To Slide or Not to Slide: Exploring Techniques for Comparing Immersive Videos

Xizi Wang
David R. Cheriton School of Computer Science
University of Waterloo
Waterloo, Ontario, Canada
l84wang@uwaterloo.ca

Yue Lyu
David R. Cheriton School of Computer Science
University of Waterloo
Waterloo, Ontario, Canada
y32lyu@uwaterloo.ca

Yalong Yang
School of Interactive Computing
Georgia Institute of Technology
Atlanta, Georgia, USA
yalong.yang@gatech.edu

Jian Zhao
David R. Cheriton School of Computer Science
University of Waterloo
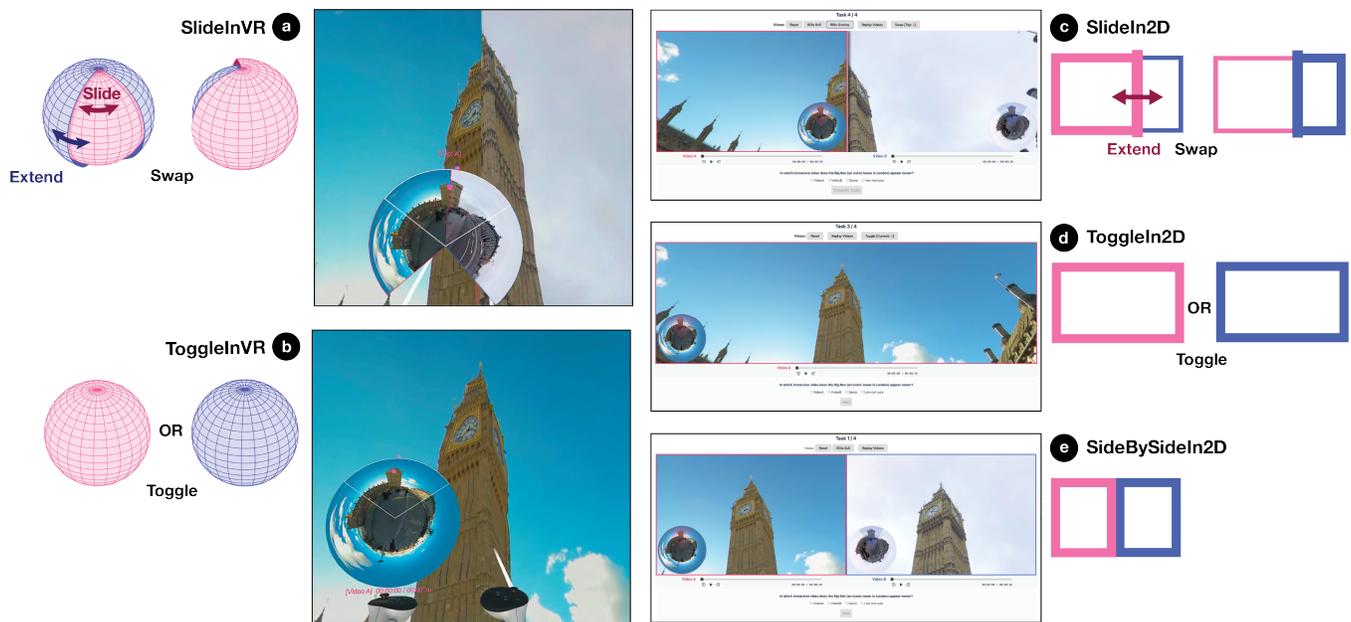Waterloo, Ontario, Canada
jianzhao@uwaterloo.ca

Figure 1: Overview of the five techniques for comparing two immersive 360° videos, each illustrated with a schematic of key interactions and a task example: (a) SlideInVR, (b) ToggleInVR, (c) SlideIn2D, (d) ToggleIn2D, and (e) SideBySideIn2D. Techniques (a) and (b) support in-headset VR comparison, and techniques (c)–(e) are presented in 2D desktop settings. All examples depict the same Task#14, where two immersive videos from © VR Gorilla (used with permission) captured the same Big Ben from two nearby camera viewpoints in 2017 [12] and 2025 [13].

Abstract

Immersive videos (IVs) provide 360° environments that create a strong sense of presence and spatial exploration. Unlike traditional videos, IVs distribute information across multiple directions, making comparison cognitively demanding and highly dependent on interaction techniques. With the growing adoption of IVs, effective comparison techniques have become an essential yet underexplored area of research. Inspired by the "sliding" concept in 2D media comparison, we integrate two established comparison strategies from the literature—toggle and side-by-side—to support IV comparison with greater flexibility. For an in-depth understanding of different strategies, we adapt and implement five IV comparison techniques across VR and 2D environments: SlideInVR, ToggleInVR, SlideIn2D, ToggleIn2D, and SideBySideIn2D. We then conduct a user study ($N = 20$) to examine how these techniques shape users' perceptions, strategies, and workflows. Our findings provide empirical insights into the strengths and limitations of each technique, underscoring

the need to switch between comparison approaches across scenarios. Notably, participants consistently rate SlideInVR and SlideIn2D as the most flexible and favorite methods for IV comparison.

CCS Concepts

• **Human-centered computing** → **Virtual reality**; **Web-based interaction**; **Empirical studies in HCI**; *Visual analytics.*

Keywords

immersive video, 360-degree video, video comparison, virtual reality, 2D, comparison approach, visual comparison

1 Introduction

Immersive video (IV), e.g., 360° video, has gained significant attention due to its ability to provide a more engaging, interactive, and realistic experience compared to traditional 2D videos, fostering greater focus, interest and active participation as well as promoting effective and low-workload learning [7, 22, 26]. It is widely used across various domains, including general education (e.g., used as lecture materials or lecture recording [6, 24, 26, 33, 38]), professional training (e.g., nursing [7] and surgeon training [5]), and entertainment (e.g., sports or concert recordings and narrative video content [2, 13]). As IV adoption grows, practitioners (and sometimes viewers) may frequently need to compare two IV clips during production (e.g., reviewing and selecting IV clips [18] across takes or viewpoints [34, 49]), editing (e.g., comparing stitched IV outputs produced by different stitching tools [3] and aligning the points of interest between two clips [35]), and evaluation (e.g., making pairwise quality selections within a larger IV collection [23], evaluating the quality and consistency of AI-generated IVs across AI models [44], and comparing the visual or motion differences of two IV recordings during learning activities [6]). However, 2D video comparison workflows do not directly translate to IV comparison, because viewers need to navigate a omnidirectional scene where maintaining ego-centric spatial context is paramount; without specialized techniques, this process often leads to significant memory strain and disorientation.

Virtual Reality (VR) is the most common modality for viewing and comparing 360° IVs, offering a direct and intuitive visual representation of the IV content. This approach allows users to fully experience the video and make effective comparisons and decisions. However, VR typically supports only one IV at a time, forcing users to toggle between videos during comparison. This disrupts workflows, increases reliance on short-term memory to retain visual details between toggles, and ultimately reduces both the accuracy and efficiency of IV comparisons [19]. While prior research has explored the navigation and editing of IVs in VR [20, 35], these efforts largely focus on single-video interaction rather than supporting direct comparison of different IVs.

In contrast, projecting IVs onto a 2D screen treats them as traditional flat videos but sacrifices essential spatial information. Flattening spherical content introduces distortions, which can obscure important details and hinder accurate comparisons [35]. Previous work on 2D video comparison has explored approaches such as side-by-side, toggle/overlay, and explicit encoding [9, 14], demonstrating their efficiency for traditional videos. However, their applicability to immersive 360° content remains underexplored, leaving open questions about how to effectively support IV comparisons.

These limitations in both VR and 2D highlight the lack of a general approach that supports direct, flexible, and efficient comparison of 360° IVs. To fill in the gap, we investigate how different comparative approaches (e.g., toggle, side-by-side) and modalities (e.g., VR, 2D) affect users when comparing two IVs. Inspired by the techniques in 2D video comparison [9, 14], we adapt the general concept of "sliding" [9, 36, 40], which is commonly used for comparing 2D images and videos, to support pairwise comparison of IVs in both 2D and VR. This adapted sliding concept lets users dynamically transition between side-by-side and toggle views, effectively combining them into a hybrid approach. Concretely, it overlays two IVs as front and back layers and comprises three core features (*slide*, *extend*, and *swap*) that support flexible adjustments of positions and visible regions of IVs. *Slide* allows users to reposition the two IVs to place the regions of interest (ROIs) side by side or in an overlapping (toggle-like) configuration; *extend* enables users to adjust the spatial distribution by extending or shrinking the visible area of each IV within a limited view; and *swap* switches the order of the two layers, changing which IV is prioritized and visible.

We apply this concept to VR and 2D modalities, resulting in two techniques: SlideInVR (Figure 1-a) and SlideIn2D (Figure 1-c). Employing just the toggle and side-by-side approaches, we further implement three basic IV comparison techniques, including ToggleInVR (Figure 1-b), ToggleIn2D (Figure 1-d) and SideBySideIn2D (Figure 1-e). There is no SideBySideInVR because placing two omnidirectional IVs directly side by side in VR is not feasible. To better support IV comparisons, all techniques provide zoomed-in and global views as well as temporal-spatial visualizations of ROIs.

To address our overarching goal of exploring how comparative approach and modality can affect user' perceptions, strategies, and workflows, we conducted a within-subject study with 20 VR users to assess the pros and cons of these five techniques. We derive four types of IV comparison tasks, including comparisons of temporal occurrence, spatial-temporal motions, local visual differences, and global visual differences, from previous work [4, 9]. Our results highlighted the benefits and trade-offs of different comparison approaches and modalities. VR was perceived to offer more natural navigation, spatial awareness, and effectiveness in spotting visual quality and distortions, but also introduced a higher workload and a steeper learning curve compared to 2D. Both task type and technique shaped participants' workflows and strategies: side-by-side comparison was generally preferred for temporal occurrence and dynamic ROIs, while toggle was more effective for fine-grained visual differences when ROIs appeared in the same location. Notably, SlideInVR and SlideIn2D consistently emerged as the two most favored techniques, valued for their flexibility and ability to hybrid multiple comparison methods while supporting dynamic allocation of space.

In summary, our work makes the following contributions: (1) we extend the concept of sliding in 2D image/video comparison to immersive video comparison, resulting in two new techniques, SlideInVR and SlideIn2D, which enable seamless transitions between toggle and side-by-side approaches; (2) we conduct a within-subjects study (N=20) evaluating five representative immersive video comparison techniques across 2D and VR as well as comparison approaches, which provide empirical findings and design guidance on how modality and task characteristics shape users' workflows and strategies.

2 Background

2.1 Immersive Videos and Applications

Immersive videos (IVs) refer to video formats that capture a scene in a wide field of view (FoV), most commonly including 180° and 360° videos (also referred to as omnidirectional, panoramic, or spherical videos [16]). Several common 2D projection formats are used to represent IVs. The equirectangular projection, which provides a panoramic overview, is the most common projection used by major platforms such as YouTube [15]. The azimuthal equidistant projection, rendered as a "planet view", has been applied for visualization and minimap applications, but is less common as a primary distribution format [15, 17, 35]. While these projections make spherical video compatible with conventional 2D players, they also introduce geometric distortion, which can affect perceived visual quality and usability [15, 35]. IVs can be captured and displayed in either monoscopic or stereoscopic modes [10]. The monoscopic is one of the most common formats and supported by most consumer devices; while stereoscopic provides a stronger sense of depth perception, it requires more complex capture. Therefore, in this paper, our focus is comparing monoscopic 360° IVs.

IVs have been widely applied in various domains, including higher education [6, 24, 26, 38], virtual field trips [1, 33], professional training [7], and entrainment [2]. When viewed through VR devices, IVs provide immersive environments that foster a strong sense of presence and spatial exploration [26]. Compared to traditional 2D videos, they also enable viewers, especially learners, to feel more focused, motivated, and joyful, encouraging active engaging and interactions [5, 7, 26]. These benefits and applications highlight the need for better support in comparing different types of IVs and understanding their differences. For example, Jin et al. found that instructors expressed their concerns about the quantity and quality of IV when selecting and producing educational materials [24]. Motivated by these opportunities and challenges, our work explores how different modalities can support experts and general users in comparing different types of IVs.

2.2 Immersive Videos Editing

There exist several commercial applications, along with powerful plugins, for IV editing, such as Insta360 Studio[1] and Adobe Premiere Pro[2]. For example, Mettle's Skybox VR Player allowed users to preview spherical video within Premiere, while Adobe later showcased CloverVR, an in-VR editing extension for Premiere, at Adobe MAX 2016 [39]. The research community has also explored various tools

and techniques to support IV editing. For instance, Nguyen et al. proposed Vremiere that allows users to browse and edit monoscopic IVs directly in a VR headset while still relying on mouse and keyboard as input devices [35]. This in-headset interface introduced a new VR editing workflow and demonstrated several helpful features, including the overview of IVs in planet view that we later adopted in our design. They also highlighted the importance of WYSIWYG ("What You See Is What You Get") during editing for reducing the gulf between traditional 2D desktop editing and immersive VR viewing. Building on this work, 6DIVE is a system supporting fully in-VR stereoscopic IV editing and multi-track spatial editing that enables users to blend one video clip on top of another similar to picture-in-picture [20]. A comparison of 6DIVE with a desktop editor showed that VR allowed more intuitive spatial alignment and received higher attractiveness and stimulation. While these systems focus on IV editing and preview, they highlight recurring moments where editors compare alternative clips or edit variants (e.g., clip selection, alignment, and placement) [18, 20, 35] and imply a need for comparison techniques that work directly in VR, motivating a more systematic study of IV comparison techniques.

2.3 Navigation in Immersive Videos

Researchers have also explored how to improve the navigation in immersive videos, which is a fundamental task in IV editing, browsing, and other applications. The first body of research has focused on physical navigation in IVs, which is how viewers use natural body movements, such as head rotations, leaning, or walking, to explore IV content. For example, shot orientation techniques were proposed to automatically reorient the viewpoint during IV viewing to keep important content in view [37]. Similarly, Lin et al. investigated automatic rotation and visual guidance techniques, showing that they improved users' ability to maintain focus across different types of IVs [29]. Kang and Cho introduced an interactive navigation system for IV playback that automatically generates a viewing path based on saliency and optical flow, allowing users to adjust the view interactively to improve comfort and reduce missed content [25].

Some works have also explored alternative projections and additional views of IVs to support navigation and reduce the fear of missing out. For example, Li et al. presented an algorithm for generating route tapestries (a new projection showing left and right side views) to support navigating tour-specific IVs and comparing vacation destinations in a 2D interface [28]. Vermast and Hürst proposed two 3D thumbnail designs for providing overviews of IVs, and found that spherical 3D thumbnails improved user experience and supported information searching in VR [42]. Lin et al. presented Outside-In, a technique that visualizes out-of-sight ROIs in IV by reintroducing them as spatial picture-in-picture previews [30]. Prior studies also explored the use of panoramic thumbnails and highlighted the importance of providing an overview during IV viewing [46].

Besides physical navigation, other work has focused on temporal navigation, which is how viewers control their position in time within an IV, such as rewinding, fast-forwarding, or jumping to specific moments. Yu and Bi studied different types of IV timelines in VR and found that 3D wrist-based timeline offered better user

---

experiences [47]. Also, Liu et al. presented RadarVR that provides spatiotemporal visual guidance for ROIs to support IV viewing [31]. These findings inspired our design of temporal navigation around the controller and encouraged us to further explore more natural forms of temporal navigation in IVs.

Together, these systems demonstrate a variety of approaches for physical and temporal navigation. However, as another critical task, little work has explored IV comparison across 2D or VR modalities. To address this gap, we examine five IV comparison techniques across 2D and VR, revealing how these techniques, task type and modality influence users' navigation, strategies, and overall perceptions.

2.4 Visual Comparison

Prior work in visualization research has established taxonomies of comparative designs. Gleicher et al. summarized a taxonomy of comparative designs and identified three main comparison techniques: juxtaposition (side-by-side), superposition (overlay), and explicit encoding [19]. In juxtaposition, items are shown in separate views for direct comparison, which we applied directly in our design of SideBySideIn2D. Superposition overlays items in the same space and often relies on techniques such as semi-transparency or quick toggles to animate between views for easier comparison. We adopted this approach in our ToggleIn2D and ToggleInVR techniques. Explicit encoding visualizes the relationships or differences between these two items directly. These categories provide a foundation for situating subsequent work on visual comparison in both 2D and immersive environments, regardless of the data format.

2.4.1 Visual Comparison in 2D.
Visual comparison has been widely explored in 2D visualization. For example, Tominski, Forsell, and Johansson investigated a natural folding interaction alongside other comparative interactions (i.e., side-by-side and shine-through) for comparing printed complex data on a 2D interface [40]. Our "sliding" concept was initially inspired by the natural interaction with "paper" demonstrated in this work, which provided high flexibility for switching between different comparison approaches by moving papers around.

Beyond this work, other studies also explored different comparison approaches in 2D interfaces, especially traditional 2D video comparison. Cherukuru and Scheitlin demonstrated a visual comparator that supports juxtaposition, superposition, and explicit encoding for comparing dynamic spatiotemporal data. Their system allows users to superimpose up to three synchronized animations and compare them through a slider-based interface, helping reduce visual clutter while drawing attention to temporal changes. Furthermore, these "before and after" slider widgets have been widely adopted in commercial 2D image comparison tools and web platforms (e.g., Juxtapose [36]), where users drag a divider to reveal more or less of each image. Our IV comparison techniques aligned with this design rationale, particularly the SlideIn2D where juxtaposition, superposition, and sliding interactions are adapted to IVs to support both overview and subtle difference detection. Recently, Baker et al. introduced a suite of interaction techniques for 2D video comparison [9]. Their system supports equalizing the reference frames of videos, juxtaposition techniques such as

side-by-side and small multiples, superposition techniques including transparency and window-blind overlays, explicit encodings of differences, and timeline-based translations between temporal and linear representations. These techniques reduce the manual effort required for comparing 2D videos and improve usability across diverse datasets. Our IV comparison techniques are closely related in that we also adapt juxtaposition and superposition comparison. However, we extend this design space to immersive 360° video formats and emphasize flexible switching between comparison approaches. Unlike 2D videos in their study, which often share a fixed perspective, IV introduces additional challenges, such as the need to process information from all directions, viewing multiple non-distorted 360° videos simultaneously, tracking moving objects in space, balancing local and global comparisons, and comparing across different viewpoints, which our work specifically explored.

2.4.2 Visual Comparison in VR.
Visual comparison in VR has received growing attention, as immersive environments introduce unique challenges and opportunities beyond traditional 2D settings. Recently, Yuan et al. explored different VR techniques for comparing color differences in 360° images, including a side-by-side layout that placed 2D images next to each other in VR, a sequential approach that allowed users to switch between full images, and a partial approach that displayed portions of each 360° image [48]. While their design rationale aligns with ours, their techniques focus primarily on detecting subtle differences of static images in VR and offer less flexibility than our IV comparison approaches. Another recent work by Zhu et al. evaluated three 3D visual comparison techniques for change detection in VR [50]. These techniques, including Sliding Window, 3D Slider and Switch Back, were derived from the classic paradigms of juxtaposition and superposition comparison. Their design concept of toggle-style switching (Switch Back) and slider-based partitioning (Sliding Window and 3D Slider) also parallels our IV comparison techniques (toggle and sliding) and features (peeking). However, whereas Zhu et al. focused on 3D scene change detection, our work explores how different comparison approaches can be flexibly adapted for IV comparison tasks.

In motor-skill learning, Ma et al. introduced avaTTAR, an AR system that combines detached and on-body visualizations to support table tennis stroke learning [32]. Their system juxtaposes expert and user avatars for detached (side-by-side) comparison while also overlaying on-body cues to guide posture correction. This highlights suitable usage cases for different comparison approaches: side-by-side supports overall comparisons, whereas overlay cues help users perceive local subtle differences, consistent with findings from prior work [45]. While their focus is on motor-skill acquisition, our work investigates immersive video comparison techniques that similarly build on juxtaposition and superposition, but emphasize flexible modalities for analyzing visual differences across 360° video content.

3 The Sliding Concept

Prior work on 2D comparison systems has demonstrated a variety of approaches, including toggle, side-by-side, and explicit encoding [9, 19, 41]. When applied to 360° videos, each approach has its own limitations. Toggle enables users to overlay two IVs and switch
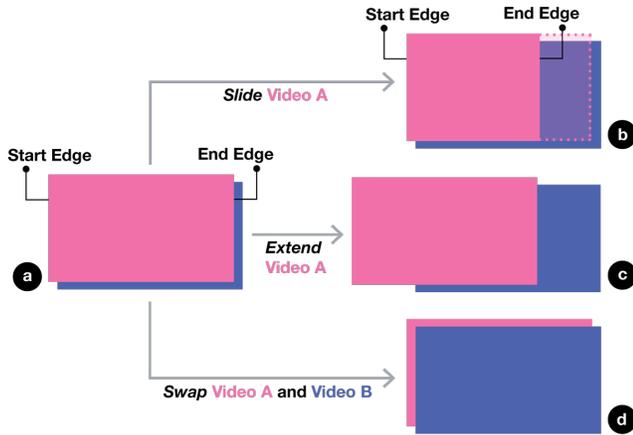
Figure 2: Abstract illustration of the three core features of our adapted sliding concept in panels (a)–(d). Two IVs, Video A (pink) and Video B (blue), are shown in overlapping display areas in (a). The concept supports three interactions: *extend* ◂|▸ , from (a) to (b); *slide* ↔ , from (a) to (c); and *swap* ⇄ , from (a) where the front layer is Video A to (d) where the front layer is Video B.

between them temporally. This preserves spatial alignment and supports fine-grained comparison when the ROIs align, but meanwhile places a heavy burden on short-term visual memory [19], especially for long IVs. This challenge is amplified in VR: if the ROIs are not in the same viewing direction, users must rotate their viewpoint between toggles, further increasing memory load and also physical demanding. Side-by-side, in contrast, allows holistic comparisons by showing both contexts simultaneously in 2D. However, splitting the FoV into separate views disrupts spatial continuity and forces users to mentally align disparate perspectives [19]. In VR, a true side-by-side presentation is not feasible for 360° IVs: a viewer can only be immersed inside one omnidirectional sphere at a time, so we cannot simply place two full 360° environments "next to each other." Instead, side-by-side designs for 360° IVs necessarily present each IV as a partial view (e.g., in windows or cropped regions), which shrinks each view and limits peripheral context and immersion. Additionally, explicit encoding is effective when relationships can be computed and visualized, especially in data visualizations [19]. But in texture-rich 360° video, such relationships are difficult to define or calculate. Thus, in this work, we are not considering explicit encoding for 360° videos comparison. These limitations motivate our "sliding" approach for IV comparison, which combines the strengths of toggle and side-by-side for IV comparison tasks.

While sliding is not a new interaction in 2D media comparison, we significantly extend and adapt this concept to IV comparison, particularly in VR. In sliding, each IV is represented as a display area defined by a start and an end edge (Figure 2-a), and these display areas are layered in terms of rendering. Users can configure the static layout of the display areas by adjusting their sizes and positions relative to others. Sliding supports three core interactions (Figure 2-a,b,c,d):

1. *extend* ◂|▸ : extending the edges of each IV display area to control how much of the 360° field is revealed;

2. *slide* ↔ : repositioning the display area around the polar axis of the spherical video to adjust the layout;

3. *swap* ⇄ : changing the layer order to foreground a different IV.

Together, these interactions enable users to flexibly transition between overlay and side-by-side views, supporting both fine-grained and holistic comparisons of media. While the sliding concept can in principle support comparison of multiple IVs, in this work, we focus on two-IV comparison and implement features specifically for this case.

3.1 Adopting Sliding in VR

Browsing one IV in VR is intuitive: the IV is projected onto the inside surface of an omnidirectional sphere, and the user is located in the center of the sphere to view the whole IV and navigate by rotating their VR headset. When adopting sliding in VR, the two IVs are layered, with the top-layer IV "nested inside" the back-layer IV (Figure 3-a). A start edge and an end edge of each IV—each resembling a great-circle arc on a sphere that passes through both poles—together define a spherical-lune display area, while the complementary region (from the end edge back to the start edge) is not rendered. These edges determine which portion of each IV is displayed, e.g., 75% (Figure 3-a) or 100% when the two edges of an IV overlap (Figure 3-e). The size and placement of each IV display area can be adjusted through *slide* ↔ (Figure 3-b) and *extend* ◂|▸ (Figure 3-c), and the layer order can be adjusted through *swap* ⇄ (Figure 3-d).

3.2 Adopting Sliding in 2D

When viewing IVs in 2D environments, the IV is projected to a physical monitor. As users cannot move the physical monitor, changing the FoV is commonly achieved by dragging the IV within a display area that shows only a partial view of the content. Therefore, in 2D, each IV display area corresponds to the current FoV of that IV. When the sliding concept is adopted in 2D, two IV display areas are overlaid, with the front-layer IV rendered on top of the back-layer IV (Figure 4-a). Each IV display area has a left edge (start edge) and a right edge (end edge). The size and placement of each IV display area can be adjusted through *extend* ◂|▸ , which moves its edges (Figure 4-b), and *slide* ↔ , which moves the entire display area (Figure 4-c). The layer order can be adjusted through *swap* ⇄ (Figure 4-d). In our final implementation, we simplified this concept in 2D by removing the *slide* ↔ interaction and modifying *extend* ◂|▸ and *swap* ⇄ to better use screen space and reduce interaction steps, which will be explained in Section 4.3.

4 Technique Design and Implementation

In this work, we aim to investigate how the two modalities—2D and VR—and the three comparison approaches—side-by-side, toggle, and sliding (i.e., flexibly supporting both side-by-side and toggle)—affect users' behaviors and experiences in comparing two IVs. We design and implement five techniques: two applied the sliding concept (i.e., SlideInVR and SlideIn2D) and three other techniques (i.e., ToggleInVR, ToggleIn2D, and SideBySideIn2D). Because full 360° IVs cannot be placed side by side in VR without distortion, we only implemented the side-by-side version in 2D (i.e., SideBySideIn2D). We focus only on IV comparison tasks in our study (Section 5.2),
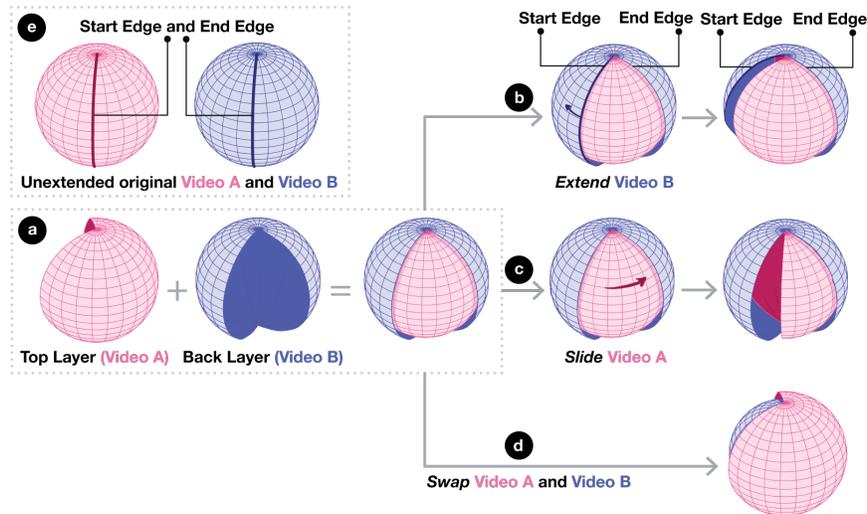
Figure 3: Semantic illustration of our sliding concept in VR. In (e), the two original 360° IVs, Video A (pink) and Video B (blue), are shown as full spheres, with their start and end edges aligned. In (a), two spherical-lune IV display areas are combined. Panels (b)–(d) demo the three interactions supported by sliding: *extend* ◄|► **, which shrinks Video B's display area by moving its start edge (b);** *slide* ⬌ **, which rotates Video A so that the overlapping area gradually increases (c); and** *swap* ⇆ **, which changes the front layer from Video A to Video B (d).**
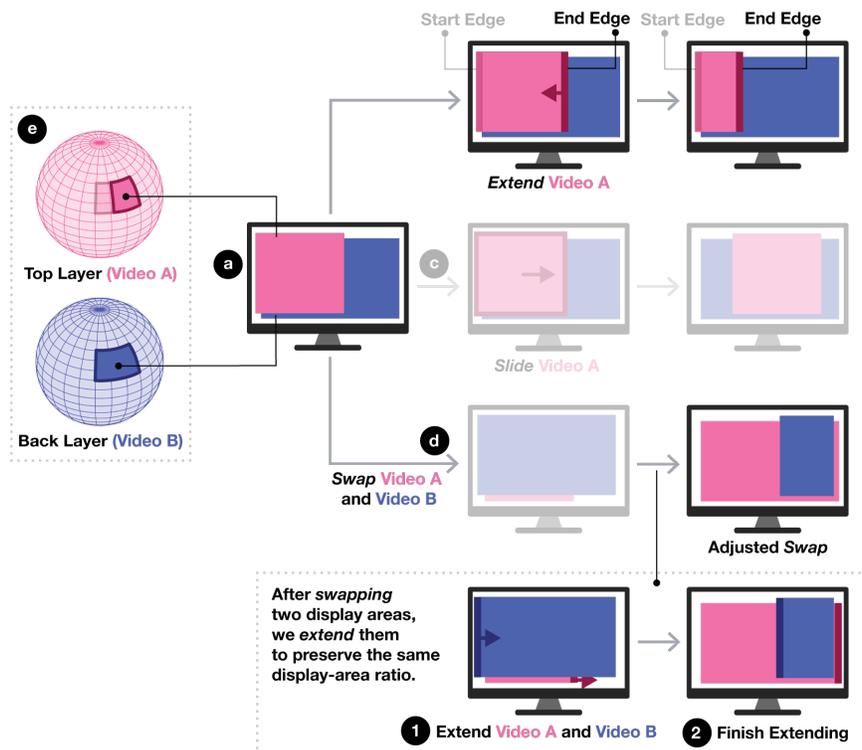

Figure 4: Semantic illustration of our sliding concept in 2D. Partial original 360° IV in (e), Video A (pink) and Video B (blue), are shown in two overlapping IV display areas on the monitor in (a). Panels (b)–(d) demo the three interactions supported by sliding: *slide* ⬌ **(b),** *extend* ◄|► **(c), and** *swap* ⇆ **(d). To optimize monitor space and minimize interactions, during the implementations, we removed the** *slide* ⬌ **(b) and made adjustments to** *extend* ◄|► **(c) and** *swap* ⇆ **(d), which will be explained in the Section 4.3.**

rather than supporting a complete IV comparison workflow, which may include capturing, organizing, annotating, and editing IV contents. Thus, users do not define their own ROIs in our techniques. Instead, all techniques visualize one predefined ROI per IV, representing the primary area of interest in each IV.

For all techniques except SideBySideIn2D, the two IVs can be fully or partially overlaid; therefore, we provide a *peek* 👁 interaction [9], which temporarily reveals a small region of the back-layer IV or the other IV. All techniques offer a zoomed-in local view and a global minimap (azimuthal equidistant projection), visualize ROI trajectories, and include playback controls (play, pause, seek).

For techniques that support independent movement of each IV display area (i.e., SlideInVR, SlideIn2D, and SideBySideIn2D) , we implemented on-demand, ROI-based adjustments of the display areas to facilitate the comparison process. To help users quickly distinguish the two IVs, we applied consistent color coding across techniques: pink for Video A and blue for Video B.

We used Unity (2022.3.16f1) and the Meta XR Interaction SDK (v66.0) to implement the two VR techniques (i.e., SlideInVR and ToggleInVR) and run the system on a Meta Quest 3, which has a 96° vertical field of view (vFoV) and a 104° horizontal field of view (hFoV). To implement the three 2D techniques, we used Vite and React together with Three.js for WebGL-based rendering of the different IV views.

4.1 SlideInVR

Following the sliding concept (Figure 3), SlideInVR enables users to view two IVs simultaneously in VR. Custom shaders render both IVs in spherical-lune display areas (Figure 5) with corresponding minimaps, and users interact via VR controllers (Appendix Figure 14-a).
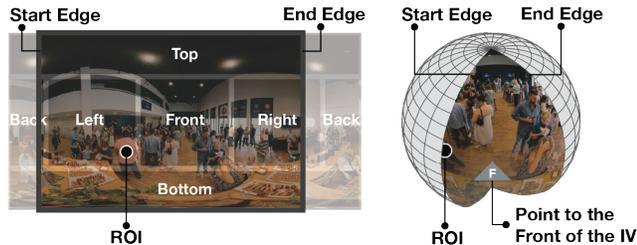
Figure 5: Mapping between the equirectangular IV frame and the spherical-lune display area in SlideInVR for the second tutorial task. Left: a 360° IV [11] from © Grace Church (used with permission) in equirectangular projection, annotated with directions (Front, Back, Left, Right, Top, Bottom) and the predefined ROI (the man in pink shirt). The darker thick rectangle indicates the current display area. Right: the corresponding display area rendered on the lune sphere.

4.1.1 Extend, Slide, Swap, and Peek. SlideInVR exposes four core interactions: *extend* ↔ , *slide* ↔ , *swap* ⇄ , and *peek* 👁 . With either controller, users grab and drag an edge to extend ↔ an IV's spherical-lune display area (Figure 6-b), and the size is updated in real time during extension. The edges of the back-layer display areas are rendered thinner, helping users distinguish the order of

layers and reducing visual distraction. A band-shaped selector (pink above the equator for Video A, blue below for Video B) supports *slide* ↔ : users grab and drag the band to rotate the display area around the polar axis, with a gradient indicating the amount of sliding while updates occur on button release to reduce motion sickness (Figure 6-d,e). Pressing any trigger button activates *peek* 👁 , temporarily revealing a small circular region (15°both vertically and horizontally) of the back-layer IV (Figure 7-b).
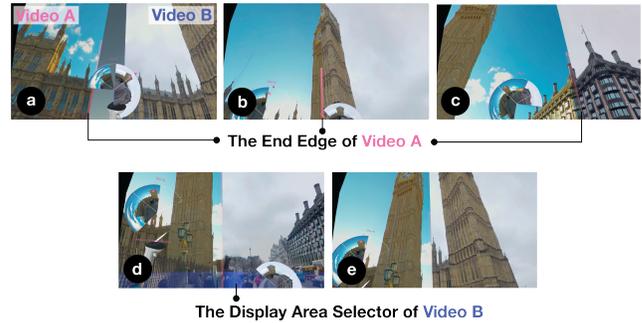
Figure 6: In (a), two display areas do not overlap, and the user plans to *extend* ↔ Video A. In (b), the user is *extending* ↔ Video A by moving the end edge to the right. In (c), the user stops *extending* ↔ , and the Big Ben in Video A (top layer) is overlaid on the Big Ben in Video B (back layer). In (d), the user is *sliding* ↔ the display area of Video B by moving it clockwisely. In (e), the position of Video B is updated once the user stops *sliding* ↔ . The two IVs are from © VR Gorilla, used with permission [12, 13].

4.1.2 ROI-Based Display Manipulation. A hidden radial menu provides ROI-based interactions and view presets (Figure 7-a). ROIs O verlay and ROIs SxS automatically extend ↔ and slide ↔ the display areas so that the ROIs are overlaid (Figure 7-b) or placed side by side (Figure 7-c) in front of the user. Additional items include *Reset Views* (default half-and-half layout with Video A on top), Set Views 360° (full overlay layout, similar to ToggleInVR viewing mode), *Restart Videos* (play all VIs from the beginning), and *Open/Close Panel.*

4.1.3 Minimap. A minimap of Video A or Video B in azimuthal equidistant projection is attached above the left or right controller. It shows the current display area, a FoV indicator, a front-direction arrow ("F"), and an ROI indicator (Figure 8-b). SlideInVR also visualizes ROI trajectories on the minimap, with depth encoding time, and a 3D progress bar that provides better user awareness of video time [47] through the center to show video time. The progress bar is hidden when the minimap faces the user to avoid occlusion and reappears when the controller is rotated. When the IV is playing, the trajectory and progress bar move along the depth axis while the minimap itself stays fixed to the controller to reduce motion sickness. Below the minimap, we show the current and total duration. When the two controllers are close together, their minimaps, progress bars, and ROI trajectories overlay spatially, with the top-layer IV's minimap rendered above the back layer.
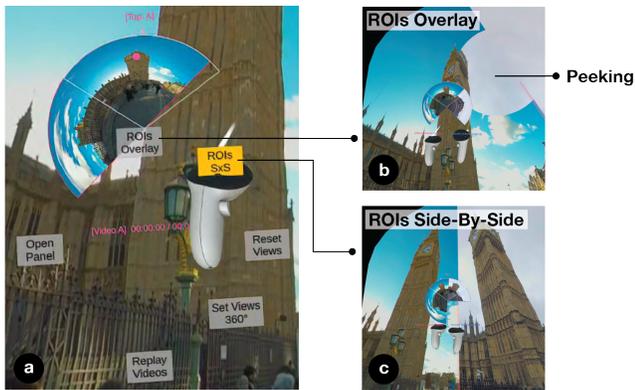
Figure 7: The hidden radial menu of SlideInVR advanced features surrounding the left controller in (a): *ROIs Overlay* (top), *ROIs SxS* (top right; selected and highlighted), *Reset Views* (right), *Set Views 360 °* (bottom right), *Restart Videos* (bottom), and *Open/Close Panel* (left). (b) ROIs SxS aligns the ROIs (Big Ben) side by side. (c) ROIs Overlay aligns the same ROIs overlaid. The two IVs are from © VR Gorilla, used with permission [12, 13].

4.1.4 Temporal Navigation. Users navigate an IV using natural body movements. Navigation for Video A and Video B can be controlled independently using the left and right controllers, respectively, also allowing for navigating both IVs simultaneously. They can *seek* to a specific time in the corresponding IV(Figure 8-a). Moving the controller downward toggles play and pause for the corresponding IV. In addition, moving the joystick forward or backward on either controller triggers a 5-second rewind or fast-forward, respectively, for the corresponding IV.

4.2 ToggleInVR

ToggleInVR displays a single IV on the entire sphere, similar to how users watch IVs in state-of-the-art systems. With overlapping IVs, *peek* 👁 reveals a small region of the other IV, and "B" toggles between them. The minimap, progress bar, and spatially visualized ROI trajectory follow the same design as in SlideInVR. However, only the current IV's minimap is visible, and the projection always shows its full overview. All other interactions for activating advanced features and navigation are the same as in SlideInVR, except that users can only navigate the current viewing video. In the hidden menu, only *Restart Videos* (bottom) and *Open/Close Panel* (left) are provided. No ROI-based advanced features are supported because ToggleInVR presents only one IV at a time, and users are not able to reposition the IV. A full overview of the interaction mappings is provided in Appendix Figure 14-b.

4.3 SlideIn2D

All 2D-based techniques follow the same interface: a row of clickable features, the canvas for displaying IVs, and then temporal navigation controls as shown in Figures 9 and 10.

4.3.1 Physical Navigation. In SlideIn2D, the canvas consists of two display areas. Each IV is shown in its own display area, with the
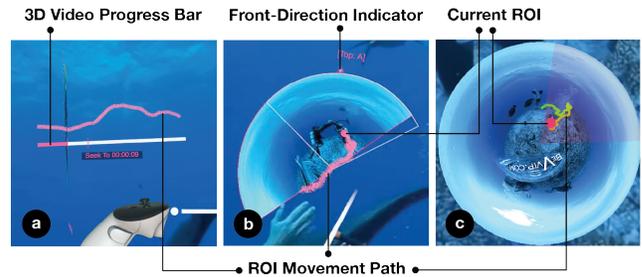

Figure 8: In (a), the user seeks within Video A in SlideInVR by holding the "X" button and moving the left controller forward and backward. (b) demonstrates the spatial-temporal trajectory of the ROI of Video A in SlideInVR; (c) demonstrates the gradient-color trajectory of the ROI of Video A in any 2D techniques (from red to green). Example IV is from © Luxury Travel VR, licensed under CC BY 4.0 [43].

highlighted top-layer display area (partial) overlaid on the back-layer display area. Each IV viewport has a fixed vFoV of 40°, while the hFoV is adjusted according to the size of the browser and the size of the display area. In our implementation, the maximum hFoV of SlideIn2D is ≈ 98°, derived from the symmetric perspective frustum of THREE.js:

$$hFoV = 2 \cdot \arctan\left(\tan\left(\frac{vFoV}{2}\right) \cdot \frac{W}{H}\right),$$

where W/H is the aspect ratio of the display area.

Users can drag the display area to navigate the corresponding IV (Figure 9), where horizontal dragging adjusts the yaw of the viewpoint and vertical dragging adjusts the pitch. They can then use the *Reset* button to return both IVs to their default front-facing views.

4.3.2 Extend, Swap, and Peek. Users drag the slider (one edge of the top-layer IV) left or right to *extend* ◀▶ its display area and reallocate horizontal space (Figure 9-a). To simplify the screen real estate and simplify interactions, SlideIn2D instantiates only a subset of sliding features: users can extend only the top-layer IV via a single edge, while the other edges are fixed to keep the back-layer IV fully visible (Figure 4), ensuring back-layer display area is always fully visible. Thus, the back layer cannot be *extended* ◀▶ or *slid* ⟷ . Users can *swap* ⇄ the layer order to foreground the other IV. Pressing "P" activates *peek* 👁 , temporarily revealing a 300px × 300px (about 21° × 21°) region of the back-layer IV at the mouse position; a "Peeking" label in the upper-right corner indicates this mode.

4.3.3 ROI-Based Display Manipulation. We provide two features for automatically adjusting both IV views based on the current ROI positions. ROIs SxS recenters each ROI within its display area, placing them side by side on the canvas (Figure 9-b), while ROIs Overlay recenters both views so the ROIs overlap on the canvas (Figure 9-c).

4.3.4 Minimap. In the corner of each display area, a minimap shows an azimuthal equidistant overview of the IV, with the current ROI position and a gradient-colored trajectory (red to green) depicting its motion over time, following prior designs [8] (Figure 8-c).
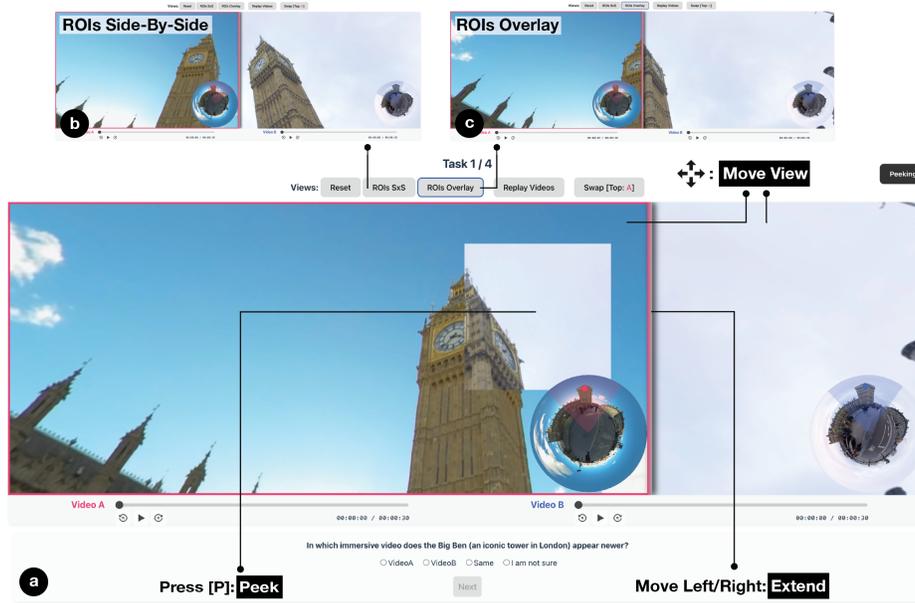
Figure 9: (a) Overview of SlideIn2D showing two IVs [12, 13] from © VR Gorilla (used with permission) and core controls. (b) ROIs SxS aligns the ROI (Big Ben) side by side. (c) ROIs Overlay aligns the same ROI for overlay-based comparison.

4.3.5 Temporal Navigation. Below the IV display areas, temporal controls for Video A (left) and Video B (right) allow users to drag a progress bar to seek, as in a traditional video player. Additional controls include play/pause, 5-second rewind/forward, current/total duration, and a *Replay Videos* button to restart both IVs.

4.4 ToggleIn2D

ToggleIn2D (Figure 10) presents a single FoV that shows one IV at a time. The view (40 vFoV and 98° hFoV) is shared by both IVs, so dragging the view updates the viewpoint of both videos simultaneously. Users can reset to the default front-facing view with *Reset* and switch between IVs using the *Toggle* button (Figure 10). Because a single shared view applied to both IVs, ROI-based advance features, including ROIs SxS and ROIs Overlay , are not supported. Since the IVs are fully overlaid, *peek* 👁 temporarily reveals a small region of the other IV. Navigation controls and minimaps follow SlideIn2D, but users can navigate only the currently visible IV, consistent with ToggleInVR. *Replay Videos* restarts both IVs.

4.5 SideBySideIn2D

SideBySideIn2D (Figure 10) splits the canvas into two equal views, each with 40° vFoV and 59° hFoV: left for Video A and right for Video B. Users can drag each view independently to navigate its IV. ROIs SxS automatically centers each ROI within its FoV. Because views never overlay, *peek* 👁 and ROIs Overlay are not supported. The minimap and temporal navigation controls follow the same design as in SlideIn2D.

5 User Study

5.1 Participants and Apparatus

We recruited 20 participants (eight females and 12 males) where eight were aged 18–24, and 12 were aged 25–34. Recruitment was conducted via university mailing lists, posts in relevant local VR Discord communities, and word of mouth. We screened for VR and immersive video (IV) viewing experience and excluded individuals with no prior experience. Participants reported diverse VR experience (from < 1 year to > 5 years) and IV viewing frequency (Appendix Figure 15). Most participants (14 out of 20) had at least one year of VR experience, while 11 out of 20 participants reported using VR occasionally (*N* = 7), regularly (*N* = 2), or frequently (*N* = 2). In contrast, the majority (16 out of 20) reported watching IVs rarely (*N* = 5) or very rarely (*N* = 9). Participants received 20 CAD per hour as a remuneration, and the study was approved by our institutional research ethic office.

All participants completed study tasks on a wheeled office chair. For the three 2D techniques, they used a 28-inch monitor with a keyboard and mouse, and for the two VR techniques, they wore a Quest 3 HMD and used two controllers.

5.2 Tasks

We designed four domain-independent IV comparison tasks [T1–T4] that were derived based on previous work on low-level comparison tasks in information visualization [4] and extended from 2D video comparison tasks [9] (Table 1):

[T1] Identify differences in the temporal occurrences of frames and ROIs. For example, a coach may compare two 360° training recordings of the same drill to determine in which attempt a player enters a key position earlier, performs a move more often, or stays in a critical area for longer.
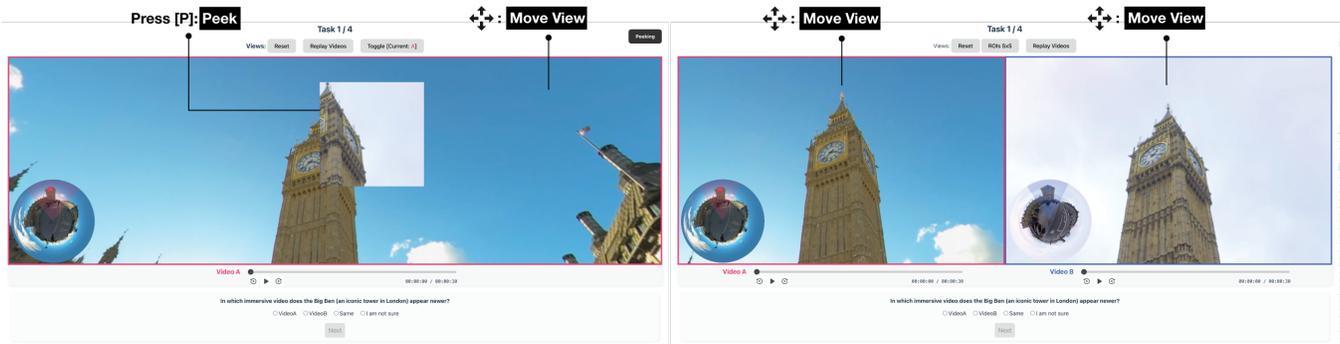
Figure 10: The overview of ToggleIn2D (Left) and SideBySideIn2D (Right) showing two IVs from © VR Gorilla (used with permission) [12, 13].

[T2] **Compare spatial-temporal motion of ROIs.** For example, an IV film director or editor may watch two trials of a scene and compares the actors' movements and transition of postures [34, 49].

[T3] **Find visual differences in the visual attributes of ROIs, such as size, color and static posture.** For example, a historian or environmental scientists may view and compare IVs captured at different times to identify the changes of building appearances or surrounding environments [1, 33].

[T4] **Identify the visual difference between the full immersive video frames.** For instance, an AI scientist may compare video graphics or panorama quality of two AI-generated panorama videos to determine whether frames in one IV are more distorted or more consistent and have less stitches at the edges [44].

5.3 Study Data: Immersive Videos Clips

We selected 20 pairs of 30-second 360° IV clips for the within-subject comparison tasks, covering four comparison tasks [T1–T4] across five techniques. Each pair was associated with a meaningful question requiring participants to compare and identify video content (Appendix Table 2).

Most study stimuli were natural, real-world IVs sourced from YouTube: 18 of the 20 pairs were natural clips. We additionally included 2 pairs of simulated AI-generated IVs specifically for T4, because T4 focuses on assessing global visual differences (e.g., distortions, blurring, and stitching artifacts) that are difficult to obtain with controlled severity from in-the-wild videos. Due to limitations of current generative AI (GenAI) models for IVs (e.g., graphics quality) [44], we simulated "AI-generated" IVs by re-recording real scenes in Unity and applying controlled filters to introduce different levels of distortion, blur, and stitching artifacts.

Across 20 pairs, clips were drawn from (i) the same environment and perspective, (ii) the same environment but different perspectives, or (iii) entirely different environments. We also selected ROIs exhibiting diverse visibility patterns: ROIs may remain visible throughout an entire video, disappear and reappear in the same or different locations (e.g., in clips with transitions), or appear and disappear multiple times. To reduce ambiguity in identifying the intended ROI, six clips were augmented with a brief arrow indicator in the first five seconds to highlight visually similar ROIs.

For each [T1–T3] task , we pre-defined exactly one ROI per IV based on the corresponding question. For ROI visualization, we used a pre-trained open-source 360° tracking model[3] to track and annotate the chosen ROI in each IV clip. In clips where the ROI disappeared and reappeared (e.g., with transitions), separate tracking segments were combined into a single trajectory. No ROI tracking and visualizations were performed for [T4], which required assessing global visual differences rather than comparing localized content.

5.4 Procedure

Participants provided informed consent and completed a demographic pre-survey. We then introduced 360° IVs, azimuthal equidistant projection, and four comparison tasks [T1–T4]. Next, participants completed all five comparison techniques (i.e., SlideInVR, ToggleInVR, SlideIn2D, ToggleIn2D and SideBySideIn2D) in a balanced Latin square order.

In each condition, participants received a brief slide-based introduction to the technique interactions, followed by a short tutorial with four practice tasks using the same four IV pairs, and were encouraged to explore their own comparison strategies. After the tutorial, participants performed four experimental tasks in random order, each corresponding to a unique, non-repeated IV pair from the 20 prepared pairs (Appendix Table 2). After completing four tasks, participants filled out a post-questionnaire on their perceptions of the technique and strategies used. Participants could take short breaks between conditions to reduce fatigue.

After all five conditions, participants took part in a 15-minute semi-structured interview to discuss their perceptions of five techniques, preferences, workflows, and strategies for comparing IVs. The study lasted an average of 2.27 hours per participant (min = 1.70 and max = 2.77). We conducted the study in a single session to keep participants' task and technique experiences fresh for supporting more reliable comparisons.

5.5 Measures

During each comparison task, we recorded participants' performance, including accuracy. In the post-questionnaire, participants reported their strategies and rated their perceptions of task workload and comparison techniques on 7-point Likert scales. These

---

[3]https://github.com/VitaAmbroz/360Tracking

Table 1: Four types of comparison tasks with abstract question templates and selected actual questions.

Task Type	Abstract Question Templates	Selected Actual Questions
[T1]	• In which IV, does the [ROI] appearing or disappearing happen first? • In which IV, does the [ROI] in red color appear more or less frequently? • Does one IV have a accumulated longer duration for the [ROI] (e.g., an object staying in view longer before exiting)?	• In which immersive video does the person in white greet earlier? • In which immersive video does the shark attack more frequently?
[T2]	• In which IV, does the [ROI] move toward or away from a reference or an area (e.g., the front area of an IV)? • In which IV, does the [ROI] moves faster or slower? • In which IV, does the [ROI] (e.g., a person) change posture or facial expression more frequently or more dynamically?	• In which immersive video does the diver wave the paddle more consistently? • In which immersive video does the lemur (indicated by the arrow) move along a more circular path?
[T3]	• Are the two [ROIs] the same object or person? • In which IV, does the [ROI] show more diversity in appearance (e.g., color or shape)? • In which IV, does the [ROI] (e.g., a person) exhibit a more theatrical posture or facial expression at a specific moment?	• Are the two female lions near the camera in each immersive video the same individual? • In which immersive video does the Big Ben (an iconic tower in London) appear newer?
[T4]	• Which IV has better graphics quality (e.g., less noticeable distortion, blurring and stitches)? • Which IV has better frame consistency? • Do the two IVs depict the same environment or location?	• Do the two immersive videos depict the same location? • Which immersive video has better quality (e.g., less noticeable distortion, stitches or blurring)?

ratings included standardized instruments (NASA-TLX [21], UMUX-Lite [27]) as well as custom self-defined questions. In the interview, participants ranked the five techniques by dragging their names on a slide and provided rationales for their rankings.

5.5.1 Qualitative Analysis. In addition to quantitative measures, we also conducted a thematic analysis of participants' free-text responses in the post-questionnaires and the transcripts of the semi-structured interviews. The first author, who ran all the study, generated an initial codebook related to (i) workflows of IV comparison, (ii) comparison strategies (e.g., toggle vs. side-by-side, parallel vs. sequential), and (iii) preferences and perceived benefits and drawbacks of each technique and modality. The first two authors then independently coded the responses and transcripts by assigning relevant excerpts to these codes and adding, splitting, or modifying codes when new patterns emerged (i.e., a hybrid inductive–deductive approach). Finally, the same two authors grouped related codes into higher-level themes, resolving disagreements through discussion; a third author acted as a tiebreaker when needed.

6 Results

In this section, we report how modality and comparative approach shaped participants' workflows, strategies, and subjective experience when comparing IVs. Our analysis combines qualitative observations and interviews with quantitative ratings and accuracy data. Across techniques, we did not observe significant differences

in task accuracy (Section 6.4); instead, the main effects of the techniques were reflected in how participants navigated, compared, and experienced the tasks.

6.1 Overall Workflows

Across techniques, participants followed a two-phase workflow during IV comparison: navigation and comparison. Participants first moved within the 360° space to "find," "locate," "navigate," or "track" predefined or self-defined ROIs and then applied the suitable comparison strategies. For [T1–3], most participants typically located the ROI(s) in one or both IVs and then compared using parallel or sequential viewing (Section 6.3). For [T4], participants often treated the task as iterating [T3] multiple times. Some participants also created their own anchor points to support global comparison (e.g., P18 *"tried to manually select anchors and then compare for a fair comparison for global visual comparison."*).

6.1.1 Phase 1: Spatial Navigation. In the study, participants showed different workflows with different techniques. In SlideInVR, SlideIn2D and SideBySideIn2D, participants heavily used ROI SxS and ROIs Overlay features to automatically find the ROIs before comparing. Many described ROI SxS as a favorite feature because it reduced search time and effort to locate ROIs (P1,7-9,11,12,14,16,20). *"I liked the ROIs side by side (ROIs SxS) a lot, that was the MVP. It just fixes everything for you, as long as the ROI isn't moving that much."* -P11

After initial ROI alignment, we observed two common behaviors: some participants continued using ROI-based view adjustment to follow moving ROIs, while others preferred manual dragging to

maintain spatial context in 2D. *"When I was dragging (in 2D), I got a little more understanding about the scenery around me ... If I directly only rely on the button. I feel like it kind of locks me into one of viewpoint, and I have no understanding."* -P20

For ToggleIn2D and ToggleInVR, because the two IVs share the same view, there was no way to make the two ROIs overlaid or side by side. Thus, participants heavily relied on the minimap and manual navigation to locate ROIs. P8 described, *"they don't have ROIs overlay, which means that you need to have a lot of navigation."*

6.1.2 Phase 2: Comparison. Once ROIs were located, the technique strongly shaped comparison style. Participants described SlideIn2D and SlideInVR as the most "powerful" and "flexible" because they supported fluid switching between side-by-side and overlay views, along with space reallocation via extending ◀▶ . As P2 summarized, *"SlideInVR is the one with more possibility."* -P2

By contrast, SideBySideIn2D enforced persistent side-by-side viewing, while ToggleIn2D and ToggleInVR often pushed participants toward sequential viewing, especially when ROIs moved or appeared in different locations. As P9 described, *"Toggling both in VR and in 2D was somewhat horrible for tracking motion. When objects are moving in a very different way or show in drastically different locations, toggling back and forth doesn't let you track both. You still have to go through each video individually."* -P9

The peek ◉ feature partially mitigated this limitation by providing a concurrent local view when ROIs were not aligned, enabling a form of real-time side-by-side comparison (P3,6): *"For ToggleInVR, I find the peeking was extremely helpful. Without it I had to toggle all the time and could not compare in real time. But with the peeking, I could just compare the point of interest simultaneously."* -P3

6.2 Physical and Temporal Navigation

We next describe how techniques shaped navigation, both physically (viewpoint changes in 360° space) and temporally (timeline control).

6.2.1 Physical Navigation. Few participants found 2D navigation easier due to less body movement and familiar mouse input (P5,7,8), whereas the majority described VR navigation as more natural and intuitive (P1-7,10-12,14,16-18,20) especially for shifting attention across directions or following fast-moving ROIs (P1,5,7,13,16,20). As P18 described, *"for the shark (fast-moving ROIs), in 2D, it was very hard to keep track of the ROI, but in VR it would have been much easier because I can just turn my head a bit."*

VR also fostered a stronger sense of spatial understanding through embodied interaction (P9,10,14–16,19). As P14 described, *"... on a flat screen (2D techniques), I can forget where things are because I drag to get there. But in VR, I just know it's back there, and I just look there."* In addition, the larger FoV in VR reduced the need for manual navigation during comparison, because participants could capture more relevant content with "a little bit" (P8) head rotations.

However, the larger FoV in VR also increased visual information and could raise cognitive demand (P3,8,9). In addition, some participants reported a "dizziness" feeling (P9) and disorientation during full-scene toggling in ToggleInVR (P9,11,15). As P15 noted, *"... when it completely changes, I have that brief half second moment was like, 'where am I?' And then I have to focus on it again... At least*

the SlideInVR lets you control how much changes, because there's that idea of multiple sections."

Many participants found the minimap helpful for locating and verifying ROIs, especially for moving targets (P4-7,11,14,15,17,20) and when using 2D techniques (P11,14). As P20 reflected, *"The lemur jumped off the screen, and I couldn't figure out whether it was the same one sitting there or another one. Then I had to look at the minimap and see that I was following the wrong lemur."* Some also used the minimap for comparing distances or paths through visualized trajectories (P4-6,12,15,18), setting up display areas in SlideInVR (P3) and providing a global overview of either the entire frame (P6,19) or large ROIs (P10,18,19).

6.2.2 Temporal Navigation. Several participants reported that temporal control felt more intuitive in VR, allowing play/pause/seek with less attentional shift (P1,5,8,12). For instance, P1 noted, *"For the 2D, I had to pause or play when I'm focusing on what I'm doing on the screen, It takes more effort and it causes more delay... But in VR... I can just unintentionally use my hands to do that motion, and I don't have to put my attention towards my hands and the replay buttons."*

6.3 Comparison Strategies

Participants' strategy choices were strongly task-dependent rather than uniformly tied to a particular technique. Almost all participants (except P10) noted that they used or would use different comparison approaches depending on the task. In this section, we describe how different task characteristics shaped participants' preferences and the strategies they employed.

6.3.1 Side-by-side vs. Toggle. In most tasks, many participants preferred to use side-by-side (P1,4-7,9,11,12,15,16), as perceived as *"a natural and straightforward comparison method that they are already familiar with"* (P3,6,14-16). They also explained that side-by-side reduced the fear of *"missing something"* (P7-9,11,14,16) and lowered memory workload by eliminating the need to recall details (P8,11) since both videos remained visible during comparison. As P11 noted, *"it is beneficial, because sometimes if you're doing it one at a time, you're relying on your brain to remember everything correctly, and sometimes it just doesn't. But seeing them side by side sometimes reveals information that you might have missed."* Side-by-side also supported time efficiency (P3,6,14) and was effective for comparing time steps or tracking differences across temporal sequences (P3) by playing two IVs simultaneously.

Toggle was often preferred for fine-grained visual differences (e.g., subtle shape or quality differences) (P3,6,8,9,12,15-18). Reflecting on a [T3] task, P8 stated: *"I remember the lion posture task clearly. I used the overlay (toggle) very quickly, because it gave me very clear differences between how [the lion posture in] the top layer is different from [the lion posture in] the second layer."*P8 This approach was especially preferred when comparing ROIs with similar shapes appearing in the same location or perspective (P1,5-8,13,16,19,20). However, toggle became ineffective when ROIs were moving or when videos were captured from different perspectives, and participants switched to side-by-side (P1-3,5-7,9,16,18-20). As P16 summarized, *"for example, for the video quality in the same location, peek (toggle) works quite nicely. But for a lot of the videos that actually*

don't match together, peeking (toggling) a specific region wouldn't show much difference." These results align with Yuan et al. 's findings that for detecting subtle local and global visual differences in static 360° images, toggle is preferred and perceived as more effective than placing ROIs side by side.

Overall, no single approach was sufficient across all tasks. Participants therefore stressed the importance of supporting flexibility in switching between comparison approaches (P1-4,6,8,11,12,15-19). As P11 noted, *"each of these tasks would shine in one of the tools ... some of them are just better if you watch them side by side ... some of them are very good if you can use a peek with the overlay (toggle)."*

6.3.2 Parallel vs. Sequential. Participants exhibited two multitasking approaches: parallel and sequential. While some favored parallel (P9,11,14,15) and others preferred sequential (P7,13,19), most reported that their choice depended on the task. Participants used parallel viewing by rapidly shifting attention when differences were subtle (P3,12), ROIs were static (P7,16,17), or videos are long (P12,14). As P12 explained, *"the good thing is the video is short, so you don't need to memorize a lot ... But when there are many things to remember with only tiny differences, like in video A I took six donuts, but in video B I took seven, it's really hard to memorize. That's why I don't prefer watching only one video at a time."*

However, when ROIs were moving substantially or quickly, participants switched to watching IVs sequentially (P7,9,11,13,16,17,20). As P17 reflected on a [T2] task: *"If the movement is very mild, then parallel is so good... but the lemur was very hard because it was moving a lot, and I had to drag my mouse to move my view to follow it. "* Besides tasks that demanded more physical movements, participants also preferred sequential multitasking when tasks imposed high mental workload, such as memorization, counting, or comparing multiple objects (P3,7,9,12-14,16,18,20). As P3 explained: *"For really, really simple tasks I can still view the two videos at the same time. But if it's a bit harder, like if I need to count the donuts or compare how many times the shark attacks, I need to carefully watch one video after another. In that case, I would prefer to use toggle, or even if they are side by side, I would only play one video at a time."*

6.4 Quantitative Results

We report quantitative results for perceived helpfulness, workload, usability, and task accuracy. Because Likert and NASA-TLX ratings deviated from normality (Shapiro–Wilk, $p < .05$), we used Friedman tests with Wilcoxon signed-rank post-hoc tests with Holm–Bonferroni correction. Trial-level accuracy (binary) was analyzed using a mixed-effects logistic regression with participant as a random effect. In line with common HCI practice, we treat $p < .05$ as evidence against the null hypothesis and consider larger p-values as descriptive rather than inferential.

6.4.1 Helpfulness. Participants rated how helpful they found each technique for each task type Figure 11. For [T1] (temporal occurrence), a Friedman test indicated an omnibus effect of Technique ($p = .017$), but no pairwise comparisons reached significance ($p = .064$) after Holm–Bonferroni correction. For [T2] (spatial location), [T3] (visual difference), and [T4] (event sequence), Friedman

tests did not reveal significant differences in helpfulness across techniques ($p \geq .19$). Detailed statistical results and additional rating distributions are provided in Appendix Table 4 and Figure 16.

6.4.2 Workload. We collected participants' perceptions of their experience using the NASA-TLX questionnaire (Figure 12). Friedman tests reveal no significant differences for most subscales or average ratings, except for physical demand ($p < 0.001^{***}$). Post-hoc tests did not yield statistically significant pairwise differences for physical demand, including SlideInVR vs. SlideIn2D ($p = .072$), SlideInVR vs. ToggleIn2D ($p = .061$), and SlideInVR vs. SideBySideIn2D ($p = .072$). Descriptively, mean physical demand ratings for SlideInVR were higher than for the 2D techniques (Appendix Table 5), which is consistent with participants' reports that VR required more physical effort (e.g., head and body movements).

These results are also reflected in the interviews. Participants appreciated its flexibility and power, but many also described it as demanding to use. Many participants reported that SlideInVR imposed a high workload to operate, remember and recall its controls (P2,4,8,10,12,15,20). As P2 explained: *"SlideInVR is the one with more possibility, but it's the most overwhelming... SlideInVR is amazing, but the workload is too high. Maybe if I was working with it for a couple of weeks, that would be automatic."* Participants also reported difficulty in distinguishing and managing the two video layers (P5,7), which requires more mental and physical effort during IV comparison. As P5 noted, *"For SlideInVR, it's really not that easy to tell apart which layer is on top [layer]. When you're looking at a video, you don't know which layer it is without moving it away to see what is below."*

6.4.3 Usability. We assessed perceived usability with the UMUX-Lite scale. The Friedman test did not reveal a significant difference in usability ratings across the five techniques ($p = .08$). Descriptively, SideBySideIn2D tended to receive the highest median usability ratings and ToggleIn2D the lowest (Appendix Table 6), but these differences did not meet our criterion for statistical significance.

6.4.4 Accuracy. In our accuracy analysis, "I am not sure" responses (2.75% of answers) were treated as incorrect, as they indicate that participants could not determine which IV better satisfied the comparison criterion. Item-level inspection (Table 2) showed that three tasks (8, 10, and 19) had the lowest accuracy (30–40%). Per-technique accuracy on these tasks also varied considerably, and some Technique × Task combinations were represented by only 2–7 trials, reflecting the limited sample size and randomized task assignment.

We modeled trial-level accuracy using a mixed-effects logistic regression with Technique and TaskType as within-subject fixed effects and a random intercept for participant. The model revealed no significant main effect of Technique ($p \geq .313$), no significant main effect of TaskType ($p \geq .19$), and no significant Technique × TaskType interaction ($p \geq .24$). Mean accuracies ranged from 0.64 to 0.75 across techniques (Appendix Table 3), suggesting that, within our study, the different techniques and task types did not lead to systematic differences in correctness.

6.5 Learnability and Preferences

6.5.1 Learning Curve. In general, participants perceived 2D techniques and controls as more familiar, while VR required additional adaptation. As P9 described: *"I think if I grew up using VR, then*
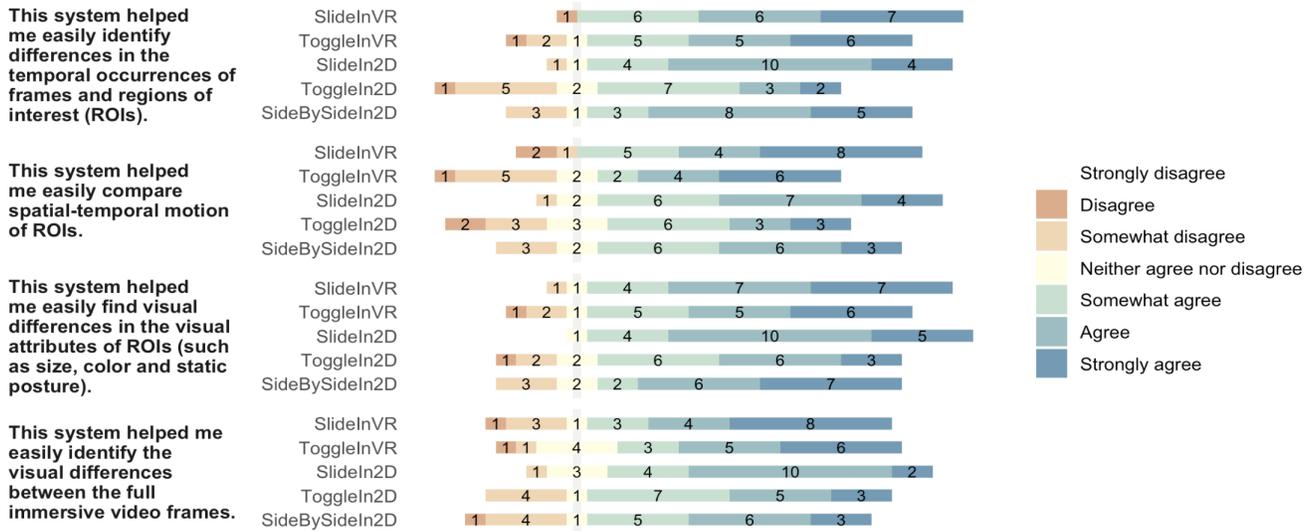
Figure 11: Stacked bar charts of how participants thought that each comparison techniques was helpful for four types of comparison tasks.
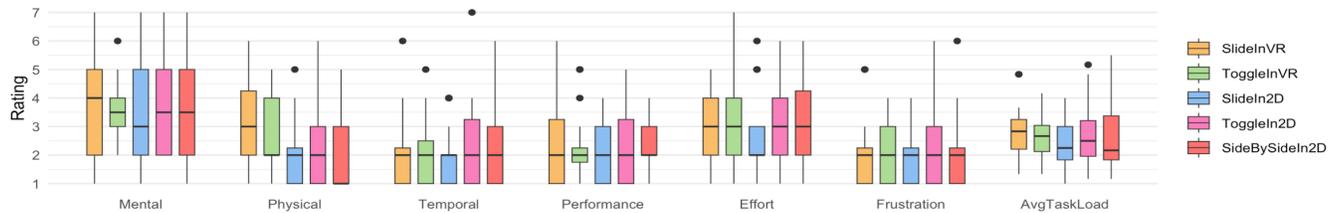

Figure 12: Box-and-whisker diagrams of unweighted NASA-TLX ratings on a 7-point Likert scale (the lower the better).

I'd feel less comfortable with 2D. But for now, it's a little bit hard for me to adapt [to VR]." This challenge was particularly evident in SlideInVR. Although we provided sufficient time for participants to familiarize themselves with each technique, most participants reported that SlideInVR involved a steep learning curve (P1–3,5,7,8,10,12,13,15,17,18). Many described feeling overwhelmed at first, as they had to grasp a new interaction concept while also learning to operate numerous features. As P12 explained: *"I think that (SlideInVR) is a promising technique, but for me, it's a little bit overload, because I didn't have past experience working with that, and I spent a really long time to figure out which button I should press, or how can I achieve my goal. But I think that is a decent technique to help me to catch the differences between two videos."*

6.5.2 User Preference. For the overall experience, most participants ranked SlideInVR and SlideIn2D as their top two favorite comparison techniques, and ToggleIn2D as the least. Specifically, SlideInVR received 40% first-place votes and 70% top-2 placements; SlideIn2D received 65% top-2 placements, while ToggleIn2D was ranked last by 50% of participants, as shown in Figure 13. A Friedman test reveals a significant effect of technique on preference ($p < 0.001^{***}$). The post-hoc pairwise tests using Wilcoxon signed rank test with Holm–Bonferroni correction indicate a significant difference in preference between SlideInVR and ToggleIn2D ($p = 0.04^{*}$), SlideIn2D

and ToggleIn2D ($p = 0.006^{**}$), and SideBySideIn2D and ToggleIn2D ($p = 0.04^{*}$).

In addition, participants' explanations clustered around several recurring rationales: powerfulness (P1,5,6,8,9,11,14-20) and flexibility (P1,3-6,9,10,12-15,17,19,20); ease of use (P3-5,7,8,11-20); familiarity with the technique (P2,4,7,10-13) and its input devices (P7,12,20); learning curve (P2,3,8,12,13,18); effectiveness of comparison (P3,8,12,19,20); naturalness of navigation (P1,4,6,11-14,16-18); workload (P1,2,4,7,8,16,17,20); and immersive (P5) and intuitive (P9,10,15,16) IV viewing experience brought by VR.

Many participants perceived VR as better supporting comparisons of graphics quality and distortion than 2D (P1,3,6,8-10,12,16,19, 20). This advantage relates to the WYSIWYG nature of VR, where distortions and artifacts felt more immediate and noticeable. As P6 explained: *"It is more natural if you see a distortion in the VR, and you feel that fake immediately."* Participants also emphasized that VR's immersive qualities and sense of presence made subtle differences easier to detect. As P19 described: *"In VR, everything I can view more carefully. I think I'm a part of it so I'm able to view things just better... Texture is more clear in the VR, I'm able to understand things better just because there's more depth to it."* Similarly, beyond VR-specific benefits, some participants also valued SlideIn2D for allowing adjustable screen-space allocation during side-by-side comparison
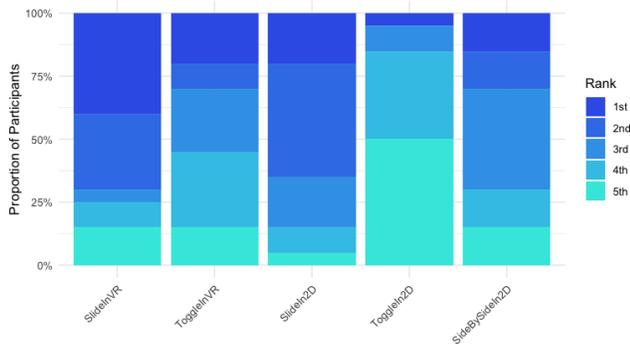
Figure 13: Stacked bar charts of comparison technique rankings from the most favorite (1st) to the least favorite (5th).

(P3,7,9,11,15,16,20). As P15 pointed out, *"SideBySideIn2D is already an upgrade... but sometimes I don't really care about a video that much. So I want to make sure that I'm looking at 80% of one video, instead of 50/50 all the time. So it was nice to have that option, and the good thing about SlideIn2D is that even though there was an extra option, it felt very natural and it didn't add mental demand."*

In summary, our results indicate that IV comparison follows a two-phase workflow: users first navigate to relevant ROIs and then apply comparison strategies. During navigation, users' perceptions were shaped by different modalities and techniques: VR made spatial and temporal navigation feel more natural and helped participants judge quality and distortion, but it also came with higher physical demand and a steeper learning curve. Participants' preferences for comparison strategies were strongly task-dependent. They switched between side-by-side and toggle viewing, and between parallel and sequential viewing, based on how fine-grained the comparison was, where and how the ROI moved, task complexity, and memory demands, rather than relying on a single approach. SlideInVR and SlideIn2D were perceived as more powerful and flexible because they enabled fluid switching between side-by-side and toggle-style comparisons, whereas ToggleIn2D was consistently the least preferred. Within the scope of our study, these differences in workflows, strategies, and user experience did not translate into systematic accuracy differences across techniques and task types, suggesting that the primary benefits of the techniques lie in how they shape comparison processes rather than raw performance outcomes.

7 Discussion

7.1 When Sliding Helps (and When It Does Not)

Our study indicates that supporting both side-by-side and overlay-style comparison is important for IV comparison tasks. The sliding concept provides a flexible way to move between these comparison approaches, enabling users to adjust how much of each video is visible and to choose the comparison style that best fits the current task. Note that our study was conducted only on tasks comparing two IVs, and thus future studies might be needed to assess the insights below for multiple IV comparison.

In 2D, sliding helped participants make better use of limited screen space and attention by allocating more area to the view that

was most important for the current comparison. Participants also found sliding particularly useful when tracking moving ROIs, since they could adjust how much of each video was visible instead of relying on a fixed 50/50 split. Although we used a relatively large monitor (28 inches), several participants still used sliding to hide less relevant content and focus on key regions or fast movements. We believe these benefits to be even more pronounced on smaller displays.

In VR, the sliding concept helps to reduce the physical movements of heads and bodies to switch between different views during comparison tasks, especially for those objects located in different locations. Also, as P15 and several other participants reported, it could help reduce large visual changes and motion sickness when comparing IVs in VR.

These qualitative findings help explain why sliding techniques were often preferred and perceived as powerful and flexible, even though our quantitative analysis did not show clear accuracy differences between techniques. However, SlideInVR is still a novel prototype that introduces a multi-layer representation of IVs. On the one hand, having multiple display areas can partially break immersion, as participants' attention is drawn to the interface rather than the scene itself. On the other hand, our current design only supports comparing two IVs at a time, and many participants already found it difficult to distinguish the layer order in VR. Extending sliding to more than two simultaneous IVs in VR would therefore require additional design work to avoid visual clutter and confusion, and the benefits are not guaranteed [48].

In addition, sliding primarily supports *how* users allocate visual attention, manage motion, and control transitions, rather than directly improving correctness on short, well-defined comparison questions. For designers, this suggests that sliding is most valuable when users must repeatedly reframe views, track moving ROIs, or work within limited visual space, whereas simpler techniques may remain sufficient for short, low-workload comparisons.

7.2 Technique-Specific Trade-offs

In order to understand how different approaches (i.e., toggle, side-by-side, and combined) support IV comparison, we implemented five different techniques in our study. In the following, we synthesize our key findings into technique-level strengths and limitations, which could shed light on future technique development.

- **SlideInVR and SlideIn2D** offered the greatest *flexibility* and perceived *power*, allowing participants to fluidly switch between side-by-side and toggle comparison and to reallocate space for ROIs. Their ROI-based advanced features also facilitated faster navigation. This flexibility allowed them to support a wide range of comparison tasks and made them particularly suitable for more complex or unfamiliar IV comparisons that required switching between different comparison strategies. They were strongly preferred overall, but the flexibility came with higher workload and a steeper learning curve, especially in VR. Their advantages therefore manifested primarily in richer workflows and user preference, rather than in significantly higher accuracy.
- **SideBySideIn2D** provided a simple, familiar, and usable baseline for many tasks. It supported continuous side-by-side comparison and was perceived as efficient and low-risk in terms of "missing

something," but lacked the ability to easily switch to overlay-style comparisons. Thus, SideBySideIn2D was particularly suitable for easy, low-workload, or long-duration comparison tasks, as well as for comparing trends in relative large monitor.

- **ToggleIn2D and ToggleInVR** were effective for detecting fine-grained local differences when ROIs were aligned in space, but became cumbersome when ROIs moved or scenes differed substantially. ToggleInVR also introduced disorientation due to abrupt scene changes. Participants rarely preferred these techniques overall, despite their usefulness for specific tasks. Thus, both ToggleIn techniques were particularly suitable for comparing short IVs captured from the same camera or for high-workload, detail-focused comparison tasks.

8 Limitations and Future Work

Our study examined how users adopted strategies across techniques and tasks, revealing both the promise and current challenges of five IV comparison techniques. We next highlight key limitations and opportunities for future research.

First, although we designed tasks to be as objective as possible, some comparisons remained inherently subjective and were interpreted differently across participants. For example, in Task#13, participants disagreed on whether a lion using its paw to hold the meat constituted a distinct eating posture. In Task#10, participants also varied in how they interpreted "more circular" (e.g., a more geometrically circular shape vs. a more complete circle). Such interpretive differences made some questions more difficult, increased task performance time and limited strictly objective performance evaluation.

Second, participants frequently relied on ROI side-by-side and ROI overlay features to quickly locate relevant content, suggesting strong potential for more autonomous ROI support. A promising direction is to integrate AI that can pre-analyze videos (e.g., track multiple ROIs that may intermittently appear and disappear) and adapt layouts based on task context and users' prior interaction patterns. This could reduce manual tool switching, better maintain ROIs in view, and lower workload during immersive comparison.

Third, our work focused on monoscopic 360° IVs and primarily on *pairwise* comparison between two clips at a time. Future work should explore how these techniques generalize to other IV formats (e.g., stereoscopic 360° and 180°), as well as how interaction and layout should scale to *multi-video* comparison scenarios involving three or more IVs.

Finally, our evaluation used short (30-second) real-world IV clips with a single predefined ROI per clip. Future studies could validate the techniques in domain-specific workflows (e.g., training, film-making, editing, game scene design, and AI-generated IV analysis), and examine richer ROI pipelines where users can annotate, refine, and manage multiple ROIs over longer and more complex videos.

9 Conclusion

In this work, we adapt the concept of *sliding* from 2D image and video comparison and apply it to VR and IV comparison, enabling users to flexibly switch between the side-by-side and toggle comparative approaches. We implement five comparison techniques across 2D and VR modalities and investigated them in a user study

involving four types of comparison tasks. We also examined how different techniques and modalities shaped users' workflows and strategies. Our findings show that users preferred and applied different techniques depending on the task, with SlideInVR and SlideIn2D rated as the two most favored techniques. These two techniques were valued for their powerfulness and flexibility, as they supported smooth switching between comparative approaches and provided better control for allocate space. At the same time, our results reveal challenges, particularly with SlideInVR, which highlights opportunities for refining interaction design to reduce complexity, manage workload, and improve comfort.

References

[1] Rahman Abidin, Nunuk Suryani, and . Sariyatun. 2020. Students' Perceptions of 360 Degree Virtual Tour-Based Historical Learning About The Cultural Heritage Area of The Kapitan and Al-Munawar Villages in Palembang City. *International Journal of Social Sciences and Management* 7, 3 (Jul. 2020), 105–112. https://doi.org/10.3126/ijssm.v7i3.29764

[2] Shahryar Afzal, Jiasi Chen, and K. K. Ramakrishnan. 2017. Characterization of 360-degree Videos. In *Proceedings of the Workshop on Virtual Reality and Augmented Reality Network* (Los Angeles, CA, USA) *(VR/AR Network '17)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3097895.3097896

[3] Randi Altman. 2017. Editing 360 Video in VR (Part 2). https://postperspective.com/editing-360-video-in-vr-part-2/

[4] Robert Amar, James Eagan, and John Stasko. 2005. Low-Level Components of Analytic Activity in Information Visualization. In *Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization (INFOVIS '05)*. IEEE Computer Society, USA, 15. https://doi.org/10.1109/INFOVIS.2005.24

[5] Vera Arents, Pieter C. M. De Groot, Veerle M. D. Struben, and Karlijn J. Van Stralen. 2021. Use of 360° virtual reality video in medical obstetrical education: a quasi-experimental design. *BMC Medical Education* 21, 1 (Dec. 2021), 202. https://doi.org/10.1186/s12909-021-02628-5

[6] Knut Aukland. 2025. 360-video as ethnographic material in the classroom: student teachers' reflection on representations with 360-videos from Mecca. *British Journal of Religious Education* 0, 0 (2025), 1–13. https://doi.org/10.1080/01416200.2025.2575393 arXiv:https://doi.org/10.1080/01416200.2025.2575393

[7] Abdulfatai Olamilekan Babaita, Mayumi Kako, Chie Teramoto, Miho Okamoto, Yoko Hayashi, Shinichiro Ohshimo, Takuma Sadamori, Minoru Hattori, and Michiko Moriyama. 2024. Face-to-face versus 360° VR video: a comparative study of two teaching methods in nursing education. *BMC Nursing* 23, 1 (March 2024), 199. https://doi.org/10.1186/s12912-024-01866-4

[8] Benjamin Bach, Conglei Shi, Nicolas Heulot, Tara Madhyastha, Tom Grabowski, and Pierre Dragicevic. 2016. Time Curves: Folding Time to Visualize Patterns of Temporal Evolution in Data. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 559–568. https://doi.org/10.1109/TVCG.2015.2467851

[9] Adam Baker, Carl Gutwin, Justin Matejka, and Ian Stavness. 2024. Interaction Techniques for Comparing Video. In *Proceedings of the 50th Graphics Interface Conference* (Halifax, NS, Canada) *(GI '24)*. Association for Computing Machinery, New York, NY, USA, Article 5, 13 pages. https://doi.org/10.1145/3670947.3670948

[10] Yoshihiro Banchi, Keisuke Yoshikawa, and Takashi Kawai. 2020. Evaluating user experience of 180 and 360 degree images. *Electronic Imaging* 32, 2 (2020), 244–1–244–1. https://doi.org/10.2352/ISSN.2470-1173.2020.2.SDA-244

[11] Grace Church (YouTube channel: "Grace Church Orlando"). 2025. *Grace 360 | [donut emoji] Step into our lobby | (360° Video) | Grace Church Orlando.* YouTube. https://www.youtube.com/watch?v=R4ks43FvpXM

[12] VR Gorilla (YouTube channel: "VR Gorilla Virtual Reality & 360 Videos"). 2017. *A London City Guided Tour - 360 VR Video (Teaser)*. YouTube. https://www.youtube.com/watch?v=m9EClKA1VeQ

[13] VR Gorilla (YouTube channel: "VR Gorilla Virtual Reality & 360 Videos"). 2025. *A London City Guided Tour in VR -(Teaser) Virtual City Trip - 8K 360 3D*. YouTube. https://www.youtube.com/watch?v=t64Wb8dfmRg

[14] Nihanth W. Cherukuru and Tim Scheitlin. 2020. Visual Comparator: An Interactive Tool for Dynamic Spatiotemporal Comparative Visualizations. *Bulletin of the American Meteorological Society* 101, 10 (2020), E1861 – E1869. https://doi.org/10.1175/BAMS-D-19-0266.1

[15] Federico Chiariotti. 2021. A survey on 360-degree video: Coding, quality of experience and streaming. *Computer Communications* 177 (2021), 133–155. https://doi.org/10.1016/j.comcom.2021.06.029

[16] Ching-Ling Fan, Wen-Chih Lo, Yu-Tung Pai, and Cheng-Hsin Hsu. 2019. A Survey on 360° Video Streaming: Acquisition, Transmission, and Display. *ACM Comput. Surv.* 52, 4, Article 71 (Aug. 2019), 36 pages. https://doi.org/10.1145/3329119

[17] Kevin Fan, Liwei Chan, Daiya Kato, Kouta Minamizawa, and Masahiko Inami. 2016. VR planet: interface for meta-view and feet interaction of VR contents. In *ACM SIGGRAPH 2016 VR Village* (Anaheim, California) (SIGGRAPH '16). Association for Computing Machinery, New York, NY, USA, Article 24, 2 pages. https://doi.org/10.1145/2929490.2931001

[18] Dustin Freeman, Stephanie Santosa, Fanny Chevalier, Ravin Balakrishnan, and Karan Singh. 2014. LACES: live authoring through compositing and editing of streaming video. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) (CHI '14). Association for Computing Machinery, New York, NY, USA, 1207–1216. https://doi.org/10.1145/2556288.2557304

[19] Michael Gleicher, Danielle Albers, Rick Walker, Ilir Jusufi, Charles D. Hansen, and Jonathan C. Roberts. 2011. Visual comparison for information visualization. *Information Visualization* 10, 4 (Oct. 2011), 289–309. https://doi.org/10.1177/1473871611416549

[20] Ruairi Griffin, Tobias Langlotz, and Stefanie Zollmann. 2021. 6DIVE: 6 Degrees-of-Freedom Immersive Video Editor. *Frontiers in Virtual Reality* 2 (2021). https://doi.org/10.3389/frvir.2021.676895

[21] 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. https://doi.org/10.1177/154193120605000909 arXiv:https://doi.org/10.1177/154193120605000909

[22] Kevin Huang, Jiannan Li, Mauricio Sousa, and Tovi Grossman. 2022. immersivePOV: Filming How-To Videos with a Head-Mounted 360° Action Camera. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 91, 13 pages. https://doi.org/10.1145/3491102.3517468

[23] International Telecommunication Union. 2020. *Subjective test methodologies for 360° video on head-mounted displays*. Recommendation P.919. ITU-T. https://www.itu.int/rec/T-REC-P.919/en

[24] Qiao Jin, Yu Liu, Ye Yuan, Bo Han, Feng Qian, and Svetlana Yarosh. 2024. Virtual Reality, Real Pedagogy: A Contextual Inquiry of Instructor Practices with VR Video. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 665, 21 pages. https://doi.org/10.1145/3613904.3642510

[25] Kyoungkook Kang and Sunghyun Cho. 2019. Interactive and automatic navigation for 360° video playback. *ACM Transactions on Graphics* 38 (07 2019), 1–11. https://doi.org/10.1145/3306346.3323046

[26] Jihyung Kim, Kyeongsun Kim, and Wooksung Kim. 2022. Impact of Immersive Virtual Reality Content Using 360-Degree Videos in Undergraduate Education. *IEEE Transactions on Learning Technologies* 15, 1 (2022), 137–149. https://doi.org/10.1109/TLT.2022.3157250

[27] James R. Lewis, Brian S. Utesch, and Deborah E. Maher. 2013. UMUX-LITE: when there's no time for the SUS. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) (CHI '13). Association for Computing Machinery, New York, NY, USA, 2099–2102. https://doi.org/10.1145/2470654.2481287

[28] Jiannan Li, Jiahe Lyu, Mauricio Sousa, Ravin Balakrishnan, Anthony Tang, and Tovi Grossman. 2021. Route Tapestries: Navigating 360° Virtual Tour Videos Using Slit-Scan Visualizations. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) (UIST '21). Association for Computing Machinery, New York, NY, USA, 223–238. https://doi.org/10.1145/3472749.3474746

[29] Yen-Chen Lin, Yung-Ju Chang, Hou-Ning Hu, Hsien-Tzu Cheng, Chi-Wen Huang, and Min Sun. 2017. Tell Me Where to Look: Investigating Ways for Assisting Focus in 360° Video. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (CHI '17). Association for Computing Machinery, New York, NY, USA, 2535–2545. https://doi.org/10.1145/3025453.3025757

[30] 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* (Québec City, QC, Canada) (UIST '17). Association for Computing Machinery, New York, NY, USA, 255–265. https://doi.org/10.1145/3126594.3126656

[31] Sean J. Liu, Rorik Henrikson, Tovi Grossman, Michael Glueck, and Mark Parent. 2023. RadarVR: Exploring Spatiotemporal Visual Guidance in Cinematic VR. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) (UIST '23). Association for Computing Machinery, New York, NY, USA, Article 86, 14 pages. https://doi.org/10.1145/3586183.3606734

[32] Dizhi Ma, Xiyun Hu, Jingyu Shi, Mayank Patel, Rahul Jain, Ziyi Liu, Zhengzhe Zhu, and Karthik Ramani. 2024. avaTTAR: Table Tennis Stroke Training with Embodied and Detached Visualization in Augmented Reality. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) (UIST '24). Association for Computing Machinery, New York, NY, USA, Article 35, 16 pages. https://doi.org/10.1145/3654777.3676400

[33] David M. Markowitz, Rob Laha, Brian P. Perone, Roy D. Pea, and Jeremy N. Bailenson. 2018. Immersive Virtual Reality Field Trips Facilitate Learning About Climate Change. *Frontiers in Psychology* Volume 9 - 2018 (2018). https://doi.org/10.3389/fpsyg.2018.02364

[34] W. Murch. 2001. *In the Blink of an Eye: A Perspective on Film Editing*. Silman-James Press. https://books.google.ca/books?id=wB7cAAAAMAAJ

[35] Cuong Nguyen, Stephen DiVerdi, Aaron Hertzmann, and Feng Liu. 2017. Vremiere: In-Headset Virtual Reality Video Editing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (CHI '17). Association for Computing Machinery, New York, NY, USA, 5428–5438. https://doi.org/10.1145/3025453.3025675

[36] Northwestern University Knight Lab. [n. d.]. Juxtapose JS: Easy-to-make frame comparisons. Web tool. https://juxtapose.knightlab.com/ Accessed: 2026-01-21.

[37] Amy Pavel, Björn Hartmann, and Maneesh Agrawala. 2017. Shot Orientation Controls for Interactive Cinematography with 360 Video. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (Québec City, QC, Canada) (UIST '17). Association for Computing Machinery, New York, NY, USA, 289–297. https://doi.org/10.1145/3126594.3126636

[38] Philipp Rosendahl and Ingo Wagner. 2023. 360° videos in education – A systematic literature review on application areas and future potentials. *Education and Information Technologies* 29, 2 (Jan. 2023), 1319–1355. https://doi.org/10.1007/s10639-022-11549-9

[39] Adobe Communications Team. 2016. Let's Get Experimental: Behind the Adobe MAX Sneaks. https://blog.adobe.com/en/publish/2016/11/04/lets-get-experimental-behind-the-adobe-max-sneaks

[40] C. Tominski, C. Forsell, and J. Johansson. 2012. Interaction Support for Visual Comparison Inspired by Natural Behavior. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (Dec. 2012), 2719–2728. https://doi.org/10.1109/TVCG.2012.237

[41] Christian Tominski, Camilla Forsell, and Jimmy Johansson. 2012. Interaction Support for Visual Comparison Inspired by Natural Behavior. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2719–2728. https://doi.org/10.1109/TVCG.2012.237

[42] Alissa Vermast and Wolfgang Hürst. 2023. Introducing 3D Thumbnails to Access 360-Degree Videos in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (2023), 2547–2556. https://doi.org/10.1109/TVCG.2023.3247462

[43] Luxury Travel VR. 2025. *Amazing DOLPHIN ENCOUNTER in VR! Swim with Wild Dolphins in Stunning 360° 6K Virtual Reality ! [dolphin emoji][water wave emoji]*. YouTube. https://www.youtube.com/watch?v=-YABpx1D5j0

[44] Qian Wang, Weiqi Li, Chong Mou, Xinhua Cheng, and Jian Zhang. 2024. 360DVD: Controllable Panorama Video Generation with 360-Degree Video Diffusion Model . In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 6913–6923. https://doi.org/10.1109/CVPR52733.2024.00660

[45] Xizi Wang, Ben Lafreniere, and Jian Zhao. 2024. Exploring Visualizations for Precisely Guiding Bare Hand Gestures in Virtual Reality. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 636, 19 pages. https://doi.org/10.1145/3613904.3642935

[46] Shoma Yamaguchi, Nami Ogawa, and Takuji Narumi. 2021. Now I'm Not Afraid: Reducing Fear of Missing Out in 360° Videos on a Head-Mounted Display using a Panoramic Thumbnail. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 176–183. https://doi.org/10.1109/ISMAR52148.2021.00032

[47] Mengmeng Yu and Chongke Bi. 2023. Timeline Exploration in 360° Video. In *Proceedings of the 2023 ACM Symposium on Spatial User Interaction* (Sydney, NSW, Australia) (SUI '23). Association for Computing Machinery, New York, NY, USA, Article 24, 12 pages. https://doi.org/10.1145/3607822.3614533

[48] Lin-Ping Yuan, John J. Dudley, Per Ola Kristensson, and Huamin Qu. 2025. Personalized Dual-Level Color Grading for 360-degree Images in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 31, 5 (2025), 2435–2444.

https://doi.org/10.1109/TVCG.2025.3549886

[49] Junjie Zhang, Lik-Hang Lee, Yuyang Wang, Shan Jin, Dan-Lu Fei, and Pan Hui. 2024. Jump Cut Effects in Cinematic Virtual Reality: Editing with the 30-degree Rule and 180-degree Rule. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. 51–60. https://doi.org/10.1109/VR58804.2024.00029

[50] Changrui Zhu, Ernst Kruijff, Vijay M. Pawar, and Simon Julier. 2025. Evaluating 3D Visual Comparison Techniques for Change Detection in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 31, 5 (2025), 3245–3255. https://doi.org/10.1109/TVCG.2025.3549578
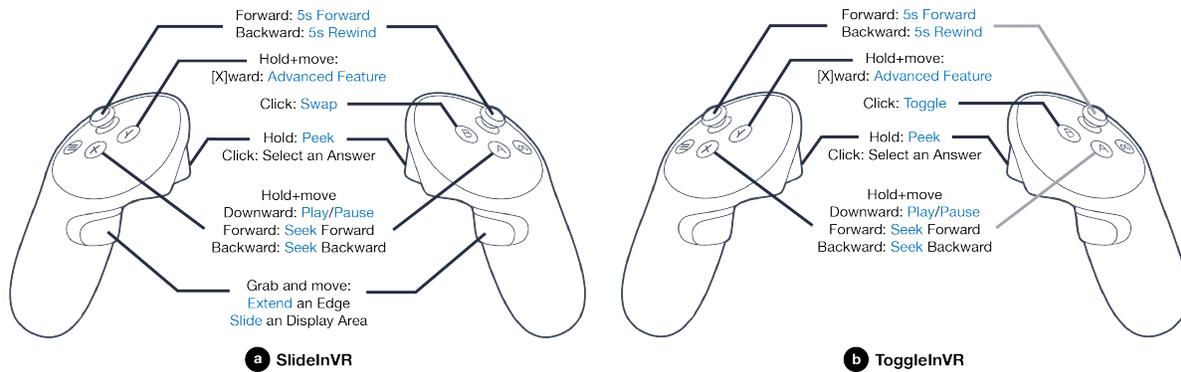
A Appendix

Figure 14: An overview of VR controller interactions in SlideInVR (a) and ToggleInVR (b), illustrating button functions.
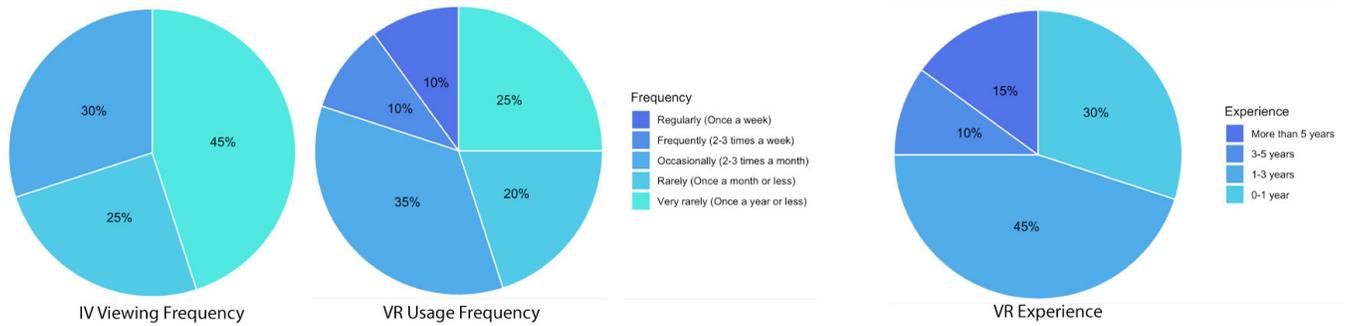

Figure 15: Three bar charts showing distribution of participants' IV viewing frequency (left), VR usage frequency (middle), and VR experience (right).

Table 2: A full task list used in the user study. Each task contains task ID, task type, comparison question, options of the answer, correct answer, whether two videos were applied with filters, and average accuracy in the study. Each video lasts for 30 seconds.

TaskID	TaskType	Question	Options	Answer	Applied Filters	Avg. Accuracy
1	T1	In which immersive video does the person in white greet earlier?	[VideoA, VideoB, Same, I am not sure]	VideoB		100%
2	T1	In which immersive video does the shark attack more frequently?	[VideoA, VideoB, Same, I am not sure]	VideoA		85%
3	T1	In which immersive video does the lemur (indicated by the arrow) spend more time facing away from the viewer?	[VideoA, VideoB, Same, I am not sure]	VideoA		50%
4	T1	In which immersive video does the bear (indicated by the arrow) stop more frequently?	[VideoA, VideoB, Same, I am not sure]	VideoA		55%
5	T1	In which immersive video do more white cars than black cars pass by Sagrada Familia (a famous church in Spain)?	[VideoA, VideoB, Same, I am not sure]	VideoB		75%
6	T2	In which immersive video does the female lion move closer to the observer camera for a longer duration?	[VideoA, VideoB, Same, I am not sure]	VideoA		80%
7	T2	In which immersive video does the aurora move faster in the sky?	[VideoA, VideoB, Same, I am not sure]	VideoB		80%
8	T2	In which immersive video does the cloud move faster in general in the sky?	[VideoA, VideoB, Same, I am not sure]	Same		30%
9	T2	In which immersive video does the diver wave the paddle more consistently?	[VideoA, VideoB, Same, I am not sure]	VideoA		70%
10	T2	In which immersive video does the lemur (indicated by the arrow) move along a more circular path?	[VideoA, VideoB, Same, I am not sure]	VideoA		40%
11	T3	In which video do the fireworks show significantly more diverse or vibrant color changes?	[VideoA, VideoB, Same, I am not sure]	VideoB		70%
12	T3	Are the two female lions near the camera in each immersive video the same individual?	[Yes, No, I am not sure]	No		45%
13	T3	Does the male lion eat the meat in a similar posture in two immersive videos?	[Yes, No, I am not sure]	Yes		80%
14	T3	In which immersive video does the Big Ben (an iconic tower in London) appear newer?	[VideoA, VideoB, Same, I am not sure]	VideoA		65%
15	T3	In which immersive video does the aurora appear more intense at its peak (e.g., brighter)?	[VideoA, VideoB, Same, I am not sure]	VideoB		85%
16	T4	Do all the people who appear in immersive video A also appear in immersive video B, even if their accessories (e.g., headwear) may differ?	[Yes, No, I am not sure]	Yes		80%
17	T4	Do the two immersive videos depict the same location?	[Yes, No, I am not sure]	Yes		95%
18	T4	Which immersive video has better quality (e.g., less noticeable distortion, stitches or blurring)?	[VideoA, VideoB, Same, I am not sure]	VideoB	✓	45%
19	T4	Which immersive video has better quality (e.g., less noticeable distortion, stitches or blurring)?	[VideoA, VideoB, Same, I am not sure]	VideoA	✓	40%
20	T4	Were the two immersive videos recorded around the same time?	[Yes, No, I am not sure]	Yes		75%
Tutorial1	T1	In which immersive video are more donuts taken directly from the left three plates (indicated by the arrow)?	[VideoA, VideoB, Same, I am not sure]	VideoA		
Tutorial2	T2	In which immersive video does the person move a greater distance — the woman in a rose-colored dress with sunglasses in Video A, or the man in a pink shirt with sunglasses in Video B?	[VideoA, VideoB, Same, I am not sure]	VideoB		
Tutorial3	T3	Does the facing direction of the man in a blue shirt and brown trousers during the conversation differ between the two immersive videos?	[Yes, No, I am not sure]	Yes		
Tutorial4	T4	Which immersive video has better quality (e.g., less noticeable distortion, stitches or blurring)?	[VideoA, VideoB, Same, I am not sure]	VideoB	✓	

Table 3: Mean accuracy (proportion correct), standard deviation (SD), and sample size (n) per technique.

Technique	Mean Accuracy	SD	n
SlideInVR	0.64	0.48	80
ToggleInVR	0.65	0.48	80
SlideIn2D	0.75	0.44	80
ToggleIn2D	0.69	0.47	80
SideBySideIn2D	0.64	0.48	80

Table 4: Friedman test results for perceived helpfulness of techniques across the four task types.

Task	p-value
[T1] Identify differences in the temporal occurrences of frames and ROIs.	0.017*
[T2] Compare spatial-temporal motion of ROIs.	0.194
[T3] Find visual differences in the visual attributes of ROIs, such as size, color and static posture.	0.519
[T4] Identify the visual difference between the full immersive video frames.	0.433

Table 5: Friedman test results for NASA-TLX subscales and average ratings across five comparison techniques.

Subscale	p-value
Mental demand	0.948
Temporal demand	0.917
Performance	0.925
Effort	0.365
Frustration	0.935
Physical demand	< 0.001***
Overall workload (average)	0.381

Table 6: Mean UMUX-LITE scores for each technique and pairwise Wilcoxon signed-rank test p-values (continuity correction). Higher scores indicate better perceived usability.

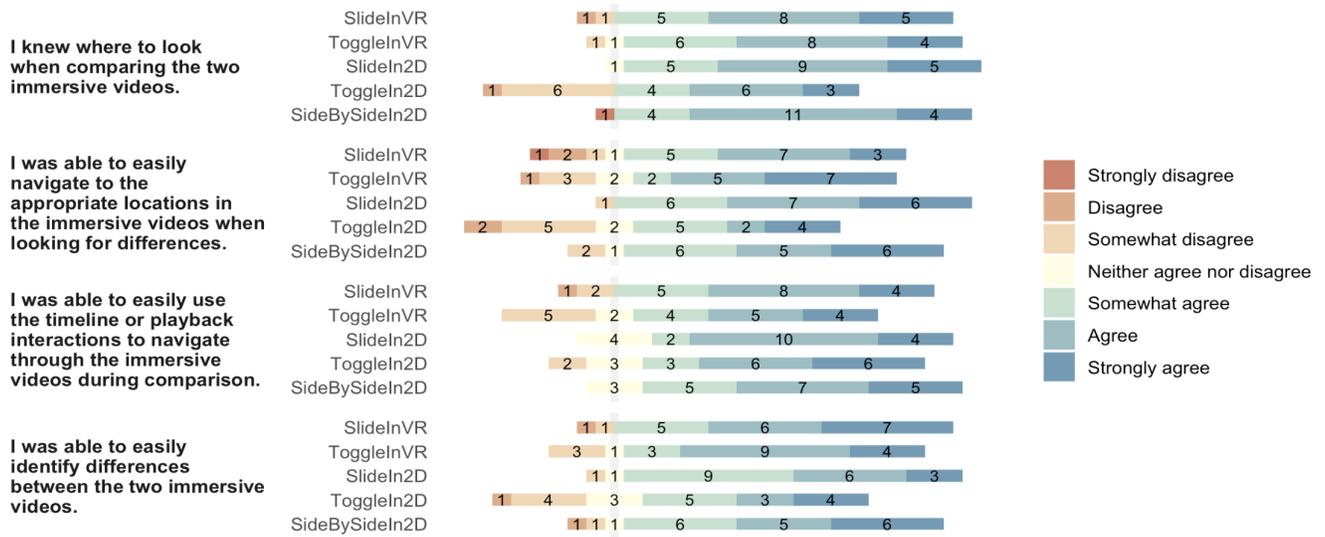| Technique	Mean	SideBySideIn2D	SlideIn2D	SlideInVR	ToggleIn2D
SideBySideIn2D	76.53	-	-	-	-
SlideIn2D	75.98	1.000	-	-	-
SlideInVR	69.48	1.000	1.000	-	-
ToggleIn2D	66.23	0.038*	0.419	1.000	-
ToggleInVR	72.19	1.000	1.000	1.000	0.604

Figure 16: Stacked bar charts of how participants perceived each techniques' physical and temporal navigation experience and overall effectiveness of comparing visual difference.