores and subjective ratings of the four conditions (TWO-HANDED with CONTROLLER, TWO-HANDED with FRONT, SCROLL with CONTROLLER, and SCROLL with FRONT), with significant results highlighted.

in our results between left-handed and right-handed participants. While this sample is too small to generalize preferences and user tendencies beyond these demographics, further research could investigate how specific user profiles may exhibit distinct behaviors with certain interaction techniques.

6 LIMITATIONS

We acknowledge our participant pool was limited. Our study involved university students, who may have more experience with the latest technologies than the general public. It would be worthwhile to study the investigated techniques with a broader and more diverse range of users to ensure the generalizability of our findings. Moreover, we did not incorporate common virtual rotation techniques, such as orbiting or snap turn, in our navigation interface since orientation control via scene rotation was beyond the main scope of this study. It would be beneficial to study the effect of such commonly implemented functionality in VR on multiscale navigation with the proposed scaling techniques in the future.

7 CONCLUSION AND FUTURE WORK

We studied scaling techniques for navigation interfaces in multiscale virtual environments with nested structures, where smaller levels of scale are nested within multiple larger levels. In our user study, we evaluated two components of scaling techniques, the input method and the scaling center. For the input method, we found that our SCROLL method significantly reduces the task completion time, error rate, and overall task load and improves usability compared with the state-of-the-art baseline method, the TWO-HANDED pinch/grab method. We also found that the *ScalingCenter* affects the usability of the scaling technique differently according to which *Input* it was matched with, with the CONTROLLER scaling center paired with SCROLL providing the best user experience for navigating MVEs with nested structures.

Some open questions remain. First, investigating the effect of the input device on the navigation performance of scaling techniques is warranted. Using other types of 3D controllers or different input modalities, such as mid-air hand gestures and/or eye gaze, could offer different options for 3D locomotion in MVEs with nested structures. Specifically, some mixed and virtual reality devices, such as the Meta Quest and Valve Index, lack trackpad input. There, potential solutions for scale control with position-based input could include a mid-air swiping gesture with the non-dominant hand, a click-and-drag gesture using the trigger button on a controller held with the non-dominant hand, or a single-hand microgesture [9] to scroll between levels. It would be interesting to study these adaptations on such platforms. Second, evaluating scaling techniques in collaborative scenarios would also be worthwhile. Several design components of the scaling technique, such as the scaling center and visualization techniques, could potentially have positive effects on the spatial understanding between collaborators, who may have different roles.

REFERENCES

[1] P.-A. Albinsson and S. Zhai. High precision touch screen interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 105–112, 2003. 2

[2] F. Argelaguet and M. Maignant. Giant: stereoscopic-compliant multi-scale navigation in ves. In *Proceedings of the 22nd acm conference on virtual reality software and technology*, pp. 269–277, 2016. 1, 3

[3] C. R. Austin, B. Ens, K. A. Satriadi, and B. Jenny. Elicitation study investigating hand and foot gesture interaction for immersive maps in augmented reality. *Cartography and Geographic Information Science*, 47(3):214–228, 2020. 3, 4, 8

[4] J. Avery, M. Choi, D. Vogel, and E. Lank. Pinch-to-zoom-plus: an enhanced pinch-to-zoom that reduces clutching and panning. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pp. 595–604, 2014. 2

[5] J. Avery, S. Malacria, M. Nancel, G. Casiez, and E. Lank. Introducing transient gestures to improve pan and zoom on touch surfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–8, 2018. 2

[6] F. Bacim, D. Bowman, and M. Pinho. Wayfinding techniques for multiscale virtual environments. In *2009 IEEE Symposium on 3D User Interfaces*, pp. 67–74. IEEE, 2009. 3

[7] B. B. Bederson and J. D. Hollan. Pad++ a zooming graphical interface for exploring alternate interface physics. In *Proceedings of the 7th annual ACM symposium on User interface software and technology*, pp. 17–26, 1994. 2

[8] B. Buxton et al. Multi-touch systems that i have known and loved. *Microsoft Research*, 56:1–11, 2007. 2

[9] E. Chan, T. Seyed, W. Stuerzlinger, X.-D. Yang, and F. Maurer. User elicitation on single-hand microgestures. In *Conference on Human Factors in Computing Systems*, CHI '16, p. 3403–3414, Apr 2016. doi: 2858036.2858589 9

[10] I. Cho, J. Li, and Z. Wartell. Multi-scale 7dof view adjustment. *IEEE transactions on visualization and computer graphics*, 24(3):1331–1344, 2017. 1, 2, 3, 4, 5, 6

[11] I. Cho and Z. Wartell. Evaluation of a bimanual simultaneous 7dof interaction technique in virtual environments. In *2015 IEEE symposium on 3D User Interfaces (3DUI)*, pp. 133–136. IEEE, 2015. 3, 4, 6

[12] S. Cmentowski, A. Krekhov, and J. Krüger. Outstanding: A multi-perspective travel approach for virtual reality games. In *Proceedings of the annual symposium on computer-human interaction in play*, pp. 287–299, 2019. 3

[13] G. Decay. Maquette. https://store.steampowered.com/app/762840/Maquette/, 2021. 2

[14] Dynamoid. Cosm worlds. https://store.steampowered.com/app/467480/C_O_S_M/, 2016. 1, 2

[15] D. Englmeier, W. Sajko, and A. Butz. Spherical world in miniature: Exploring the tiny planets metaphor for discrete locomotion in virtual reality. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 345–352. IEEE, 2021. 3

[16] J. Feng, I. Cho, and Z. Wartell. Comparison of device-based, one and two-handed 7dof manipulation techniques. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction*, pp. 2–9, 2015. 3, 4

[17] C. Forlines and C. Shen. Dtlens: multi-user tabletop spatial data exploration. In *Proceedings of the 18th annual ACM symposium on User interface software and technology*, pp. 119–122, 2005. 2

[18] J. Gackenbach, M. Rosie, J. Bown, and T. Sample. Dream incorporation of video-game play as a function of interactivity and fidelity. *Dreaming*, 21:32–50, 2011. doi: 10.1037/a0022868 1

[19] H. Games. No man's sky. https://store.steampowered.com/app/275850/No_Mans_Sky, 2016. 1, 2

[20] Google. Google earth vr. https://store.steampowered.com/app/348250/Google_Earth_VR/, 2016. 1

[21] Google. Google tilt brush. https://store.steampowered.com/app/327140/Tilt_Brush/, 2016. 1, 2

[22] N. N. Griffin and E. Folmer. Out-of-body locomotion: Vectionless navigation with a continuous avatar representation. In *25th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–8. ACM, 2019. doi: 10.1145/3359996.3364243 3

[23] Y. Guiard. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *Journal of motor behavior*, 19(4):486–517, 1987. 3, 6

[24] S. G. Hart. Nasa task load index (tlx). volume 1.0; paper and pencil package, 1986. 7

[25] F. Hartung, P. Hagoort, and R. M. Willems. Readers select a comprehension mode independent of pronoun: evidence from fmri during narrative comprehension. *Brain and Language*, 170:29–38, 2017. doi: 10.1016/j.bandl.2017.03.007 1

[26] A. G. Hauptmann. Speech and gestures for graphic image manipulation. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 241–245, 1989. 2, 4

[27] J. A. Henry and N. F. Polys. The effects of immersion and navigation on the acquisition of spatial knowledge of abstract data networks. *Procedia Computer Science*, 1(1):1737–1746, 2010. 2

[28] K. Hinckley, M. Czerwinski, and M. Sinclair. Interaction and modeling techniques for desktop two-handed input. In *Proceedings of the 11th annual ACM symposium on User interface software and technology*, pp. 49–58, 1998. 2

[29] D. P. Käser, M. Agrawala, and M. Pauly. Fingerglass: efficient multiscale interaction on multitouch screens. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1601–1610, 2011. 2

[30] F. Kiss, P. W. Woźniak, V. Biener, P. Knierim, and A. Schmidt. Vum: understanding requirements for a virtual ubiquitous microscope. In *Proceedings of the 19th International Conference on Mobile and Ubiquitous Multimedia*, pp. 259–266, 2020. 3

[31] R. Kopper, T. Ni, D. A. Bowman, and M. Pinho. Design and evaluation of navigation techniques for multiscale virtual environments. In *Ieee virtual reality conference (vr 2006)*, pp. 175–182. Ieee, 2006. 3

[32] D. Kouřil, T. Isenberg, B. Kozlíková, M. Meyer, M. E. Gröller, and I. Viola. Hyperlabels: Browsing of dense and hierarchical molecular 3d models. *IEEE transactions on visualization and computer graphics*, 27(8):3493–3504, 2020. 3

[33] A. Krekhov, S. Cmentowski, K. Emmerich, M. Masuch, and J. Krüger. Gullivr: A walking-oriented technique for navigation in virtual reality games based on virtual body resizing. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, pp. 243–256, 2018. 3

[34] J. J. LaViola Jr, D. A. Feliz, D. F. Keefe, and R. C. Zeleznik. Hands-free multi-scale navigation in virtual environments. In *Proceedings of the 2001 symposium on Interactive 3D graphics*, pp. 9–15, 2001. 3

[35] J.-I. Lee, P. Asente, B. Kim, Y. Kim, and W. Stuerzlinger. Evaluating automatic parameter control methods for locomotion in multiscale virtual environments. In *26th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–10, 2020. 1

[36] J.-I. Lee, P. Asente, and W. Stuerzlinger. A comparison of zoom-in transition methods for multiscale vr. In *ACM SIGGRAPH 2022 Posters*, pp. 1–2. ACM, 2022. 1, 6

[37] J.-I. Lee, P. Asente, and W. Stuerzlinger. Designing viewpoint transition techniques in multiscale virtual environments. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 680–690. IEEE, 2023. 1, 5, 6

[38] Y. Li, C.-W. Fu, and A. Hanson. Scalable wim: Effective exploration in large-scale astrophysical environments. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):1005–1012, 2006. 3

[39] S. Liao, Y. Zhou, and V. Popescu. Ar interfaces for disocclusion—a comparative study. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 530–540. IEEE, 2023. 6

[40] T. B. V. LLC. The body vr. https://thebodyvr.com, 2016. 1

[41] S. Malacria, E. Lecolinet, and Y. Guiard. Clutch-free panning and integrated pan-zoom control on touch-sensitive surfaces: the cyclostar approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2615–2624, 2010. 2

[42] D. P. Mapes and J. Moshell. A two-handed interface for object manipulation in virtual environments. *Presence: Teleoperators & Virtual Environments*, 4:403–416, 1995. 2, 5

[43] J. McCrae, I. Mordatch, M. Glueck, and A. Khan. Multiscale 3d navigation. *Proceedings of the 2009 symposium on Interactive 3D graphics and games - I3D '09*, pp. 7–14, 2009. doi: 10.1145/1507149.1507151 3

[44] T. Menzner, T. Gesslein, A. Otte, and J. Grubert. Above surface interaction for multiscale navigation in mobile virtual reality. In *2020 IEEE conference on virtual reality and 3D user interfaces (VR)*, pp. 372–381. IEEE, 2020. 3

[45] Meta. Gravity sketch. https://www.meta.com/experiences/1587090851394426, 2019. 2, 4, 5

[46] Meta. Meta horizon world. https://www.meta.com/experiences/2532035600194083/, 2020. 2, 4, 5

[47] M. Nancel, J. Wagner, E. Pietriga, O. Chapuis, and W. Mackay. Mid-air pan-and-zoom on wall-sized displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 177–186, 2011. 2

[48] M. Negulescu, J. Ruiz, and E. Lank. Zoompointing revisited: supporting mixed-resolution gesturing on interactive surfaces. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, pp. 150–153, 2011. 2

[49] I. Oakley and S. O'Modhrain. Tilt to scroll: Evaluating a motion based vibrotactile mobile interface. In *First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics Conference*, pp. 40–49. IEEE, 2005. 4

[50] Oculus. Oculus quill. https://www.oculus.com/experiences/rift/1118609381580656, 2016. 1, 2, 4, 5

[51] Y. S. Pai, Z. Chen, L. Chan, M. Isogai, H. Kimata, and K. Kunze. Pinchmove: Improved accuracy of user mobility for near-field navigation in virtual environments. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '18, article no. 7, 11 pages. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3229434.3229470 2, 6

[52] K. Perlin and D. Fox. Pad: an alternative approach to the computer interface. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, pp. 57–64, 1993. 2

[53] J. S. Pierce and R. Pausch. Navigation with place representations and visible landmarks. In *IEEE Virtual Reality 2004*, pp. 173–288. IEEE, 2004. 3

[54] W. Robinett and R. Holloway. Implementation of flying, scaling and grabbing in virtual worlds. In *Proceedings of the 1992 symposium on Interactive 3D graphics*, pp. 189–192, 1992. 3, 5

[55] K. A. Satriadi, B. Ens, M. Cordeil, B. Jenny, T. Czauderna, and W. Willett. Augmented reality map navigation with freehand gestures. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 593–603. IEEE, 2019. 2, 3, 4, 6

[56] K. L. Schultz, D. M. Batten, and T. J. Sluchak. Optimal viewing angle for touch-screen displays: Is there such a thing? *International journal of industrial ergonomics*, 22(4-5):343–350, 1998. 5

[57] S. Sekiguchi and T. Komuro. Volgrab: Realizing 3d view navigation by aerial hand gestures. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 2064–2069, 2017. 2

[58] B. Softworks. The elder scrolls v: Skyrim vr. https://store.steampowered.com/app/611670/The_Elder_Scrolls_V_Skyrim_VR/, 2018. 6

[59] S. Stellmach, M. Jüttner, C. Nywelt, J. Schneider, and R. Dachselt. Investigating freehand pan and zoom. *Mensch & Computer 2012: interaktiv informiert–allgegenwärtig und allumfassend!?*, 2012. 2

[60] I. E. Sutherland. Sketch pad a man-machine graphical communication system. In *Proceedings of the SHARE design automation workshop*, pp. 6–329, 1964. 2

[61] TechViz. Techviz xl. https://www.techviz.net/en/techviz-xl, 2021. 2

[62] D. R. Trindade and A. B. Raposo. Improving 3d navigation techniques in multiscale environments: a cubemap-based approach. *Multimedia Tools and Applications*, 73(2):939–959, Nov. 2014. doi: 10.1007/s11042-012-1127-8 1

[63] Valve. Half-life: Alyx. https://store.steampowered.com/app/546560/HalfLife_Alyx/, 2020. 6

[64] M. Warburton, C. Campagnoli, M. Mon-Williams, F. Mushtaq, and J. R. Morehead. Kinematic markers of skill in first-person shooter video games. *PNAS Nexus*, 2, 2023. doi: 10.1093/pnasnexus/pgad249 4

[65] C. A. Wingrave, Y. Haciahmetoglu, and D. A. Bowman. Overcoming world in miniature limitations by a scaled and scrolling wim. In *3D User Interfaces (3DUI'06)*, pp. 11–16. IEEE, 2006. 3

[66] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only ANOVA procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, 4 pages, pp. 143–146. ACM, New York, NY, USA, 2011. doi: 10.1145/1978942.1978963 7

[67] Y. Yang, M. Cordeil, J. Beyer, T. Dwyer, K. Marriott, and H. Pfister. Embodied navigation in immersive abstract data visualization: Is overview+ detail or zooming better for 3d scatterplots? *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1214–1224, 2020. 2, 4, 5

[68] X. L. Zhang. Multiscale traveling: crossing the boundary between space and scale. *Virtual reality*, 13(2):101, 2009. 1, 3, 4, 5, 6, 8

[69] J. Zhao and A. Klippel. Scale-unexplored opportunities for immersive technologies in place-based learning. In *2019 IEEE Conference on virtual reality and 3d user interfaces (VR)*, pp. 155–162. IEEE, 2019. 3