RESEARCH ARTICLE

Impact of Alignment Edits on the Quality of Experience of 360° Videos

LUCAS S. ALTHOFF[1], **MYLÉNE C. Q. FARIAS**[2], **(Senior Member, IEEE),**
ALESSANDRO RODRIGUES SILVA[1], **AND MARCELO M. CARVALHO**[3], **(Member, IEEE)**

[1]Department of Computer Science, University of Brasília, Brasília 70297400, Brazil
[2]Department of Computer Science, Texas State University, San Marcos, TX 78666, USA
[3]Ingram School of Engineering, Texas State University, San Marcos, TX 78666, USA

Corresponding author: Lucas S. Althoff (190051612@aluno.unb.br)

ABSTRACT Optimizing user quality of experience (QoE) for 360° videos faces two major roadblocks: inaccurate viewport prediction and viewers missing the plot of a story. To tackle these issues simultaneously, alignment edits have emerged as a promising solution. These edits, also known as "re-targeting edits," work by aligning the user's field of view with a specific region of interest in the video content. Despite their potential benefits, there is limited knowledge about the actual impacts of alignment edits on user experience (UX). Therefore, we conducted subjective experiments following ITU-T P.919 methodology to explore their effects on QoE. We proposed an alignment edit based on gradual rotation of the 360° frame, aiming to replicate natural viewing behavior. We tested this approach under various conditions and thoroughly analyzed its impacts using both head motion data and feedback from observers, focusing on their sense of presence, comfort, and perceived experience. The results of our experiments are encouraging. Our proposed gradual alignment technique achieved a level of comfort and presence comparable to that of instant edits. Furthermore, all alignment edits tested led to a noticeable reduction in head movement speed after the edit, affirming their potential utility for on-demand video streaming. Notably, the gradual edits, in particular, induced a significant reduction of approximately 8% in head movement speeds when compared to the instant alignment technique. These findings shed light on the positive effects of alignment edits on user experience and firmly establish the viability of the proposed gradual alignment technique to enhance QoE during video consumption.

INDEX TERMS 360° videos, alignment edits, quality of experience, user experience.

I. INTRODUCTION

Virtual reality (VR) has gained popularity due to its immersive and realistic experiences, with the VR market expected to grow from US$ 28.42 billion in 2022 to US$ 87 billion in 2030 [4]. This growth can be attributed to the affordability of head-mounted displays (HMDs), the expansion of metaverse

The associate editor coordinating the review of this manuscript and approving it for publication was Alessandro Floris.

solutions, and the increasing production of high-quality VR content, particularly 360° videos [5]. Although the VR industry has reached a certain level of maturity, the streaming of 360° videos is still in the early stages of development, mainly due to the challenges associated with the streaming of such content over today's typical broadband residential Internet connections. However, there is significant potential for expansion in this area [11]. To provide a high-quality viewing experience of 360° videos to end-users, two key

questions must be addressed. First, to what extent can the design of 360° video content be optimized to enhance the user's quality of experience (QoE)? Second, how can we improve the delivery of resource-demanding 360° videos over the Internet to increase the perceived QoE?

The first question can be addressed through cinematography studies that provide guidelines for content creators to enhance the impact of the story, user engagement, and user sense of presence [6], [7], [9], [67]. Although it is primarily based on empirical information, content design can be further examined through quantitative user studies that collect behavioral or subjective data [28], [37]. From a cinematographic point of view, it is important to develop techniques to guide viewers across scenes, so they can follow the intended storyline [12]. Regarding the second question, streaming videos with the available bandwidth requires adapting the video resolution during streaming time using adaptive bit rate (ABR) algorithms. To stream 360° videos, ABR algorithms are frequently used given the way these videos are visualized. The spatial and temporal content that must be transmitted depends on the user's gaze direction, which requires the implementation of complex functionalities, such as gaze prediction and content recognition [1], [2], [3].

This study focuses on exploring video edits as a content design mechanism to enhance the streaming experience of 360-degree videos. Specifically, we investigate a specialized category of video edits termed "alignment edits," which effectively redirect the user's field of view during video playback. Figure 1 visually demonstrates the impact of an alignment edit on the user's field of view. There are two fundamental types of alignment edits considered in this research: instant and gradual edits. Both types work by aligning the user's field of view (FOV) with a predetermined timestamp's significant region of interest (ROI). Employing content alignment has the potential to enhance gaze prediction, leading to more efficient utilization of network resources and potentially elevating user Quality of Experience (QoE) [28]. These alignment edits can be triggered either in real-time by the video player system or incorporated directly into the original video. The ultimate objective of this technique is to guide users by manipulating their view, thereby altering the content presented within their FOV.

Alignment edits were first investigated by Dambra et al. [28]. In their study, alignment edits were used to instantaneously rotate the 360° frame, aligning the user's FOV with the specific ROI inside the content. The authors proposed their technique based on research on attention coordination in mult-user VR narratives [12], with a focus on 360° streaming optimization. They concluded that those "instant alignment edits" improve streaming indicators by reducing bandwidth consumption and the user's average head motion, without a noticeable decay in the user's feeling of immersion.

Instant alignment edits may not always be the optimal choice, as they can potentially disorient users, as discussed in [12]. Despite the usefulness of alignment edits, there

has been a lack of in-depth investigation into the impact they have on the overall user experience. Previous research by Dambra et al. [28] focused on a limited number of video contents and solely explored instant alignment edits, disregarding other possible methods to redirect the user's field of view. Their argument was that gradual rotations in alignment edits could induce cybersickness. However, in this study, we challenge this assumption and consider the "blinking eye" cybersickness reduction technique [29] to propose a gradual alignment edit for 360° videos. Our investigation aims to assess the feasibility of these gradual alignment edits, which can potentially prevent user disorientation and negative effects on the sense of presence, as observed in other studies involving 2D video cuts or VR teleportation [56], [65], [66]. Figure 1 illustrates the alignment edits, demonstrating how horizontal rotation of the 360° frame can achieve alignment between the region of interest (ROI) and the user's field of view (FOV).

To our knowledge, this is the first work to explore the effects of alignment edits on several aspects of the user experience, such as head motion, sense of presence, discomfort, and overall experience. Figure 1a shows the two basic types of gradual and instant alignment edits included in our study, instant alignment is conducted as frame-to-frame rotation, while gradual alignment is conducted by 360° frame rotations inside a time interval. Moreover, our user study includes a variety of content and editing conditions, such as content motion, semantic information, and scene environment [9], [64]. In addition, our proposed alignment edit was investigated with various rotation speeds in a reproducible montage scheme, where the edits are located at the same video timestamp, enabling data-based parameter tuning.

We structure the paper as follows. In Section II, we provide background information on storytelling in 360-degree videos, showing the technical challenges related to alignment edits. Next, in Section III, we describe the proposed gradual alignment edit technique and formulate the parameters of the alignment edits. In Section IV, we define the research hypothesis, present the user study preparation, and detail the user study methodology. In Section V, we describe the analysis procedure and results, first considering the opinion scores given by participants in Section V-A, following the results considering the head tracking data in Section V-B. Finally, Section VI offers evidence-based recommendations, identifies future research directions, and discusses the limitations of our findings.

II. STORYTELLING AND EDITING FOR 360° VIDEOS

Cinematography guidelines serve as a critical tool in establishing rules to achieve scene continuity and aesthetic coherence in traditional filmmaking [67]. These guidelines include some rules, for example, the 180° rule, which restricts camera positioning across the action axis to achieve scene coherence. To maintain continuity of action, directors typically begin an action in one shot and immediately
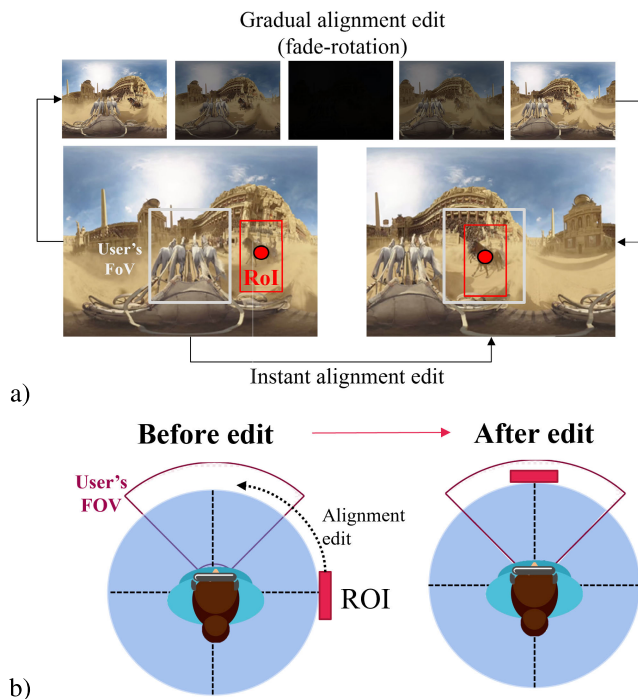
Gradual alignment edit
(fade-rotation)

a)

Before edit → After edit

b)

FIGURE 1. Illustration of the two types of alignment edits investigated: a) video frames prior to the alignment edits and right after it lining up the user FOV with a specific ROI; b) top-down perspective of the ROI motion across an alignment edit.

FIGURE 2. Main stages of the rendering process of a 360° video, with the user positioned in the center of the render sphere.

continue it after a cut. The 180° rule also creates a virtual stage in which the action unfolds [14]. But in immersive storytelling production, the role of the director changes as viewers have control of the camera and the freedom to explore the scene [68], [69]. This presents challenges in creating a coherent narrative in 360° videos due to spatial displacements between regions of interest [16], [17]. Although continuity editing still applies to Cinematic Virtual Reality (CVR) [18], many traditional editing techniques (e.g., camera angles, zooms, fades, cuts) may be ineffective in the 360° scenario, leading to questions on how to create effective narratives for this type of immersive media. Figure 2 depicts the main stages of the rendering process of a 360° video, showing that the displayed visual inside the user's FOV is a small portion of the complete render sphere.

When watching 360-degree videos, viewers typically exhibit two primary behaviors: *exploratory*, where they freely navigate the content in search of interesting spots, and *fixation*, where they focus on a particular region of interest (ROI). These behaviors affect gaze directions and the overall user experience [37], [38]. For example, if something captures the viewer's attention in a way that the filmmaker did not expect, viewers can miss notable events that help them understand the story and enjoy the content, thereby potentially degrading the user experience. In the last decade, several studies have been conducted to help directors improve the user experience for 360-degree videos [7], [19]. For example, Gödde et al. [21] performed a user study with
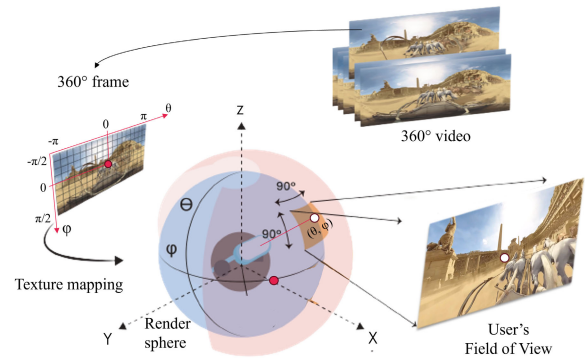
50 participants that investigated the viewer's tolerance for spatio-temporal story density, which refers to the quantity, position, and frequency of ROIs over a given timeline of the story. They found that in scenes with high spatial-temporal semantic density, a significant portion of the audience missed the story plot, with 80% of participants unable to correctly answer story-based questions such as "What happened to the main character?" or "Why did the character become aggressive?." Aitamurto et al. [40] examined variations in spatio-temporal viewing conditions in CVR and tested how they trigger a psychological condition known as fear of missing out (FOMO), resulting in anxiety and reducing viewer enjoyment.

Cinematography studies suggest that real-time editing and guiding techniques can prevent plot mismatches and promote storytelling engagement in CVR [7], [12]. At the same time, these methods reduce the probability of missing important or notable events; they also reduce user viewing freedom and the sensation of immersion [40]. There are two types of viewing guidance techniques: active and passive [7], [73]. Passive techniques can use either diegetic or non-diegetic attractors. Diegetic attractors are elements that are part of the story world, such as a character or object of the story, and are inserted at specific times to capture the viewer's attention [7]. Non-diegetic attractors, on the other hand, are elements outside the story world, such as visual guidance effects like arrows or radars, that are inserted in viewers' displays. Active techniques, on the other hand, support viewers in real-time, for example, by making the camera follow specific targets [13], [74] or manipulating the luminosity or saturation of the scene [70]. In this work, we focus on active techniques that can provide full predictability of the gaze direction, which is often required for streaming applications. It is worth noting that if the active technique is too intrusive, it can be irritating or disturbing to the viewers. For example, abrupt camera movements can degrade the sense of presence and cause discomfort. Therefore, it is crucial to consider the trade-off between discomfort (for example, cyber-sickness) and presence when designing any mechanism that alters the viewing experience [32], [33].

Researchers in the area of virtual reality and cinematography have proposed several active techniques for 360-degree videos. For example, Brilhart [20] proposed a novel editing principle for CVR, which generates scene edits by estimating content areas that are more salient or perceptually important to the storyline. Other studies, such as the work of Gödde et al. [21], have focused on creating a narrative for virtual reality content. In addition, Fearghail et al. [19], [62] investigated how predicted visual attention can help directors perform automatic content analysis, determining where the user's gaze should be directed. Finally, Dambra et al. [28] proposed a technique that allows the VR content creator to estimate user attention and determine the ROI of the scene.

From current studies, it seems clear that manipulating the field-of-view (FOV) orientation can significantly impact the user quality of experience (QoE). Pavel et al. [27] recognized this and explored different shot-orientation techniques to help viewers visualize critical information in 360-degree video stories. Although it is important to consider cybersickness associated with FOV orientation, Farmani and Teather [29] proposed a technique that addresses cybersickness caused by visual-vestibular conflicts in stationary VR. Their technique reduces the illumination of the screen when rapid head movements occur, simulating the blinking of the eyes to avoid cybersickness. This bioinspired solution was able to reduce cybersickness by up to 40% in a first-person VR shooting game application. In general, these studies demonstrate the importance of FOV orientation and provide valuable information to improve the QoE of VR users.

More recently, some studies have also explored the impact of video editing and guidance on viewing behavior. For example, Cao et al. [22] investigated the effects of three transition effects (portal, fade, cut) and found no conclusive reduction in story recall. Serrano et al. [18] analyzed head and eye tracking data to examine the effects of content factors and concluded that the number of ROIs and their displacements play a critical role in user behavior. Specifically, the time to find an ROI and the stabilization of gaze within an ROI are related to these factors. Marañes et al. [37] studied the impact of the number of ROIs before and after cuts and proposed area-based behavior metrics for head tracking data. Kjaer et al. [23] studied the effect of cut frequency on viewer disorientation and found evidence that editing does not pose a problem in relation to cinematic VR.

It is clear that real-time edits offer opportunities to customize the CVR for people susceptible to cybersickness [41], [42] or those prone to diverge from the predesigned storyline, such as individuals with a low reaction time [43]. Furthermore, real-time edits also enable system optimizations, expanding the effectiveness of streaming applications. For example, Dambra et al. [28] and Sassateli et al. [44] have examined how the instantaneous alignment technique can reduce exploratory behavior, improving streaming by reducing average bandwidth consumption. Proper alignment of ROIs is essential in 360-degree videos, since it directly affects the immersion level experienced by viewers. The

correct alignment of ROIs allows for story deployment without limiting viewer freedom, while improper alignment can cause viewers to miss critical events, degrading the overall experience [40]. To make informed alignment decisions, it is crucial to understand two main viewer behaviors: exploration and fixation. During exploration, the viewer gaze searches for ROIs in the content, while during fixation, the viewer focuses on a specific ROI [37], [38].

As previously mentioned, achieving smooth 360-degree camera movement without compromising the user experience (UX) presents a challenge, as it may disrupt the observer's ear-vision system and induce discomfort. Dambra et al. [28] conducted an investigation into the user's tolerance for the instant alignment edit and found it to be generally well received in terms of comfort. However, the use of instant rotations (or offsets) in these alignment edits controls user motion without providing any indication of the camera's direction, potentially leading to users losing their orientation and thus negatively impacting their sense of presence. In particular, there is a lack of empirical testing that directly addresses this assumption. Gradual rotation can offer an advantage over instant rotation in terms of immersion, despite possible comfort implications. This assumption is grounded in the idea that incorporating a scene transition with scene motion can maintain the sense of presence. Recent research also suggests that gradual rotations can be as comfortable as instant alignments [29], [35], [36].

III. PROPOSED ALIGNMENT EDIT DESIGN

Alignment edits involve video transitions that incorporate a frame rotation to align the user's view with a specific region of interest (ROI). Our goal is to create a smooth and uninterrupted edit that reduces cybersickness, inspired by the natural human action of blinking when viewing visual material. To accomplish this, we introduce a novel alignment edit called fade-rotation, which combines horizontal rotation of the 360° frame with fade-in/fade-out effects. The fade-rotation transition represents a type of gradual alignment edit, with the rotation occurring over a time interval. Figure 1a shows an illustration of this gradual rotation in a sequence of frames, where the distance between the user's view and the ROI is altered. Note that the ROI gradually moves into the user's field of view (FOV), while simultaneously a blink-of-the-eye effect is applied.

Figure 3 shows the reference coordinate system used in this work, which has an origin in the center of the render sphere. In this system, the HMD is positioned at the origin, and the 360° frame is projected in the shell of the render sphere with a fixed radius (R). We anchor the center point of the 360° frame at $x = R$, $y = 0$, $z = 0$. With this anchoring, we have a single coordinate system to describe the content positions, the head directions, and the camera FOV consistently. To map back-and-forth the positions from the render sphere and the 360° frame, we use the equirectangular projection. In this projection, the azimuthal angle (θ) varies within the interval $[-\pi, \pi]$
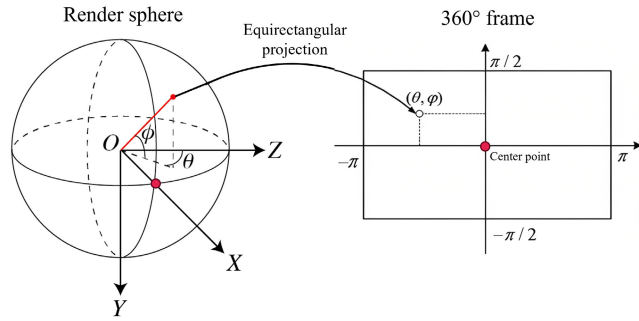
FIGURE 3. Reference coordinate system, defined in terms of the render sphere, the red dot represents the origin of the equirectangular frame.

FIGURE 4. Two fade-rotations included in a video timeline, representing the temporal edit structure of a video with multiple alignment edits.

(in radians) and corresponds to horizontal (side-to-side) head movements around the y-axis. The polar angle (ϕ) varies within the interval $[-\pi/2, \pi/2]$ (in radians) corresponding to vertical (up-down) head movements. The center of the 360° frame is fixed at $\theta = 0$, $\phi = 0$, and corresponds directly to a fixed position at the reference coordinate system $x = R$, $y = 0$, $z = 0$; establishing a fixed reference.

The alignment edit method has three parameters: the total duration of the rotation edit (ΔT_{edit} in seconds), the duration of the fade-in/fade-out effect (ΔT_{fade} in seconds), and the angular speed of the 360°-frame rotation (ω in degrees/s). These parameters can be adjusted and combined to obtain the desired transition behavior. Some examples of alignment edits include:

1) instant alignment (snap-change): $\Delta T_{edit} = 0$, $\Delta T_{fade} = 0$, and high values of $|\omega|$;
2) gradual-rotation alignment: $\Delta T_{edit} > 0$, $\Delta T_{fade} = 0$, and $|\omega| > 0$;
3) fade-rotation alignment: $\Delta T_{edit} \geq \Delta T_{fade} > 0$ and $|\omega| > 0$.

In this study, our investigation focuses on cases (1) and (3). We deliberately excluded case (2) because studies have shown that it has a negative impact on user comfort [29]. We analyze the variation of the rotation speed, without examining the parameters ΔT_{fade} and ΔT_{fade}. By concentrating on these specific aspects, we aim to provide a more comprehensive understanding of the feasibility of using fade-rotation edits in 360° videos.

The fade-rotation edit should be implemented at pre-selected timestamps of the video ($t_1, \ldots, t_N$). Figure 4 illustrates the video structure, which contains alignment edits between video shots. Applying the edit can be a player's decision to enable real-time streaming optimization. Furthermore, optimization models can automatically determine whether to trigger alignment edits, while still respecting the cinematographic choices of content creators.

Figure 5 provides an illustration of the snap-change (SC) and fade-rotation (FR) edit techniques. When editing the source videos, we considered the following visual equivalence rule. If two viewers were looking at the same frame location at the start of an edit, they should end up at the
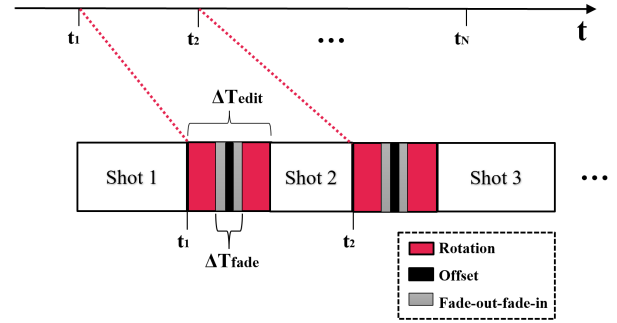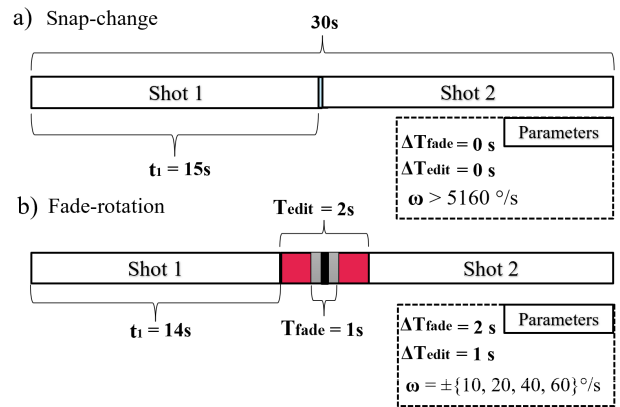

FIGURE 5. Applied parameters to the video stimuli of the user study: a) instant alignment (snap-change) settings; b) gradual alignment (fade-rotation) settings.

same frame location regardless of the type of edit executed. We define θ_T as the target total angular displacement between an initial ROI and a target ROI after an edit.

In the current study, we tested only offline FR, which means that the edits were applied to the source videos before they were watched. Online FR, which is applied in video playback time, is an object of future work. For the offline case, we cannot control where users are looking at the video when the edit is executed. Because of this, we selected the initial ROI as the center point of the frame. To increase the chance that users watched the initial ROI, we selected videos in such a way that there was one meaningful object in the center of the frame at the time of editing. Section IV describes the preparation of the videos in detail.

The gradual alignment edit tested in this work has several rotation speeds, which affects the angular displacement that can be achieved in a given time interval. For a video, the angular displacement achieved with the FR edit method is given by $\theta_r = \omega \cdot \Delta T_{edit}$, which may be higher or lower than the target total angular displacement θ_T. This requires that a small offset rotation be applied to the video, which is done at the exact moment the frame is completely black. The value of the offset rotation is simply the difference between θ_T and θ_r.
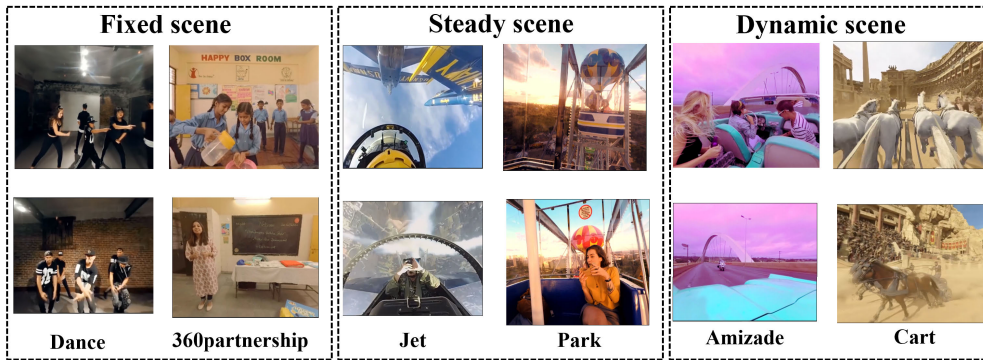
FIGURE 6. Video-stimuli of the subjective experiment, organized by camera motion type. Top: the user FOV at the center point (initial head position). Bottom: the pre-defined target ROI.

IV. USER STUDY

The primary objective of this research is to investigate the effects of offline alignment edits on user QoE and behavior, based on previous studies [28], [29], [33]. Our specific focus lies in understanding the acceptability of the fade-rotation technique among users and determining the optimal rotation speed that enhances the sensation of 'presence' while minimizing feelings of 'discomfort.' Additionally, our objective is to explore the comparative impact of instantaneous and gradual alignment edits on the overall quality of user experience.

To carry out this study, we formulate the following hypotheses:

H1 : The degree of comfort of fade-rotation is equivalent to that of snap-change;

H2 : The snap-change has a higher negative effect in presence than fade-rotation;

H3 : The ROI alignment impacts presence, comfort, and experience scores;

H4 : Alignment edits reduce the viewer's head movement speed after the edit.

To test these hypotheses, we performed a subjective experiment. Next, we detail the video stimuli and the experimental methodology used in the study.

A. VIDEO STIMULI

To test alignment edits under a variety of conditions, we chose videos that have three types of camera motions: static, steady, and dynamic. Static refers to videos that were shot with a fixed camera, steady refers to videos where the camera is in motion for most of the scene (independent of direction), and dynamic refers to videos that contain camera acceleration and content motion [35], [61]. Figure 6 shows snapshots of the six videos selected for the experiment, where four videos were chosen from the datasets Directors Cut [62] and UTD [63] ("360partnership," "Jet," "Dance," and "Cart") and the other two videos were provided by filmmakers from Caixote XR studio[1] ("Amizade" and "Park").

The chosen set of video stimuli covers a wide range of spatial and temporal information, including outdoor and indoor content. Figure 7 shows the spatial perceptual information (SI) versus the temporal perceptual information (TI) for each video [75]. These metrics indicate the amount of spatial and temporal changes in a video sequence. No criteria were used for the number of ROIs in the scene, but we looked for scenes that had relevant moving objects that could capture the viewer's attention until the intervention occurred. To select the target ROI for the alignment edit, we watched the original videos with an HMD and decided which parts of the content were perceptually important. To prevent temporal and content bias, we avoided ROIs at the beginning of the video and made sure that transitions occurred within the same duration and started at the same video timestamp. The audio track was removed from the videos, which were encoded with the H.264 codec at 40 kbps (target quality), 60 frames per second (fps), using equirectangular projection at 3840 × 1920 resolution.

The edits were manually implemented and added to the source videos using Adobe Premiere software[2] and the VR projection plugin.[3] The latter is specific for editing 360° videos. To ensure fairness and avoid potential bias, all videos had a duration of 30 seconds. The alignment edit parameters used in the user study (see Figure 5) were defined as follows.

- Snap-Change (SC): $t_1 = 15\ s$, $\Delta T_{edit} = 0s$, $\Delta T_{fade} = 0s$, and the angular speed (performed) being equal to $\theta_T \cdot 60°/s$.[4]
- Fade-Rotation (FR): $t_1 = 14\ s$, $\Delta T_{edit} = 2\ s$, $\Delta T_{fade} = 200ms$, and $\omega = 10°/s, 20°/s, 40°/s, 60°/s$.

We assume that at t_1 the viewer's FOV is centered at the frame center ($\theta = 0°$, $\phi = 0°$). When the viewer starts to watch the video, generally he/she looks towards the center of the frame, regardless of the content [18]. To increase

---

[1] https://caixotexr.com/

[2] https://www.adobe.com/br/products/premiere

[3] https://helpx.adobe.com/premiere-pro/how-to/vr-projection

[4] To compute ω, consider that the rotation is performed in the interval between two frames. Since the video has 60 frames per second, a single frame occurs in $1/60 = 16.7$ms. For the target angular displacement θ_T, the angular speed is $\theta_T \cdot 60°/s$.
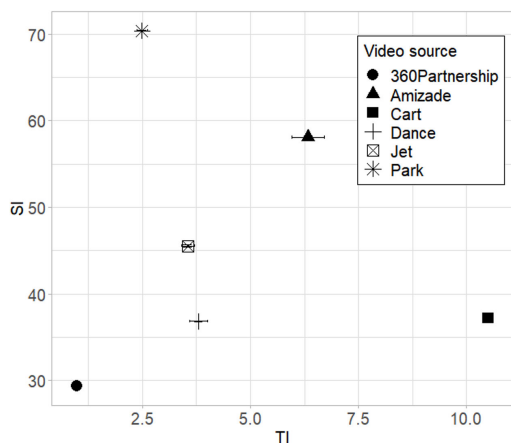
FIGURE 7. Spatial and temporal activity indexes of videos from the user study.

the probability of having the viewers looking at the desired frame center at t_1, all videos had an ROI at the frame center (center point) at t_1. To achieve that, we adjusted the initial frame center of two videos by applying an initial offset, namely Park ($-180°$) and Dance ($-86°$). The other four videos already had an ROI in the frame center point at t_1. To allow accommodation time, we set a time interval between 14 and 16 seconds before the edit [18]. In the edit software, we applied the transition called "to black", centered at $t = 15s$. We selected a 30-second long clip from the original videos. The total alignment angle θ_T for each video is: Park = $180°$, Jet = $180°$, 360Partnership = $170°$, Dance = $86°$, Amizade = $-120°$, and Cart = $120°$.

B. EXPERIMENTAL METHODOLOGY
We use the experimental methodology described in ITU-T Recommendation P.919 [50]. A full run of the experiment took approximately 37-40 minutes. The experiment was spread over two periods of time, with two different HMDs in each period. At first, participants used Oculus Rift S, while in the second participants used Meta Quest 2. During the test, participants were seated in a swivel chair. Participants who wore glasses or lenses kept them on throughout the session. As shown in Figure 8, the experiment had eight phases: (1) instructions, (2) training, (3) first session, (4) first cybersickness questionnaire (SSQ), (5) rest, (6) second session, (7) second SSQ, and (8) finalization.

In the instructions phase, participants had to select the language of the experiment (Portuguese or English), sign a consent form, read safety protocols, and complete a screening pre-questionnaire and a consent form. Figure 9 shows screenshots of the instructions phase. The prequestionnaire contained demographic and visual aptitude questions and was based on the recommendations of the VQEG immersive media group.[5] The consent form can be found in Appendix A. Following the instructions, the participants participated

[5]https://vqeg.org/projects/immersive-media-group

in a training session, where they watched a training video and simulated rating the videos by interacting with the interface. The participants were given the opportunity to repeat the training until they felt confident to proceed to the experimental session.

In the first and second sessions, participants watched the 36 videos in a randomized order, giving attribute scores to each watched video. More specifically, participants watched 16 videos, completed the cybersickness questionnaire, removed the HMD, took a 5-minute break to avoid excessive cognitive load [50], and watched 20 videos. In the end, participants completed the post-questionnaire with additional questions about the experiment, such as personal insights and comments about the experiment. The implementation of the questionnaires was fully automated and no intervention from the experimenter was required.

After viewing each video, participants were asked to rate attributes of the content using a controller interface embedded into the video player, as shown in Figure 10. The experiment was within-subjects, which means that all participants evaluated all test conditions. We used the Absolute Category Rating with Hidden Reference (ACR-HR), which requires participants to score the processed video sequences (PVS) and the corresponding source video sequences (SRC) using a discrete degradation scale ranging from 1 to 5 [52], [53]. As the name suggests, in the ACR-HR methodology, the reference video is not identified. The participants rated three attributes of each video: overall experience, discomfort, and presence. Table 1 presents the questions and the specific scoring scales used for each of the three attributes [50], [54].

The videos were presented in random order [75] to prevent or minimize temporal bias, memory-related impacts, among other issues. However, based on Farmani and Teather [29], who proposed a method to alleviate induced cybersickness during subjective experiments involving rotations, we refined the randomization process excluding videos with rapid fade-rotations (angular speed exceeding 40°/s) from the initial set of 8 videos.

Table 2 presents a summary of the characteristics of the pool of participants for each experiment. We recruited 40 and 23 participants for the first and second experiments, respectively. The sampled population had a wide variety of ages and HMD experience, and the proportion of women was greater than 40% in both experiments. In total, we collected 6,804 opinion scores and 1,300-2,000 head tracking samples per video watched. We prioritized recruiting participants outside of the university to improve population sampling.

The procedure of the experiment was implemented using a specially developed software platform. Since all the solutions available for subjective evaluation of 360° videos require proprietary platforms, we developed a platform to perform the experimental procedures: the monoscopic 360° video subjective assessment tool (Mono360). Mono360 is a Web application designed specifically to conduct this user study. It is based on a client-server architecture and uses only
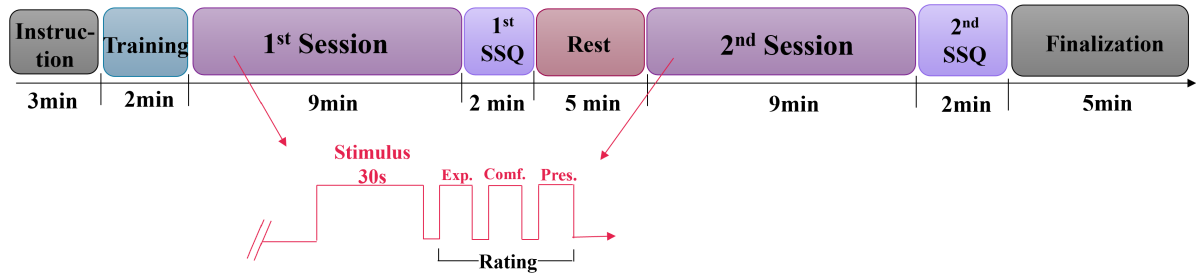
FIGURE 8. Procedure of the experiment, and the subject rating time structure.

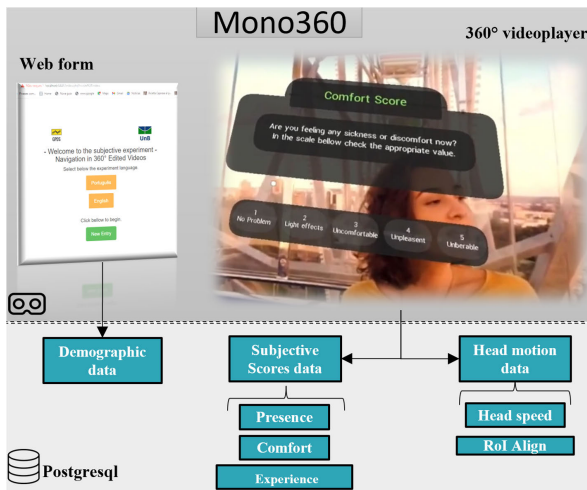**FIGURE 9. Instruction phase views.**

FIGURE 10. Tools for capturing and saving experimental data. Subjective rating scores are captured from an embedded user interface without removing the HMD.

TABLE 1. Subjective assessment measures.

Questions	Scale	Attribute
Are you feeling any sickness or discomfort now? In the scale below, check the appropriate value.	(1) Unbearable (2) Unpleasant (3) Uncomfortable (4) Light effects (5) No Problem	Comfort [52]
To which extend do you feel present in the virtual environment, as if you were really there? In the scale below, check the appropriate value.	(1) Nothing (2) Little much (3) Reasonably (4) Very much (5) Entirely	Presence [53]
Evaluate the overall experience when watching the video In scale the below, check the appropriate value.	(1) Bad (2) Poor (3) Fair (4) Good (5) Excellent	Experience [50]
Evaluate the following symptoms: Nausea, Vertigo, Sweating, Stomach awareness, Increase in salivation, Difficulty in concentration.	For each symptom: (1) None (2) Slight (3) Moderate (4) Severe	Cyber-sickness [57]

TABLE 2. Experiment population summary for both devices.

Device	Num. of Particip.	Age Avg.	Age Min.	Age Max.	Prop. of Women	1st time VR users
Rift S	40	35.62	15	65	60.0%	55.0%
Quest 2	23	29.56	18	41	43.0%	60.0%
Total	63	33.40	15	65	53.8%	56.8%

open-source technologies. The back-end of the application executes the PHP-based Yii2 framework[6] on the server side, while the front-end interface uses Bootstrap[7] framework, controlling all the user flow inside the app. The database was implemented with Postgresql, while the video player that runs on the HMD's browser uses WebXR[8] API.

For rendering 360° videos, we used the Three.js[9] library, with a WEBGL2 renderer. The rendering procedure consists of decoding the video texture into two spheres corresponding

---

[6] https://www.yiiframework.com/
[7] https://getbootstrap.com/
[8] https://immersiveweb.dev/
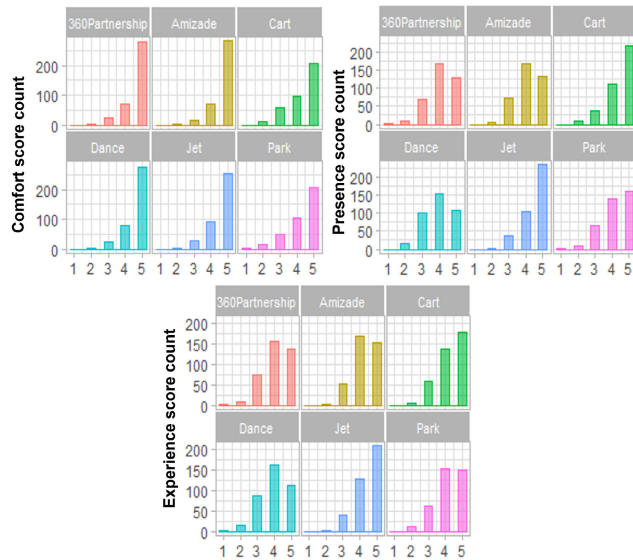
[9] https://threejs.org

FIGURE 11. Scores for the QoE attributes (presence, comfort, and experience) measured in the user study. The scores are grouped by video content. In our user study, each participant rated each video six times. Best viewed in color.

to both eye screens. For that, we implemented the render sphere with the "SphereGeometry" class from Three.js package with radius = 500, widthSegments = 60, heightSegments = 40. The video decoding process is managed by the browser. We adjusted the procedures across the two different devices.

V. RESULTS

The experiment contained six videos and six types of edits. The edits are the following:

- Fade Rotation (FR) with 4 rotation speeds: 10°/s, 20°/s, 40°/s, and 60°/s, referred as FR10, FR20, FR40, and FR60;
- Snap-Change, referred to as SC;
- No Edits, referred to as NONE.

Therefore, each participant assessed a total of 36 videos. We gathered a total of 6,804 scores, collecting scores from 63 participants for the attributes *experience*, *discomfort*, and *presence*.

We first examine the distribution[10] of the subjective scores for the three attributes. Figure 11 shows histograms containing the distribution of the presence, experience, and comfort scores grouped by video content. Applying a Shapiro test, we confirmed that the distribution is non-normal ($P < 0.05$), indicating that inference should use non-parametric tests, this signifies the need for employing non-parametric tests in our subsequent analyses.

For each studied attribute q (comfort, presence, and experience, $1 \leq q \leq 3$) and each j-th video sequence, we compute the mean opinion score (MOS) for the pool of

[10]The code for the used is available at https://gitlab.com/gpds-unb/fade-rotation-exp

N participants:

$$\text{MOS}_{j,q} = \frac{1}{N} \sum_{i=1}^{N} x_{i,j,q}, \qquad (1)$$

and the standard deviation of the opinion score (SOS)

$$SOS_{j,q} = \sqrt{\frac{\sum_{i=1}^{N} (x_{i,j,q} - MOS_{j,q})^2}{(N-1)}}, \qquad (2)$$

where $x_{i,j,q}$ represents the score given by the i-th participant to the q-th attribute corresponding to the j-th video sequence. For the tests, we consider a margin of error of 0.05 and a confidence interval of 95%, computed as follows:

$$CI_{j,q} = MOS_{j,q} \pm \frac{1,96 \cdot SOS_{j,q}}{\sqrt{N}}. \qquad (3)$$

Next, we performed the Welch t-test to identify if the attribute scores of the data acquired in October and November (with different devices, the Rift and Quest 2) are statistically different. For that, we perform a pairwise comparison between the two sub-experiment groups. The Welch t-test is adopted because the samples are not balanced and the subsets are of different sizes. The test shows that for the 'presence' scores, there are no significant differences ($P > 0.05$) between the Rift or Quest 2 groups. For the 'comfort' scores, when a pairwise comparison grouped by edit type was performed, a significant difference was found for the FR20 group. However, no significant differences were observed for all other cases.

Figure 12 shows the MOS values grouped by video for each measured attribute. Notice that 'comfort' achieved scores greater than 4 for all content, indicating that users felt high levels of comfort for the different types of content motion, and for the several alignment edits. The highest 'comfort' scores were for "Amizade," followed by "Dance,", while "Cart" had the lowest 'comfort' scores because it has a strong camera acceleration. In terms of the attribute 'presence', only "Dance" had scores less than 4, while the best scores were for "Jet," which is the only video with 'presence' higher than 'comfort'. In terms of 'experience', the scores followed the same tendency of 'presence' scores, where the highest score was for "Jet," and the lowest score for "Dance". "Dance" and "Amizade" were the only videos in which the 'experience' scores were higher than 'presence' scores.

We observed a relevant pattern in the data: videos characterized by minimal camera movements ("Dance," "360Partnership," and "Amizade") exhibited a substantial discrepancy between the 'comfort' and 'presence' scores (refer to Figure 12). In contrast, videos featuring more pronounced camera acceleration ("Jet," "Cart," and "Park") displayed a 'comfort' and 'presence' difference of less than 0.12. Notably, among these, the video "Cart" stood out, being the sole instance where the 'presence' value surpassed 'comfort'. This phenomenon indicates that videos with heightened camera motion tend to yield lower 'comfort' scores but higher 'presence' scores. This observation underscores the
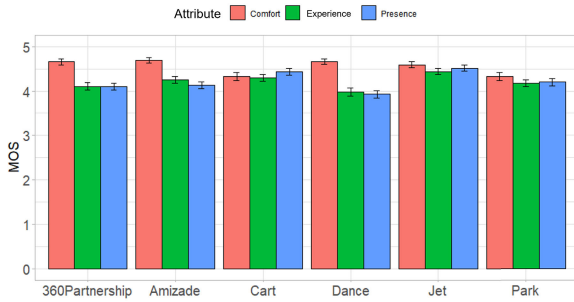
FIGURE 12. Mean opinion scores for 'presence', 'comfort', and 'experience' for each video sequence.

significance of considering scene motion when incorporating alignment edits into the video.

From the feedback provided by the participants, other aspects of the content decrease the perceived 'presence'. For example, in "Dance," ten participants reported that this video lacks realism because the dancers in the video seemed out of scale, causing strangeness. This is observed in the data that show low average scores for 'presence'. Another feedback provided by the participants is that the video content that resembled conventional 2D videos reduced their sense of presence. This was true for the videos "Dance", "360Partnership," and "Amizade", as mentioned by participants. For example, in "Amizade," some participants reported feeling outside of the car, while others reported that the content of "360Partnership" appeared artificial because they felt smaller. These situations illustrate how content can break the feeling of 'being there' (presence), corroborating recent studies on realism in VR [56].

Before performing the hypothesis analysis, we checked the reliability of the collected scores. For this, we computed the correlation between the various attribute scores (presence, comfort, and experience). As suggested by the ITU guidelines [50], the correlation between attribute pairs can be computed with the conventional Pearson linear correlation coefficient (PLCC):

$$PLCC(x, y) = \frac{\sum_{i=1}^{N}(x(i) - \bar{x})(y(i) - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x(i) - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y(i) - \bar{y})^2}}, \quad (4)$$

and the Spearman rank correlation coefficient (SRCC):

$$SRCC(x, y) = 1 - \frac{6\sum_{i=1}^{N}(R_x(i) - R_y(i))^2}{N(N^2 - 1)}, \quad (5)$$

where x and y are vectors of length N that represent the two variables being compared, $\bar{x}$ and $\bar{y}$ are the mean values of x and y, respectively, and $R_x(i)$ is the rank of the i-th value of x, and $R_y(i)$ represents the rank of the i-th value of y. To interpret the correlation values, we follow the convention of Schober et al. [58], where values below 0.1 are considered negligible, values between 0.1 and 0.69 are considered moderate, values between 0.7 and 0.89 are considered strong, and values over 0.9 are considered very strong.

TABLE 3. Correlation between QoE attributes, with data aggregated by Edit type. In bold we highlight the moderate or strong correlations.

Edit type	Comparison	PLCC	p-val	SRCC	p-val
Gradual FR10	comfort/presence	0.021	0.681	0.021	0.688
	comfort/experience	**0.136**	0.01	**0.143**	0.01
	presence/experience	**0.713**	0.001	**0.694**	0.001
Gradual FR20	comfort/presence	0.191	0.001	0.142	0.001
	comfort/experience	**0.339**	0.001	**0.319**	0.001
	presence/experience	**0.720**	0.001	**0.709**	0.001
Gradual FR40	comfort/presence	0.205	0.059	0.175	0.001
	comfort/experience	**0.396**	0.001	**0.360**	0.001
	presence/experience	**0.732**	0.001	**0.736**	0.001
Gradual FR60	comfort/presence	0.173	0,001	0.132	0.01
	comfort/experience	**0.363**	0.001	**0.328**	0.001
	presence/experience	**0.716**	0.001	**0.710**	0.001
SC	comfort/presence	0.045	0.385	0.028	0.591
	comfort/experience	0.150	0.01	0.157	0.01
	presence/experience	**0.674**	0.001	**0.658**	0.001
None	comfort/presence	0.001	0.988	0.007	0.887
	comfort/experience	0.236	0,01	0.225	0,01
	presence/experience	**0.672**	0,001	**0.675**	0,001

Table 3 shows the pairwise correlation comparison of attribute scores under the same edit conditions. A negligible correlation was found between 'presence' and 'comfort' scores for three edit types (FR10, snap-change, and NONE) and a weak correlation ($CC < .2$) for FR20, FR40, FR60. This shows that the participants were able to distinguish 'presence' from 'comfort'. A weak correlation was found between 'comfort' and 'experience', and a strong correlation between 'presence' and 'experience' for all edit types. This result appears to be due to ambiguities in the definition of the overall experience for immersive experiences [50].

Figure 13 shows MOS values for different edit types grouped by video content. We notice that the 'comfort' MOS is higher than 4 for all cases, while the 'comfort' MOS for dynamic motions tends to be lower than for the other scene motions. We used the Kruskal-Wallis (KW)[11] test to determine whether there are significant differences between two or more independent groups, verifying the effect of video-content on 'comfort' and 'presence'. We found a statistically significant effect of content on 'presence' ($\chi^2 = 20.376, df = 4, p < 0.001$) and 'comfort' ($\chi^2 = 31.423, df = 4, p < 0.001$).

Figure 14 illustrates the MOS for different edit types grouped by scene motion. Specifically, the 'comfort' MOS exhibits a discernible decline, correlating with the rotational speed of gradual edits (FR). To analyze this trend, we grouped the scores by edit types and performed a KW test to examine the relationship between comfort scores and rotation speed values. The results show a significant impact of the rotation speed on 'comfort' ($\chi^2 = 12.511, df = 3, p < 0.01$). In contrast, the effect on presence was found to be

---

[11]All statistical the statistical analysis was conducted using built-in packages from R - https://www.R-project.org/
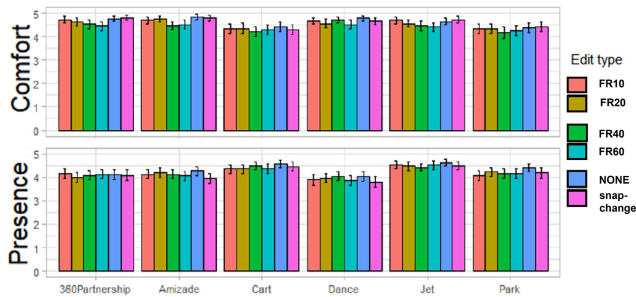
FIGURE 13. 'Presence' and 'comfort' MOS bar plots, grouped by edit type and video-content.
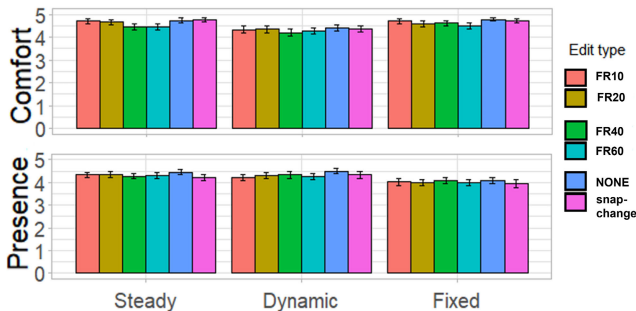
FIGURE 14. 'Presence' and 'comfort' MOS bar plots, grouped by edit type and scene motion.

TABLE 4. Paired Kruskall-Wallis test with FDR adjusted p-values for presence and comfort scores.

Camera Dynamic	Video	Comparison Video	p-value Comf.	p-value Pres.
Fixed	Dance	360Partnership	1	0.748
		Amizade	1	0.711
		Jet	1	$< \mathbf{0.001}$
	360Partnership	Amizade	0.184	0.914
		Jet	0.817	$< \mathbf{0.001}$
Steady	Amizade	Jet	0.108	$< \mathbf{0.001}$
		Cart	$< \mathbf{0.001}$	$< \mathbf{0.001}$
		Park	$< \mathbf{0.001}$	0.712
	Jet	Cart	$< \mathbf{0.05}$	0.968
		Park	$< \mathbf{0.01}$	$< \mathbf{0.01}$
Dynamic	Cart	Park	0.667	$< \mathbf{0.01}$
		Dance	$< \mathbf{0.01}$	$< \mathbf{0.001}$
		360Partnership	$< \mathbf{0.01}$	$< \mathbf{0.001}$
	Park	Dance	$< \mathbf{0.001}$	0.377
		360Partnership	$< \mathbf{0.01}$	0.677

non-significant ($\chi^2 = 0.236, df = 3, p > 0.05$), implying that the type of edit does not impact 'presence'.

Finally, we performed a multiple pairwise comparison for all groups, the post-hoc test for Kruskal-Wallis, to analyze whether the attribute scores given to a pair of videos are statistically different. The results of this test are shown in Table 4. Note that there is no significant statistical difference in 'comfort' for video-pairs with the same motion type, such as Dance/360Partnership, Amizade/Jet, and Cart/Park. This suggests that the camera's dynamic categorization (static, steady, dynamic) accurately classified the content, at least in terms of its impact on comfort. In terms of 'presence', the videos with no significant difference are 360Partnership/Amizade/Park/Dance, and separately Jet/Cart. In terms of genre, "Jet" and "Cart" are action videos.

A. OPINION SCORE ANALYSIS

To test if "the degree of comfort of fade-rotation is equivalent to that of snap-change" (hypothesis **H1**), we use the comfort scores shown in Figure 11, grouping them by edit type. We test the statistical difference between two sets of comfort scores, comparing snap-change with each fade-rotation type. For this analysis, we used Welch's t-test with FDR correction. We found a significant difference ($p < 0.05$) between snap-change and FR40, and between snap-change and FR60.

Complementing these results with a pairwise test, we also consider grouping the data in terms of the scene motion category. First, for dynamic scene motion, all comparisons had significant differences. Second, for the steady scene motion, a significant difference was found in the instant-FR40 and instant-FR60 pairs. Third, for fixed scene motion,

a significant difference was found in the instant-FR20, instant-FR40, and instant-FR60 pairs.

From all the comparison results presented in the last two paragraphs, **H1** is accepted for FR10 in fixed-scene motion videos and for FR10 and FR20 in steady motion videos. However, we reject **H1** for any video content with dynamic scene movement, and for FR with angular speed greater than 40°/s. In practical terms, for video players that lack the ability to account for scene motion in playback time, we recommend steering clear of FR20, FR40, and FR60, as they carry a higher likelihood of causing viewer discomfort. Instead, opting for FR10, or the snap-change approach, is preferable, as they exhibit a lower probability of discomfort-inducing effects. For videos characterized by steady camera motion, we suggest employing fade-rotation edits with an angular speed of less than 20°/s, as this can enhance the viewer's experience while minimizing the risk of discomfort. In essence, these findings underline the importance of selecting an appropriate FR strategy, taking into account camera motion, to optimize the viewer's experience and comfort.

Next, we investigate the fade-rotation scores relative to two baselines: the original version of the videos and the snap-change version. Figure 15 shows the scores for the four types of fade-rotation for each video-content, with the baselines shown as straight lines. This graph provides a visual comparison of multiple conditions. In terms of comfort, we see that FR10 had no significant difference ($p > 0.01$) compared to snap-change, for any video content. Furthermore, no significant differences were observed between fade rotation and snap-change for "Cart" and "Park".

Snap-change had the worst comfort scores for videos "Cart" and "Park". As the 'Cart' scene takes place, the viewer becomes a participant in a chariot race, while echoes of cheers reverberate as the race unfolds inside a coliseum. In the case of "Cart," instant edit was uncomfortable because it was combined with a strong camera translation when the chariot was turning. In the case of "Park," the viewer shares

a Ferris wheel cabin with a young woman. As the cabin ascends, the edit takes place. Comfort tends to decrease with the rotation speed for all video-content; however, specific conditions can break this trend. For example, for the video "Dance" we expected a decrease in comfort. But, surprisingly, there is a peak for the FR40 edit, showing that there is a non-trivial relationship between the rotation speed of the FR and the content; other similar cases happened for "Cart," "Park," and "Amizade" in FR60.

In terms of 'presence' for the instant edit, we notice a relatively low average score for "Dance'', "360Partnership,'' and "Amizade." These video-content had the lowest scene motion. Feedback from the participants indicated that when watching the "Dance" video, the instant alignment edit interrupted the change between the dancing groups, which caused the loss of the sense of presence. "Dance" and "Amizade" have fixed cameras and indoor scenes. It is not clear what attributes lead to a higher sense of presence; however, from the presence MOS values, we observe that videos "Cart," "Jet," and "Park" engaged them. It seems that interactions of the characters are not enough to promote a high sense of presence, given that for "Dance," and "360Partnership" there were people interacting with the camera and performing actions. However, they had a fixed camera and resulted in the lowest 'presence' scores.

To test hypothesis **H2**, we group the presence scores by edit type and apply the Welch t-test for all pairs. We did not find statistically significant differences ($p < 0.05$). Therefore, we rejected the hypothesis **H2**, confirming that the snap-change and fade rotation did not have a distinguishable effect on presence.

As mentioned above, the cybersickness questionnaire consisted of four possible levels of symptoms. Participants filled out the questionnaire after the first (pre) and second (pos) sessions. Figure 16 shows the frequency of these 4 levels of symptoms for these two instants. Note that the responses are similar results for the pre and post-questionnaires, with more than 90% of the participants reporting none to slight discomfort. Only one participant reported severe symptoms caused by the "Jet" video. This participant mentioned that he/she had a height phobia. These individual conditions are known to cause differences in comfort and a tendency to trigger cybersickness in VR [24].

B. HEAD MOTION ANALYSIS

The head motion analysis is performed using the head tracking data and two distance metrics. The two distance metrics are *i*) the distance between the gaze position and a target in the video content, and *ii*) the distance between two head tracking samples. As discussed in Section IV, gaze positions are recorded using normalized screen coordinates (X, Y), with the origin in the upper left corner of the 360° frame, spanning the interval $X, Y \in [0, 1]$. To convert the stored gaze position to the reference coordinate system (see Figure 3), we convert the normalized screen coordinates to Eulerian coordinates (ϕ, θ) by rescaling them to the

appropriate intervals: $\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ and $\theta \in [-\pi, \pi]$. With the rescaling procedure, the center point of the 360° frame matches the reference coordinate system, as shown in Figure 3.

The collected head-tracking data consists of the intersection points between the HMD's gaze direction and the spherical shell defined by the render sphere. To compute the spherical distance, we use the orthodromic distance metric, which is given by:

$$d(u, u') = 2R \cdot \arcsin\left[\frac{c(u, u')}{2R}\right], \quad (6)$$

where $c(u, u')$ is the Euclidean distance between two points on the spherical surface u, u', given by:

$$c(u, u') = \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2}.$$

Rondon et al. [76] found that the orthodromic metric is the most suitable distance metric for spherical surfaces. It can handle the periodicity of the latitude, while fitting the spherical geometry distance problem more accurately. Furthermore, Rossi et al. [39] showed that the orthodromic distance is a reliable proxy of the viewport overlap. To appropriately compute the orthodromic distance, we convert the gaze positions to 3D Cartesian points of the spherical surface. Thus, after this transformation, each data point has the form $u = (x, y, z, t)$, where t is its time coordinate.

Figure 17 depicts the empirical Cumulative Distribution Function (CDF) representing the average head speed for each video content. In our analysis, we pinpointed outliers characterized by exceptionally high head movement speeds. By examining the CDF, we ascertain that a suitable threshold for filtering out these outliers is 150° /s. This value effectively encompasses the majority of the typical head speeds recorded. Note that these outliers are rare and typically arise from inaccuracies in the head-tracking system. The HMD's tracking system is equipped with Micro Electro-Mechanical System (MEMS) sensors for orientation data collection. Although head tracking reliability has seen significant improvements in the last decade, certain issues such as drift, tilt, and stationary jitter can still affect data quality [77]. We established a head-speed threshold of 150°/s and excluded data points above this threshold, which allowed us to keep more than 99.9% of the dataset. Figure 18 shows the mean head speed for fixed-motion videos after removal of the outliers, considering the head tracking data for the entire video.

To analyze the head tracking data, we calculate the distance between the gaze direction and the ROI at any given time t. For each experiment trial, defined by the *i*th participant and the *j*th video, we collect a matrix U_{ij} of gaze positions that is expressed as follows:

$$U_{ij} = \begin{bmatrix} \mathbf{u_1} & \mathbf{u_2} & \cdots & \mathbf{u_{N_{ij}}} \end{bmatrix}, \quad (7)$$

where i corresponds to the *i*th participant, j to the *j*th video, and N_{ij} to the total number of gaze positions for a single trial
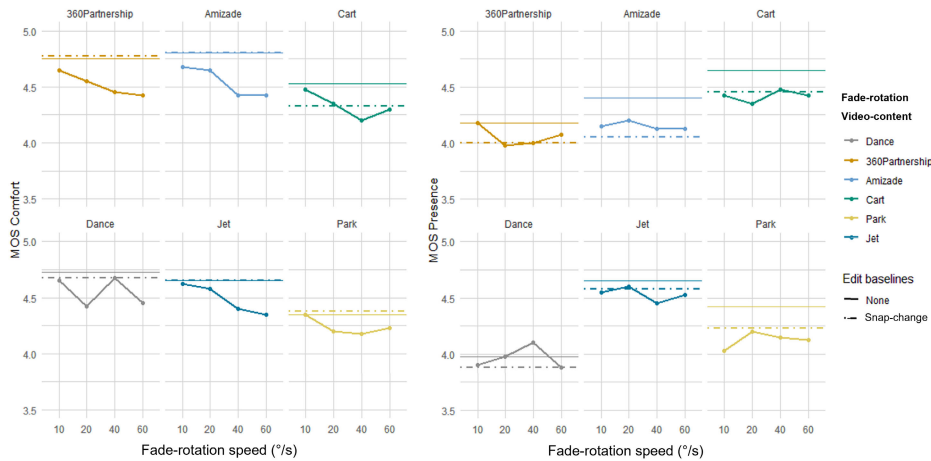
FIGURE 15. Line plot for MOS of presence and comfort. Comparing all rotation speed versions of the Fade-Rotation (FR) against the instant edit (snap-change) and no edit (none) baselines for each video.
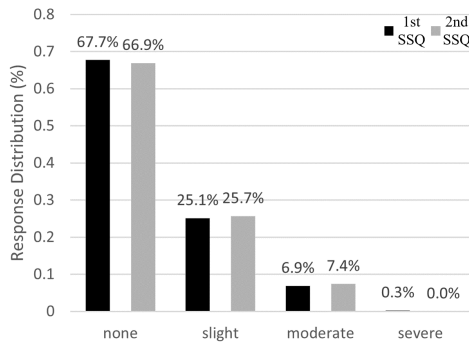

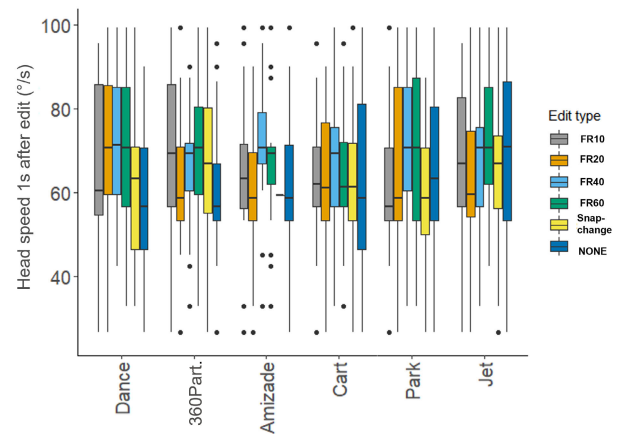**FIGURE 16.** Distribution of all cybersickness symptoms.

FIGURE 18. Boxplot the head speed measured 1s after the edit, for each video-content grouped by edit type.
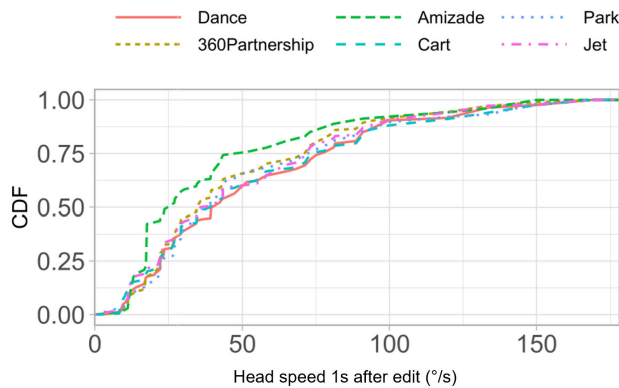

FIGURE 17. CDF of the head speed measured 1s after the edit for each video-content.

of the experiment. We can define each collected gaze position $\mathbf{u_k}$ as a 4D vector, represented as $\mathbf{u_k}$, encompassing the 3D spatial coordinates (x, y, z) and the temporal component (t) of the sample:

$$\mathbf{u_k} = \begin{bmatrix} x_k \\ y_k \\ z_k \\ t_k \end{bmatrix}.$$

To execute the analysis, we need not only the gaze positions but also the ROI positions for each time t. Thus, similarly,

we generate a matrix (V_{ij}) of ROI positions $\mathbf{v_k}$, containing the same number of samples as the user gaze matrix (U_{ij}).

Now, let $\mathcal{T}_{ij} = [\tau_1, \tau_2, \ldots, \tau_{N_{ij}}]$ be an array of time samples from the gaze position data collected in each trial. We define the observation window around a time τ as

$$[\tau_0, \tau_f] = \left[\tau - \frac{\Delta\tau}{2}, \tau + \frac{\Delta\tau}{2} \right],$$

where the time window starts at $k_0 = \text{closest}(\mathcal{T}_{ij}, \tau_0)$ and ends at $k_f = \text{closest}(\mathcal{T}_{ij}, \tau_f)$. $\Delta\tau$ corresponds to the time interval around τ to be used in the analysis. The function *closest()* returns the element x_i that minimizes $|x_i - y|$:

$$closest(x, y) = \text{argmin}_i(|x_i - y|).$$

For a single trial, defined by the ith participant and the jth video, the distance between the gaze of the user (U) and the ROI (V) at time t is given by the mean orthodromic distances around the observation window:

$$\bar{d}(U, V) = \frac{1}{N} \sum_{k=k_0}^{k_f} d(u_k, v_k). \tag{8}$$
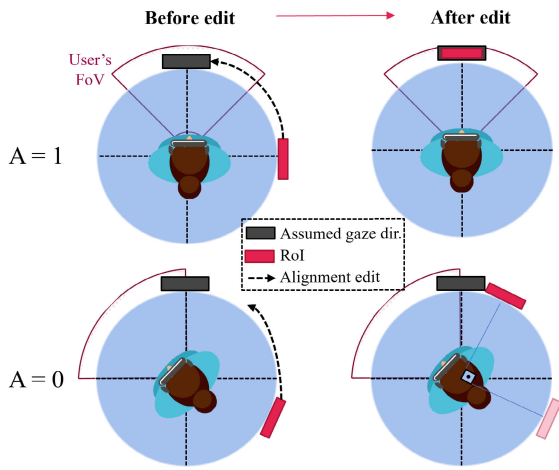
FIGURE 19. Possible states of alignment: $A = 1$ (first row) when alignment is successful, and $A = 0$ (second row) otherwise. The mean distance between user FOV and ROI just after the edit is used to compute the A. We applied a distance threshold of $\tau < 60°$ to classify each trial in terms of A.

FIGURE 20. Boxplots of participant's head speed 1 s after edit for those in the "aligned" ($A = 1$) and "nonaligned" ($A = 0$) groups. The circles show the mean values.

From the gaze-ROI distance, we analyze the research hypothesis H3 and H4. To perform statistical t-tests, we group the experimental trials according to gaze-ROI alignment. To this end, we propose an alignment function A_{ij} that attributes one of two states, "aligned" or "not aligned," to each trial according to the following equation:

$$A_{ij} = \begin{cases} 1, & \text{if } \bar{d}(U_{ij}, V_{ij}) < \delta; \\ 0, & \text{otherwise;} \end{cases} \quad (9)$$

where δ is a threshold value for the maximum distance between gaze and ROI. This value corresponds to a radius around the point of perfect intersection into which we consider gaze and ROI aligned. Figure 19 illustrates these 2 alignment states.

To classify each trial in terms of alignment, we calculate the alignment just after the edit ($t = 15 \ s$ for snap-change, $t = 16 \ s$ for fade-rotation). As shown in Figure 19, for all video rotations, if the participant was looking at the center point ($\theta = 0, \phi = 0$) at time t, she/he would be perfectly "aligned" with the target ROI at the end of the rotation. We classify each trial by computing $\bar{d}(U, V)$ at t. We fixed $\Delta t = 250$ms (equivalent to approximately 30 samples for the typical data sample frequency) and the tolerance region $\tau = 60°$. We chose this tolerance region because both devices used in the experiment have more than 90° FOV. Therefore, if the participant's gaze direction is within 60°, the ROI will be within the FOV [31], [39], [78].

To analyze the effects of alignment on subjective scores, we consider the alignment state A (see Figure 19), which can be "aligned" or "nonaligned", depending on whether the ROI was within an angular distance of 60° or not. Then we group the "aligned" or "nonaligned" cases per edit type, resulting in 2 unbalanced sets per edit type. For each condition, we perform Wilcoxon rank sum tests (with continuity correction) to analyze the differences between "aligned" and
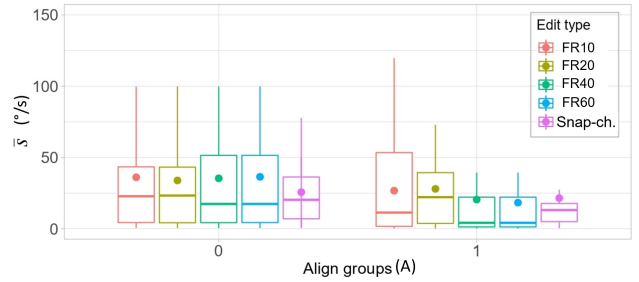
the "nonaligned" sets. There are 15 conditions, resulting from 5 types of edit ("None" not considered) and 3 attributes (presence, comfort, experience). Thus, we applied the t-test between two sets "aligned" and "nonaligned" for each condition, and for each attribute. The only condition where the pair of sets "aligned" and "nonaligned" ($p < 0.05$) had a significant difference between them was FR 10° in the experience attribute.

Therefore, except for the FR 10° experience score, the gaze-ROI alignment classification had no impact on subjective scores, partially fulfilling **H3**.

To complete the **H3** analysis, we performed a Tukey HSD post hoc test on all given combinations of A ("aligned" represented by $A = 1$ and "nonaligned" represented by $A = 0$) and contents (21 comparisons), of A and edit type (15 comparisons), as well as of A and scene motion type (6 comparisons).

In total, we performed 42 non-significant comparisons. No statistically distinguishable differences were found between the group "non-aligned" and the group "aligned". With that, we fulfill **H3**.

We tested the effect of A on the reduction in head motion. For that, first we computed the head movement speed of users at 1 second before and 1 second after the edit. The head movement speed for the i-th participant watching the j-th video at time t is given by:

$$\bar{s} = \frac{1}{N} \sum_{k=k_0}^{k_f} \frac{d(u_k, u_{k+1})}{t_{k+1} - t_k}, \quad (10)$$

where $T_{ijk} = \{t_{ij1}, \ldots, t_{ijk}, \ldots, t_{ijN_{ij}}\}$ are the timestamps inside the temporal window Δt around t, $N_{ij}(t)$ is number of samples inside Δt, and d is the orthodromic distance metric (see Eq. (6)).

Since head movement speed data are continuous, we perform an Anova Omnibus test between the "aligned" ($A = 1$) and "nonaligned" ($A = 0$) groups. The Anova Omnibus test returned $F(30.95, 1, p < 001)$, meaning an F-test with 30.95 degree of freedom and a p-value lower than that of the significance level, which can be interpreted as a significant reduction in head speed after alignment edit. Therefore, aligning with ROI just before editing reduces the speed of head movement, allowing viewers to stabilize their view in

the region. Figure 20 shows the boxplot of the head speed distribution at 1 second after editing, for the two A groups, grouped by edit type. For the aligned group, FR 40° has the lowest values (8% lower than snap-change), followed by FR 60°. The edit types that show a reduction in the average head movement speed are FR 10° $= 14.9$ °, FR 20° $= 9.5$°, FR 40° $= 26.7$°, FR 60° $= 33.1$°, snap-change $= 21.5$°. For all edit types, there is a reduction in head movement speed that may be related to a fixation on an ROI, which reduces exploratory behavior in agreement with the literature [28]. With these results, we prove **H4**, which states that alignment edits reduce head movement speed.

VI. FINAL REMARKS

This research introduces a novel alignment editing technique named ''fade-rotation'' designed specifically for enhancing the experience of watching 360° videos. Unlike comparable methods [29], which rely on user head movements as triggers, our approach employs a predetermined trigger point. This distinction makes the fade-rotation technique well-suited for enabling filmmakers to predetermine the time when edits take place. To comprehensively assess the effectiveness of this technique across varying conditions, we conducted a user study and conducted a comparative analysis against the snap-change edit method [28].

This work fits a category of video edit techniques that are specific to 360° videos, which are known as ''alignment edits.'' Our analysis centers on two specific types of alignment edits: the established snap-change and the method introduced in this work - the fade-rotation. We assess the impacts of content (scene motion, content) and edit parameters (fade-rotation speed, alignment type) on viewer experience, QoE (presence, comfort, experience), and head movement behavior.

Our main conclusions are:

1) The alignment edits tested in this study did not cause a significant degradation on users' comfort or presence. From the subjective feedback participants gave us, the majority of participants did not notice that there were edits in the videos.

2) Video content and scene motion significantly impacted users' ratings in terms of comfort, presence, and experience. This was especially true for contents with objects in motion. From the subjective feedback participants gave us, we highlight that videos with more motion imply more sense of presence and experience while reducing comfort.

3) A Fade-rotation edit with a rotation speed greater or equal than 20°/s should be avoided for dynamic scene motion contents. A fade rotation with a 10°/s rotation speed or a snap-change is a preferable option since they have a lower probability of causing discomfort.

4) The alignment between ROI and FOV impacted the user behavior right after the edit. More specifically, the head motion movement was reduced after the edit. Gradual

(a) Setup with Oculus Quest 2. (b) Setup with Oculus Rift S.

FIGURE 21. Participant wearing the HMD, and watching a experiment's video.

alignments resulted in an 8% lower head movement speed than that of an instant edit.

5) The alignment between ROI and user FOV does not impact the presence, comfort, and experience.

As future work, we plan to extend the dataset by testing the alignment edits on more video contents and using a double-stimulus methodology. In particular, we plan to investigate the effect of ΔT_{fade} on QoE. Note that in the current study, we focused on the parameter rotation speed (ω), which we believe is more relevant to users. We also plan to test a real-time version of the alignment edits, adjusting the video player to take into account the current head motion of the users.

APPENDIX A
FREE AND INFORMED CONSENT TERM

The Digital Signal Processing Group (GPDS) laboratory invites you to participate in the research entitled ''Fade Rotation: Attention-Driving Transition Mechanism for User-Centric Content-Adaptive Virtual Reality Movies''. The expected benefit of this research is to understand the degrees of acceptability of a new attention-driving mechanism in 360-degree videos that in the future should integrate a content adaptation system for optimizing the experience of viewers of immersive videos. The survey is designed to be agile and completely safe for participants. You will always be accompanied by a researcher and the instructions intend to make your participation as simple as possible.

To participate, please read the information below carefully and check ''Yes'' to consent to your participation and start your session, or check ''No'' if you do not wish to participate.

1) **Procedure**: This experiment is scheduled to last 30 minutes, and you will be shown 36 videos of 30 seconds each, giving scores on the watching videos, and answer a pre- and post-questionnaire.

2) **Possible discomfort**: Eventually while watching the immersive videos, you may experience some initial discomfort that diminishes with time. If you need to stop at any moment, just call the researcher in charge. Since one of the measures taken will be the level of discomfort,

we ask that you avoid pausing the video before the end, as this will mean losing data. However, should you wish to quit at any time, this will not cause any harm to you.

3) **Benefits and costs**: Your participation in this study will contribute important results to research in the areas of computer science and immersive media. You will not incur any expenses or burdens from your participation in the study, nor will you receive any kind of reimbursement or gratuity for participating in the research, which is entirely voluntary. This is entirely voluntary, with the exception of those participants who request a transportation stipend.

4) **Privacy and confidentiality**: All information collected in this study is confidential and your name and that of your organization will not be identified in any way. Every effort will be made during data collection to ensure your privacy and anonymity. The data collected during the study is strictly for research activities, following the procedures and rules of the UnB's ethical committee.

5) **Safety protocols for performing subject experiments during the pandemic of Covid-19**: Our experiment will be conducted respecting the safety protocol of the GPDS/ENE/UnB laboratory.

The researchers responsible for the study can provide any clarification about the study by contacting the following e-mail addresses:

- Experimenter (contact): Lucas dos Santos Althoff, 190051612@aluno.unb.br - PPGI/UnB
- Supervisor: Myléne C. Q. Farias - PPGI/UnB

Do you think you are sufficiently informed about the research that will be carried out and do you freely and spontaneously agree to participate as a collaborator?

<p align="center">NO ( )   YES ( )</p>

APPENDIX B
LABORATORY SETUP OF THE EXPERIMENT

The laboratory setup consisted by a swivel chair, a dedicated router, a server PC and the HMD.In the first experiment, the participants used the Oculus Rift S, while in the second experiment they used the Meta Quest 2. Figure 21 shows two participants wearing the two devices. The safety protocols were carried carefully with participants wearing a face-mask and the sanitation of the complete equipment were applied at the beginning and at the end of each session.

ACKNOWLEDGMENT

REFERENCES

[1] R. Marinescu, B. Ghiță, V. Negru, D. C. Mocanu, A. Iacob, and A. Liotta, "Adaptive streaming for virtual reality via reinforcement learning," in *Proc. 26th Eur. Signal Process. Conf. (EUSIPCO)*, 2018, pp. 1561–1565.

[2] Y. Ban, L. Xie, Z. Xu, X. Zhang, Z. Guo, and Y. Wang, "CUB360: Exploiting cross-users behaviors for viewport prediction in 360 video adaptive streaming," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, San Diego, CA, USA, 2018, pp. 1–6, doi: 10.1109/ICME.2018.8486606.

[3] Y. Li, W. Shen, Z. Gao, Y. Zhu, G. Zhai, and G. Guo, "Looking here or there? Gaze following in 360-degree images," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Montreal, QC, Canada, 2021, pp. 3722–3731, doi: 10.1109/ICCV48922.2021.00372.

[4] Grand View Research. (2022). *Virtual Reality Market Size & Share Report 2022-2030*. Accessed: May 31, 2023. [Online]. Available: https://www.grandviewresearch.com/industry-analysis/virtual-reality-vr-market/methodology

[5] L.-H. Lee, T. Braud, P. Zhou, L. Wang, D. Xu, Z. Lin, A. Kumar, C. Bermejo, and P. Hui, "All one needs to know about metaverse: A complete survey on technological singularity, virtual ecosystem, and research agenda," 2021, *arXiv:2110.05352*.

[6] J. S. Pillai, A. Ismail, and H. P. Charles, "Grammar of VR storytelling: Visual cues," in *Proc. Virtual Reality Int. Conf.*, 2017, pp. 1–4.

[7] S. Nassar, "Engaging by design: Utilization of VR interactive design tool in mise-en-scène design in filmmaking," *Int. Design J.*, vol. 11, no. 6, pp. 64–71, Aug. 2021, doi: 10.21608/IDJ.2021.87166.1022.

[8] F. Chiariotti, "A survey on 360-degree video: Coding, quality of experience and streaming," *Comput. Commun.*, vol. 177, pp. 133–155, Sep. 2021.

[9] A. van Kasteren, K. Brunnström, J. Hedlund, and C. Snijders, "Quality of experience of 360° video: Effects of head tracking, image content, and image quality," *Multimedia Tools Appl.*, vol. 81, pp. 9771–9802, 2022.

[10] J. Yu, H. Li, and B. Li, "Bitrate adaptation for virtual reality video streaming over HTTP," *Multimedia Tools Appl.*, vol. 76, no. 20, pp. 21467–21488, 2017.

[11] L. D. S. Althoff, H. D. Garcia, D. D. R. Morais, S. Alamgeer, M. A. Prado, G. C. Araujo, R. Prakash, M. M. Carvalho, and M. C. Q. Farias, "Designing an user-centric framework for perceptually-efficient streaming of 360° edited videos," *Electron. Imag.*, vol. 34, no. 9, pp. 394-1–394-7, Jan. 2022, doi: 10.2352/ei.2022.34.9.iqsp-394.

[12] C. Brown, G. Bhutra, M. Suhail, Q. Xu, and E. D. Ragan, "Coordinating attention and cooperation in multi-user virtual reality narratives," in *Proc. IEEE Virtual Reality (VR)*, Mar. 2017, pp. 377–378, doi: 10.1109/VR.2017.7892334.

[13] M. Wang, Y.-J. Li, W.-X. Zhang, C. Richardt, and S.-M. Hu, "Transitioning360: Content-aware NFoV virtual camera paths for 360° video playback," in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Nov. 2020, pp. 185–194, doi: 10.1109/ISMAR50242.2020.00040.

[14] D. Bordwell, *On the History of Film Style*. Cambridge, MA, USA: Harvard Univ. Press, 1997.

[15] T. Smith and J. Henderson, "Attentional synchrony in static and dynamic scenes," *J. Vis.*, vol. 8, no. 6, p. 773, Mar. 2010, doi: 10.1167/8.6.773.

[16] S. Weaving, "Evoke, don't show: Narration in cinematic virtual reality and the making of entangled," *Virtual Creativity*, vol. 11, no. 1, pp. 147–162, Jun. 2021, doi: 10.1386/vcr_00047_1.

[17] R. M. Young, "Creating interactive narrative structures: The potential for AI approaches," Spring 2000. [Online]. Available: https://aaai.org/proceeding/spring-2000-02/

[18] A. Serrano, V. Sitzmann, J. Ruiz-Borau, G. Wetzstein, D. Gutierrez, and B. Masia, "Movie editing and cognitive event segmentation in virtual reality video," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–12, Aug. 2017, doi: 10.1145/3072959.3073668.

[19] C. O. Fearghail, E. Zerman, S. B. Knorr, F.-Y. Chao, and A. Smolic. (2021). *Use of Saliency Estimation in Cinematic VR Post-Production to Assist Viewer Guidance*. Accessed: Jul. 16, 2023. [Online]. Available: https://v-sense.scss.tcd.ie

[20] J. Brilhart. (2016). *In the Blink of a Mind*. Accessed: Feb. 15, 2022. [Online]. Available: https://medium.com/the-language-of-vr/in-the-blink-of-a-mind-prologue-7864c0474a29

[21] M. Gödde, F. Gabler, D. Siegmund, and A. Braun, "Cinematic narration in VR—Rethinking film conventions for 360 degrees," in *Virtual, Augmented and Mixed Reality: Applications in Health, Cultural Heritage, and Industry*. New York, NY, USA: Springer, 2018, doi: 10.1007/978-3-319-91584-5_15.

[22] R. Cao, J. Walsh, A. Cunningham, C. Reichherzer, S. Dey, and B. Thomas, "A preliminary exploration of montage transitions in cinematic virtual reality," in *Proc. IEEE Int. Symp. Mixed Augmented Reality Adjunct (ISMAR-Adjunct)*, Oct. 2019, pp. 65–70, doi: 10.1109/ISMAR-Adjunct.2019.00031.

[23] T. Kjær, C. B. Lillelund, M. Moth-Poulsen, N. C. Nilsson, R. Nordahl, and S. Serafin, "Can you cut it? An exploration of the effects of editing in cinematic virtual reality," in *Proc. 23rd ACM Symp. Virtual Reality Softw. Technol.*, Nov. 2017, p. 4, doi: 10.1145/3139131.3139166.

[24] M. C. Howard and E. C. Van Zandt, "A meta-analysis of the virtual reality problem: Unequal effects of virtual reality sickness across individual differences," *Virtual Reality*, vol. 25, no. 4, pp. 1221–1246, Dec. 2021, doi: 10.1007/S10055-021-00524-3.

[25] M. Speicher, C. Rosenberg, D. Degraen, F. Daiber, and A. Krüger, "Exploring visual guidance in 360-degree videos," in *Proc. ACM Int. Conf. Interact. Experiences TV Online Video*, Jun. 2019, pp. 1–12, doi: 10.1145/3317697.3323350.

[26] M. Montagud, P. Orero, and A. Matamala, "Culture 4 all: Accessibility-enabled cultural experiences through immersive VR360 content," *Pers. Ubiquitous Comput.*, vol. 24, no. 6, pp. 887–905, Dec. 2020, doi: 10.1007/s00779-019-01357-3.

[27] A. Pavel, B. Hartmann, and M. Agrawala, "Shot orientation controls for interactive cinematography with 360° video," in *Proc. 30th Annu. ACM Symp. User Interface Softw. Technol.*, Oct. 2017, pp. 289–297, doi: 10.1145/3126594.3126636.

[28] S. Dambra, G. Samela, L. Sassatelli, R. Pighetti, R. Aparicio-Pardo, and A.-M. Pinna-Déry, "Film editing: New levers to improve VR streaming," in *Proc. 9th ACM Multimedia Syst. Conf.*, Jun. 2018, pp. 27–39, doi: 10.1145/3204949.3204962.

[29] Y. Farmani and R. J. Teather, "Evaluating discrete viewpoint control to reduce cybersickness in virtual reality," *Virtual Reality*, vol. 24, pp. 645–664, Jan. 2020, doi: 10.1007/s10055-020-00425-x.

[30] A. Serrano, D. Martin, D. Gutierrez, K. Myszkowski, and B. Masia, "Imperceptible manipulation of lateral camera motion for improved virtual reality applications," *ACM Trans. Graph.*, vol. 39, no. 6, pp. 1–14, Dec. 2020, doi: 10.1145/3414685.3417773.

[31] O. Younis, W. Al-Nuaimy, M. H. Alomari, and F. Rowe, "A hazard detection and tracking system for people with peripheral vision loss using smart glasses and augmented reality," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 2, pp. 1–9, 2019, doi: 10.14569/IJACSA.2019.0100201.

[32] J. L. Dorado and P. A. Figueroa, "Methods to reduce cybersickness and enhance presence for in-place navigation techniques," in *Proc. IEEE Symp. 3D User Interfaces (3DUI)*, Mar. 2015, pp. 145–146, doi: 10.1109/3DUI.2015.7131742.

[33] S. Weech, S. Kenny, and M. Barnett-Cowan, "Presence and cybersickness in virtual reality are negatively related: A review," *Frontiers Psychol.*, vol. 10, Feb. 2019, doi: 10.3389/fpsyg.2019.00158.

[34] F. Tian, M. Hua, T. Zhang, and W. Zhang, "Spatio-temporal editing method and application in virtual reality video," in *Proc. IEEE 4th Inf. Technol., Netw., Electron. Autom. Control Conf. (ITNEC)*, vol. 1, Jun. 2020, pp. 2290–2294, doi: 10.1109/ITNEC48623.2020.9085087.

[35] S. Eftekharifar, A. Thaler, A. O. Bebko, and N. F. Troje, "The role of binocular disparity and active motion parallax in cybersickness," *Exp. Brain Res.*, vol. 239, no. 8, pp. 2649–2660, Aug. 2021, doi: 10.1007/s00221-021-06124-6.

[36] S. Ang and J. Quarles, "GingerVR: An open source repository of cybersickness reduction techniques for unity," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces Abstr. Workshops (VRW)*, Mar. 2020, pp. 460–463, doi: 10.1109/VRW50115.2020.00097.

[37] C. Marañes, D. Gutierrez, and A. Serrano, "Towards assisting the decision-making process for content creators in cinematic virtual reality through the analysis of movie cuts and their influence on viewers' behavior," *Int. Trans. Oper. Res.*, vol. 30, no. 3, pp. 1245–1262, May 2023, doi: 10.1111/itor.13106.

[38] S. Rossi and L. Toni, "Navigation-aware adaptive streaming strategies for omnidirectional video," in *Proc. IEEE 19th Int. Workshop Multimedia Signal Process. (MMSP)*, Oct. 2017, pp. 1–6, doi: 10.1109/MMSP.2017.8122230.

[39] S. Rossi, F. De Simone, P. Frossard, and L. Toni, "Spherical clustering of users navigating 360° content," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 4020–4024, doi: 10.1109/ICASSP.2019.8683854.

[40] T. Aitamurto, A. S. Won, S. Sakshuwong, B. Kim, Y. Sadeghi, K. Stein, P. G. Royal, and C. L. Kircos, "From FOMO to JOMO: Examining the fear and joy of missing out and presence in a 360° video viewing experience," in *Proc. CHI Conf. Human Factors Comput. Syst.*, May 2021, pp. 1–14, doi: 10.1145/3411764.3445183.

[41] N. Tian, P. Lopes, and R. Boulic, "A review of cybersickness in head-mounted displays: Raising attention to individual susceptibility," *Virtual Reality*, vol. 26, no. 4, pp. 1409–1441, Dec. 2022, doi: 10.1007/s10055-022-00638-2.

[42] L. E. Garrido, M. Frías-Hiciano, M. P. Moreno-Jiménez, G. N. Cruz, Z. E. García-Batista, K. Guerra-Peña, and L. A. Medrano, "Focusing on cybersickness: Pervasiveness, latent trajectories, susceptibility, and effects on the virtual reality experience," *Virtual Reality*, vol. 26, pp. 1347–1371, Mar. 2022, doi: 10.1007/s10055-022-00636-4.

[43] H. Eom, K. K. Kim, S. Lee, Y.-J. Hong, J. Heo, J.-J. Kim, and E. Kim, "Development of virtual reality continuous performance test utilizing social cues for children and adolescents with attention-deficit/hyperactivity disorder," *Cyberpsychol., Behav., Social Netw.*, vol. 22, no. 3, pp. 198–204, Mar. 2019, doi: 10.1089/cyber.2018.0377.

[44] L. Sassatelli, M. Winckler, T. Fisichella, A. Dezarnaud, J. Lemaire, R. Aparicio-Pardo, and D. Trevisan, "New interactive strategies for virtual reality streaming in degraded context of use," *Comput. Graph.*, vol. 86, pp. 27–41, Feb. 2020, doi: 10.1016/j.cag.2019.10.005.

[45] R. Shi, H.-N. Liang, Y. Wu, D. Yu, and W. Xu, "Virtual reality sickness mitigation methods," *ACM Comput. Graph. Interact. Techn.*, vol. 4, no. 1, pp. 1–16, Apr. 2021, doi: 10.1145/3451255.

[46] Y.-X. Lin, R. Venkatakrishnan, R. Venkatakrishnan, E. Ebrahimi, W.-C. Lin, and S. V. Babu, "How the presence and size of static peripheral blur affects cybersickness in virtual reality," *ACM Trans. Appl. Perception*, vol. 17, no. 4, pp. 1–18, Oct. 2020, doi: 10.1145/3419984.

[47] C. Yildirim, "Don't make me sick: Investigating the incidence of cybersickness in commercial virtual reality headsets," *Virtual Reality*, vol. 24, no. 2, pp. 231–239, Jun. 2020, doi: 10.1007/s10055-019-00401-0.

[48] J. Teixeira and S. Palmisano, "Effects of dynamic field-of-view restriction on cybersickness and presence in HMD-based virtual reality," *Virtual Reality*, vol. 25, no. 2, pp. 433–445, Jun. 2021, doi: 10.1007/s10055-020-00466-2.

[49] ITU-T. (2020). *Subjective Test Methodologies for 360° Video on Head-Mounted Displays*. Accessed: Feb. 15, 2022. [Online]. Available: https://www.itu.int/rec/T-REC-P.919-202010-I

[50] J. Gutiérrez et al., "Subjective evaluation of visual quality and simulator sickness of short 360° videos: ITU-T Rec. P.919," *IEEE Trans. Multimedia*, vol. 24, pp. 3087–3100, 2022, doi: 10.1109/tmm.2021.3093717.

[51] D. Wolf, M. Rietzler, L. Bottner, and E. Rukzio, "Augmenting teleportation in virtual reality with discrete rotation angles," 2021, *arXiv:2106.04257*.

[52] P. Pérez, N. Oyaga, J. J. Ruiz, and A. Villegas, "Towards systematic analysis of cybersickness in high motion omnidirectional video," in *Proc. 10th Int. Conf. Quality Multimedia Exper. (QoMEX)*, May/Jun. 2018, pp. 1–3, doi: 10.1109/QoMEX.2018.8463377.

[53] S. Bouchard, G. Robillard, J. St-Jacques, S. Dumoulin, M. J. Patry, and P. Renaud, "Reliability and validity of a single-item measure of presence in VR," in *Proc. 3rd IEEE Int. Workshop Haptic, Audio Visual Environments Their Appl.*, Oct. 2004, pp. 59–61, doi: 10.1109/HAVE.2004.1391882.

[54] K. Borchert, A. Seufert, E. Gamboa, M. Hirth, and T. Hoßfeld, "In vitro vs in vivo: Does the study's interface design influence crowdsourced video QoE?" *Qual. User Exper.*, vol. 6, no. 1, p. 1, Dec. 2021, doi: 10.1007/s41233-020-00041-2.

[55] T. C. Peck, K. A. Mcmullen, and J. Quarles, "DiVRsify: Break the cycle and develop VR for everyone," *IEEE Comput. Graph. Appl.*, vol. 41, no. 6, pp. 133–142, Nov. 2021, doi: 10.1109/mcg.2021.3113455.

[56] K. Rahimi, C. Banigan, and E. D. Ragan, "Scene transitions and teleportation in virtual reality and the implications for spatial awareness and sickness," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 6, pp. 2273–2287, Jun. 2020, doi: 10.1109/TVCG.2018.2884468.

[57] V. Sevinc and M. I. Berkman, "Psychometric evaluation of simulator sickness questionnaire and its variants as a measure of cybersickness in consumer virtual environments," *Appl. Ergonom.*, vol. 82, Jan. 2020, Art. no. 102958, doi: 10.1016/j.apergo.2019.102958.

[58] P. Schober, C. Boer, and L. A. Schwarte, "Correlation coefficients: Appropriate use and interpretation," *Anesthesia Analgesia*, vol. 126, no. 5, pp. 1763–1768, 2018, doi: 10.1213/ANE.0000000000002864.

[59] R. Xu, H. Wang, S. Petrangeli, V. Swaminathan, and S. Bagchi, "Closing-the-loop: A data-driven framework for effective video summarization," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2020, pp. 201–205, doi: 10.1109/ISM.2020.00042.

[60] S. Lee, J. Sung, Y. Yu, and G. Kim, "A memory network approach for story-based temporal summarization of 360° videos," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1410–1419, doi: 10.1109/CVPR.2018.00153.

[61] S. Litleskare and G. Calogiuri, "Camera stabilization in 360° videos and its impact on cyber sickness, environmental perceptions, and psychophysiological responses to a simulated nature walk: A single-blinded randomized trial," *Frontiers Psychol.*, vol. 10, Nov. 2019, doi: 10.3389/fpsyg.2019.02436.

[62] S. Knorr, C. Ozcinar, C. O. Fearghail, and A. Smolic, "Director's cut: A combined dataset for visual attention analysis in cinematic VR content," in *Proc. 15th ACM SIGGRAPH Eur. Conf. Vis. Media Prod.*, Dec. 2018, pp. 1–10, doi: 10.1145/3278471.3278472.

[63] A. T. Nasrabadi, A. Samiei, A. Mahzari, R. P. Mcmahan, R. Prakash, M. C. Q. Farias, and M. M. Carvalho, "A taxonomy and dataset for 360° videos," in *Proc. 10th ACM Multimedia Syst. Conf.*, Jun. 2019, pp. '1–7, doi: 10.1145/3304109.3325812.

[64] N. K. Sharda, "Creating ambient multimedia experience on-demand: Opportunities, challenges and research models," in *Proc. 1st ACM Int. Workshop Semantic Ambient Media Experiences*, 2008, pp. 41–48.

[65] T. Beck and S. Rothe, "Applying diegetic cues to an interactive virtual reality experience," in *Proc. IEEE Conf. Games (CoG)*, Aug. 2021, pp. 1–8, doi: 10.1109/CoG52621.2021.9619025.

[66] J. Mateer, "Directing for cinematic virtual reality: How the traditional film director's craft applies to immersive environments and notions of presence," *J. Media Pract.*, vol. 18, no. 1, pp. 14–25, Jan. 2017, doi: 10.1080/14682753.2017.1305838.

[67] R. Sagot-Duvauroux, F. Garnier, and R. Ronfard, "(Re-)framing virtual reality," in *Proc. Eurographics/EuroVis Working Group Intell. Cinematography Editing*, 2022, pp. 1–8, doi: 10.2312/wiced.20221051.

[68] R. Carpio and J. Birt, "The role of the embodiment director in virtual reality film production," *Creative Industries J.*, vol. 15, no. 2, pp. 189–198, May 2022, doi: 10.1080/17510694.2021.2017634.

[69] A. Kroma, "The technical dilemmas of creative design and rapid prototyping for immersive storytelling," in *Proc. 14th Conf. Creativity Cognition*, 2022, pp. 698–703, doi: 10.1145/3527927.3533730.

[70] F. Danieau, A. Guillo, and R. Doré, "Attention guidance for immersive video content in head-mounted displays," in *Proc. IEEE Virtual Reality (VR)*, Mar. 2017, pp. 205–206, doi: 10.1109/VR.2017.7892248.

[71] M. Prado, L. Althoff, S. Alamgeer, A. R. E. Silva, R. Prakash, M. M. Carvalho, and M. C. Q. Farias, "360RAT: A tool for annotating regions of interest in 360-degree videos," in *Proc. Brazilian Symp. Multimedia Web*, Nov. 2022, pp. 272–280, doi: 10.1145/3539637.3557930.

[72] Cannes. (2022). *CANNES XR: Showcases*. Accessed: Feb. 19, 2022. [Online]. Available: https://www.marchedufilm.com/programs/cannes-xr/showcases/

[73] S. Rothe, D. Buschek, and H. Hußmann, "Guidance in cinematic virtual reality-taxonomy, research status and challenges," *Multimodal Technol. Interact.*, vol. 3, no. 1, p. 19, Mar. 2019, doi: 10.3390/MTI3010019.

[74] H.-N. Hu, Y.-C. Lin, M.-Y. Liu, H.-T. Cheng, Y.-J. Chang, and M. Sun, "Deep 360 pilot: Learning a deep agent for piloting through 360° sports videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1396–1405, doi: 10.1109/CVPR.2017.153.

[75] ITU-T. (1999). *Recommendation P.910: Subjective Video Quality Assessment Methods for Multimedia Applications*. [Online]. Available: https://www.itu.int/rec/T-REC-P.910

[76] M. F. R. Rondon, L. Sassatelli, R. Aparicio-Pardo, and F. Precioso, "TRACK: A new method from a re-examination of deep architectures for head motion prediction in 360° videos," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 5681–5699, Sep. 2022, doi: 10.1109/TPAMI.2021.3070520.

[77] S. M. LaValle, A. Yershova, M. Katsev, and M. Antonov, "Head tracking for the oculus rift," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2014, pp. 187–194, doi: 10.1109/ICRA.2014.6906608.

[78] J. W. Kelly, T. A. Doty, M. Ambourn, and L. A. Cherep, "Distance perception in the oculus quest and oculus quest 2," *Frontiers Virtual Reality*, vol. 3, pp. 1–7, Mar. 2022, doi: 10.3389/frvir.2022.850471.

[79] J. Jerald, T. C. Peck, F. Steinicke, and M. C. Whitton, "Sensitivity to scene motion for phases of head yaws," in *Proc. 5th Symp. Appl. Perception Graph. Vis.*, 2008, pp. 155–162.

LUCAS S. ALTHOFF received the B.S. degree in computational physics and the M.Sc. degree in physics education from the University of Brasília (UnB), Brazil, in 2014 and 2018, respectively, where he is currently pursuing the Ph.D. degree with the Computer Science Department, with a focus on quality of experience research for immersive multimedia. He joined the Department of Compute Science, University Center of Brasília (CEUB), in 2023, as an Assistant Professor. Previously, he was with the XLIM Laboratory, France. His research interests include distributed systems, content creation, and visual perception for virtual reality applications.

MYLÉNE C. Q. FARIAS (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the Federal University of Pernambuco (UFPE), Brazil, in 1995, the M.Sc. degree in electrical engineering from the State University of Campinas (UNICAMP), Brazil, in 1998, and the Ph.D. degree in electrical and computer engineering from the University of California at Santa Barbara (UCSB), Santa Barbara, CA, USA, in 2004, with a focus on no-reference video quality metrics. She was a Research Engineer of video quality assessment and validation of video quality metrics with CPqD, Brazil. She was an intern with Philips Research Laboratories, The Netherlands, in video quality assessment of sharpness algorithms, and Intel Corporation, Phoenix, USA, developing no-reference video quality metrics. She is currently an Associate Professor with the Department of Computer Science, Texas State University. Her current research interests include video quality metrics, video processing, multimedia signal processing, watermarking, and visual attention. She is a member the IEEE Signal Processing Society, ACM, and SPIE.

ALESSANDRO RODRIGUES SILVA received the master's degree in computer science from the Federal University of São Carlos, in 2003. He is currently pursuing the Ph.D. degree with UNB, under the guidance of Dr. Myléne C. Q. Farias, specializing in the area of 3D video quality. He is also a Professor with the Federal Institute of Goiás-Campus Anápolis. His work in the field of computer science focuses on image and video processing using real-time programming techniques as well as multimedia services platforms.

MARCELO M. CARVALHO (Member, IEEE) received the bachelor's degree in electrical/electronics engineering from the Federal University of Pernambuco, in 1995, the master's degree in electrical engineering from the State University of Campinas (Unicamp), in 1998, the master's degree in electrical and computer engineering from the University of California at Santa Barbara, Santa Barbara, CA, USA, in 2000, and the Ph.D. degree in computer engineering from the University of California, Santa Cruz, in 2006. He is currently an Assistant Professor with the Ingram School of Engineering, Texas State University. Previously, he was with the Department of Electrical Engineering, University of Brasília. His research interests include the Internet of Things, wireless and mobile networking, computer networks, and multimedia streaming.

● ● ●