Utrecht University

Watch your back!: Dynamic thumbnails for a 360° video player to enhance viewing experience on 2D displays

Master's Thesis in Game and Media Technology

Author: Jakub Kováč
Student Number: 4443691

Project Supervisor/First Examiner: Dr. Wolfgang Hürst
Second Examiner: Dr. Remco C. Veltkamp

July 3, 2024

ACKNOWLEDGEMENTS

First and foremost, I extend my gratitude to my supervisor, Dr. Wolfgang Hürst, for his continuous feedback and invaluable guidance throughout the course of this thesis. His insights and support were instrumental in shaping this work.

I am also grateful to my second examiner, Dr. Remco C. Veltkamp, for his time and effort in evaluating my thesis. His expertise and thorough assessment were greatly appreciated.

A special thanks to Dr. Christof van Nimwegen for trusting me with access to Lili's Proto Lab. His generosity allowed me to utilize the university's hardware for my user studies, and his support with technical issues was crucial to the success of my research.

I would also like to express my sincere appreciation to my friends and family for their unwavering support during the completion of this thesis. Their encouragement and understanding were a constant source of motivation.

Lastly, I am immensely thankful to all the participants in my user studies. Their time, enthusiasm, and valuable insights significantly contributed to the depth and quality of this research.

Contents

Watch your back!: Dynamic thumbnails for a 360° video player to enhance viewing experience on 2D displays

Jakub Kováč
Utrecht University
Utrecht, Netherlands
j.kovac@students.uu.nl

ABSTRACT

Addressing the challenge of navigating 360° content on 2D displays, we propose a novel 360° video player featuring dynamic thumbnails. Our user study compared this video player to a conventional 360° video player. The goal of this research is to investigate the extent to which dynamic thumbnails reduce fear of missing out (FOMO) and enhance timeline navigation while trying not to hinder the video player's usability. Our findings indicate that dynamic thumbnails significantly reduce FOMO, as users feel less anxious about missing content and exert reduced camera rotation during viewing sessions. Task completion times for timeline navigation showed no significant difference, and participants' gaze fixations revealed similar patterns between the two video players. The System Usability Scale (SUS) was used to evaluate our proposed video player, resulting in it being categorized as 'Good' with a score of 84.7. Participants appreciated the enhanced situational awareness provided by the thumbnails, although some usability issues were noted. These findings suggest that dynamic thumbnails have great potential for 360° video players, and further research is encouraged to optimize their integration and address encountered usability concerns.

1 INTRODUCTION

360° videos, also referred to as omnidirectional or panoramic videos, have gained widespread popularity in recent years. They are captured using specialized omnidirectional cameras that, unlike traditional cameras with limited fields of view, are capable of recording from all directions. Major video platforms such as YouTube[1] and Facebook[2] have played a pivotal role in championing their utilization as they are trying to take advantage of the new medium. Initially, 360° videos found primary use in gaming and entertainment content. However, their applications have expanded significantly, with a increasing number of uses in diverse fields such as education, tourism [22, 41], and sports, among many others.

One might naturally infer that 360° videos were primarily designed for VR head-mounted displays (HMDs) like HTC Vive or Oculus Rift [20], affordability of which is increasing for the average consumer. Despite this, HMDs are not always available and the predominant consumption of 360° videos still occurs on traditional 2D displays, such as those on desktop computers or mobile devices. Additionally, even though HMDs offer users a more immersive experience, it often comes at the expense of heightened cognitive burden, motion sickness and physical discomfort [7]. Hence, as highlighted by participants in a study by Passmore et al. [30] on the viewer experience of watching 360° videos, prolonged video consumption may potentially lead to symptoms of cybersickness. Consequently, given the choice, consumers might still lean towards watching 360° videos on 2D displays.

HMDs offer users an immersive experience by allowing them to explore their surroundings intuitively, simply by moving their heads. This creates a natural and seamless interaction with the virtual environment, as the display adjusts dynamically to their head movements. In contrast, users watching 360° videos on traditional 2D displays navigate the content differently. They interact with a normal field of view (NFOV) viewport, which represents a portion of the entire 360° scene. To explore different parts of the video, they must manually rotate this viewport using input devices such as a mouse or touchscreen. This poses a challenge as users need to constantly readjust their focus on their intended targets [23]. For instance, if a new target were to appear on the opposite side of the video or if there is a scene change, users must rotate to see what is going on, potentially losing track of their initial target in the process. Alternatively, 360° videos can be displayed in their entirety using equirectangular projection [29], but the unfamiliar format and geometrical distortions make the video difficult to watch [19, 22].

Consequently, users are restricted to glimpsing only a small fraction of a 360° scene at any given time. This limitation can trigger a fear of missing out (FOMO) among users, as they are concerned about missing important content that might lie beyond their current field of view (FOV) [12, 30, 34, 40, 44]. Additionally, when users attempt to navigate through a 360° video using the timeline, conventional 360° video players, such as the one found on YouTube, do not provide them with the ability to see the entire 360° scene. Instead, YouTube only offers a small equirectangular projection thumbnail above the timeline when hovering over a specific point in time. The small size of the thumbnail and the distortions in the projection make it difficult to explore the entire surroundings of the 360° scene while skimming through it.

This study aims to investigate the possibility of enhancing 360° video players for 2D displays to address users' FOMO during 360° video consumption. To achieve this, we propose a method of expanding the coverage of a 360° video at any given time through the incorporation of dynamic thumbnail add-ons. These thumbnails are strategically designed to enhance the user experience both during live viewing of a 360° video and when navigating through a 360° video along its timeline. To test the effectiveness of this innovative 360° video player design, we formulate the following research questions:

[1]https://www.youtube.com/
[2]https://www.facebook.com/

- **RQ1** To what extent does the proposed interface mitigate users' fear of missing out during the live viewing of 360° videos, as compared to conventional 360° video players?
- **RQ2** In what ways does the proposed interface affect user navigation (in terms of task completion time and total rotation exerted) through a video's timeline compared to conventional 360° video players?
- **RQ3** How do users perceive the usability of the proposed interface in contrast to conventional 360° video players?

2 RELATED WORK

2.1 Fear of missing out

Fear of missing out (FOMO) is a phenomenon primarily researched in the context of social media [2, 28, 33]. It is defined as "a pervasive apprehension that others might be having rewarding experiences from which one is absent" and characterized by "the desire to stay continually connected with what others are doing" [33]. Research suggests that individuals who experience FOMO tend to exhibit more physical and depressive symptoms [5, 28] and report lower general mood and reduced life satisfaction [33]. While FOMO is typically scrutinized due to its negative effects, certain studies propose that it may not be entirely negative [3, 28].

Aitamurto et al. [3] conducted a thorough examination of FOMO within the context of 360° videos. They define FOMO as fear of missing out on parts of a video outside of user's FOV that contribute to the narrative. This phenomenon is inevitable as the FOV a user looks through is limited, typically varying from 60° to 110° [21, 34, 42, 43]. Hence, when a user chooses to watch a specific segment of a 360° video scene, they are intentionally forgoing the opportunity to witness events unfolding beyond the confines of their current FOV. Their findings show that the primary contributing factor to FOMO in 360° videos is the user's awareness of events occurring in parallel. This awareness causes anxiety, frustration, and stress, as at any given moment, the user would need to adjust their FOV to keep track of important events while simultaneously getting distracted by looking around the environment or observing things happening in other sections of the video. These findings are further supported by other studies in which participants reported being distracted by their surroundings [12, 30, 34, 40], expressing feelings of confusion and fear of missing out on events happening elsewhere in the video, as they were unsure where to direct their attention [30].

Furthermore, Uslu et al. [41] examined the influence of FOMO in the context of 360° tour videos, specifically in museums. They suggest that users who value continual learning experience anxiety about missing out on information relevant to their interest. They highlight how users' intent matters when it comes to watching a video; whether they are watching casually or attempting to view it as a learning experience heavily influences the intensity of FOMO.

2.2 Viewing techniques

Lin et al. [24] introduced a visualization technique called *Outside-In*, which incorporates off-screen regions of interest (ROIs) onto the viewport through picture-in-picture (PIP) thumbnails. This technique enables users to focus on the main event while easily perceiving information that lies outside the confines of their current FOV. While effective, some participants expressed concerns about screen occlusion when numerous PIP thumbnails were present at once.

Meanwhile, Pavel et al. [31] adopted a different approach. Rather than bringing the ROI to the user, they reposition the user's perspective towards the ROI. Their *viewport-oriented* technique reorients the shot at each cut, ensuring that the most important ROI aligns with the user's FOV. With their *active reorientation* technique, users can press a button on an input device to promptly realign the shot at any given time so that the most important ROI lies within their FOV. However, these techniques have drawbacks, such as reorienting when the user might not want it and relying on additional input from the device. The system also relies on labels to mark the positions of the ROIs in the video.

Yamaguchi et al. [43] directly addressed the challenge of alleviating FOMO during the viewing of 360° videos in VR. They introduced a innovative solution in the form of a panoramic thumbnail situated at the tip of the VR controller. This thumbnail is not always visible; users can choose to display or hide it at their convenience. The panoramic thumbnail utilizes equirectangular projection, providing users with an on-demand overview of the entire video at any given time and allowing them to observe events occurring throughout the whole 360° scene.

Lin et al. [23] concentrated on the challenge of users needing to focus and refocus on their intended target when viewing 360° videos. To address this issue, they devised two assistance techniques: *Auto Pilot (AP)*, which guides the user directly to the target, and *Visual Guidance (VG)*, which indicates the location of the target through visual cues. While *AP* quickly and precisely directs the user to the intended target, some participants highly valued the freedom of self-exploration and disliked being guided to watch a specific segment of the video.

Kang et al. [19] condense and classify the aforementioned techniques into two distinct categories: automatic navigation [15, 21, 23, 36, 37] and interactive navigation [23, 24, 31, 43]. The former essentially involves transforming 360° videos into standard non-360° videos based on specific criteria. In contrast, interactive navigation allows users to freely interact with the 360° video while receiving additional assistance to ensure they do not miss important events. Kang et al. identify a drawback with automatic navigation: users lose the ability to interact with 360° videos, which is their primary function. In response, they propose a hybrid system that combines automatic and interactive navigation. This system identifies the most salient areas in the input video and computes autopilot paths based on these scores. Users retain the ability to change their viewing direction, and the system adapts accordingly by generating a new path based on the users' intent. However, a drawback of this approach is that preprocessing paths for a video requires significant computation time.

2.3 Timeline navigation

Li et al. [22] introduced the concept of *Route Tapestries*, which involves seamlessly mapping scenes along predefined camera routes using an orthographic-perspective approach. Their method is designed specifically for 360° video players intended for 2D screens. Li et al. highlight that current systems use planar thumbnails to provide an overview of entire frames for temporal navigation. These thumbnails are created by transforming spherical 360° images into 2D visualizations, such as the equirectangular projection method employed by platforms like YouTube.

While equirectangular projections can offer users an overview of the current frame, navigating through them can be challenging. Li et al. demonstrated that previews with reduced distortions significantly enhance usability, making it easier for users to identify scenes. However, a drawback of implementing *Route Tapestries* is the computational intensity required to generate them.

2.4 Design implications

Zoric et al. [44] highlight design implications in their research on panoramic videos, emphasizing the balance between active and passive viewing. While having full control over a video can be exciting, excessive interaction can also become tiresome. Constant control requires significant effort, leading to potential exhaustion and stress, especially over extended viewing sessions. It is crucial that the experience of free-form viewing and exploration remains intuitive for users. Navigating within the environment should be seamless to ensure users do not feel disoriented or lost amidst dynamic scenes.

Facilitating this can be achieved through the implementation of enhanced overviews of the entire scene, such as panoramic thumbnails by Yamaguchi et al. [43], which allow users to maintain a sense of spatial awareness. On the other hand, providing supplementary detailed views within the video enhances the user's ability to delve into specific elements of interest, encouraging a more engaging and user-friendly exploration of the content, as demonstrated by Len et al.'s *Outside-In* technique [24].

3 IMPLEMENTATION

Our proposed 360° video player interface was implemented using the Unity[3] game engine (version 2021.3.17f1). The interface features two side-view thumbnail add-ons in the top-left and top-right corners of the viewport within a standard 360° video player. The top-right side-view thumbnail provides a glimpse of the content on the right side of the video along the horizontal axis, while the top-left side-view thumbnail offers a corresponding view of the left side. The purpose of these side-view thumbnails was to expand the video's horizontal coverage for users within the constraints of the viewport, without distorting the projection or requiring them to rotate their camera. Additionally, a rear-view thumbnail add-on was included to display what was directly behind the user's current view. It was crucial to include a rear-view thumbnail, as rotating one's camera to look directly behind requires the most exertion. Consequently, the rear-view thumbnail has bigger dimensions compared to the side-view thumbnails, as shown in Figure 1. An analogy would be to envision these dynamic thumbnails as the rear-view and side-view mirrors in a car, integrated within the video player itself.

3.1 Interaction

At the beginning of the video, dynamic thumbnails are visible by default, offering users the flexibility to toggle them on or off based on their preferences. This feature draws inspiration from Yamaguchi et al.'s panoramic thumbnails [43], adapted for 2D displays given our interface does not utilize a VR controller. To control the thumbnails, users simply drag their mouse towards the top middle of the viewport. This action triggers the appearance of a lock button at the bottom of the rear-view thumbnail, as shown in Figure 1. Pressing this button unlocks the thumbnails, which disappear when the mouse returns to the middle of the viewport. The decision to use a button for unlocking, rather than a key press, simplifies interaction by reducing the reliance on multiple input devices, addressing concerns noted in Pavel et al.'s *active reorientation* technique [31].

Figure 1: Proposed layout for dynamic thumbnails within a standard 360° video player interface. Includes a lock button positioned at the bottom of the rear-view thumbnail.

If users wish to quickly scan their surroundings in the 360° scene without locking the thumbnails, they can simply drag their mouse towards the top of the viewport. However, in this mode, simultaneous rotation with the camera is disabled. This approach aims to provide users with an on-demand access to a comprehensive view of the entire 360° scene within the constraints of a 2D NFOV viewport. The objective is to reduce the need for users to constantly rotate the view to ensure they do not overlook significant events.

While related work such as Yamaguchi et al.'s panoramic thumbnails [43] focuses on providing a complete 360° scene overview using the equirectangular projection, our approach only reveals potentially important segments of the 360° scene while highlighting less accessible areas. In contrast to applications using automated navigation adaptation [15, 21, 23, 36, 37] or highlighting specific regions [23, 24, 31], our method encourages user exploration without guided assistance.

3.2 Positioning

The decision to position the dynamic thumbnails in the top horizontal third of the viewport of a 360° video player is motivated by the intent to ensure their minimally obtrusive presence, as shown in Figure 1. This choice is based on the assumption that users primarily fixate their gaze on the vertical central area of the viewport, where the primary events occur. This assumption is supported by Duanmu et al.'s study [11], which generated heatmaps for various 360° video types, indicating that users tend to concentrate their focus along the horizontal axis in the vertical center of the viewport.

Even when interesting events take place at the top portion of the video, the dynamic thumbnails are designed to be on-demand, as elaborated in Section 3.1. This allows users to hide them when

[3]https://unity.com/

they wish to view content in the upper part of the video. Additionally, conventional 360° video players position their temporal navigation widgets, such as the timeline for scrolling, fast forward, and play/pause buttons, at the bottom of the viewport. By placing the dynamic thumbnails at the top of the viewport, the goal is to establish a clear distinction between spatial and temporal navigation, mitigating potential confusion for the user.

The lock button is positioned at the bottom of the rear-view thumbnail, ensuring it is easily accessible yet unobtrusive. This placement is designed with user convenience in mind.

3.3 Field of view

While the goal is to reveal content that users cannot see while focusing on a specific segment in a conventional 360° video player, achieving comprehensive coverage without compromising viewing comfort proves unattainable [19, 22]. NFOV viewports typically offer a FOV ranging from 60° to 110° [21, 34, 42, 43]. In our implementation, we opted for a 95° FOV for the main view, as exceeding this made the video appear distorted and difficult to watch. Given our prioritization of the rear-view thumbnail as the most crucial of the three dynamic thumbnails, we allocated the same 95° FOV for it. Consequently, we had 170° remaining for the side-view thumbnails.

Initially, dividing this remaining FOV equally might seem intuitive. However, in our setup, where the 360° video is projected onto a sphere (the skybox in the Unity engine), a unique distortion occurs. The rectangular-shaped viewports, positioned alongside each other along the equator, exhibit convergence and overlap towards the edges of the sphere. This distortion stems from the spherical projection, which wraps flat surfaces around curved geometry, causing stretching and deformation towards the periphery.

To mitigate this distortion, we introduced 10° spacings between the main view and the dynamic thumbnails. This strategy aimed to alleviate the convergence and overlap observed at the edges of the sphere. With four such spacings in total, we were left with 130° for the two side-view thumbnails, evenly split into 65° each. This arrangement is shown in Figure 2.

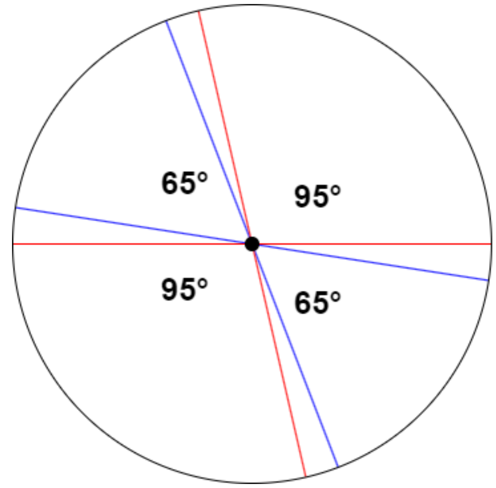


Figure 2: Field of view distribution for main view and rear-view thumbnail (both red) and side-view thumbnails (blue), with 10° spacings in between them.

4 METHODOLOGY

The user study used a within-subject design, assuming participants have prior familiarity with conventional video players used for non-360° videos. Conventional 360° video players share similar usability features with their non-360° counterparts, with the key distinction being the former's capability to alter the viewing direction. Participants were assigned the task of interacting with both a conventional and our proposed 360° video player throughout the user study. The conventional 360° video player served as a baseline for comparison when evaluating our proposed design.

The user study consisted of two separate sections: live viewing and timeline navigation. In the live viewing section, participants were instructed to watch the videos as they normally would. In the timeline navigation section, participants searched for a specific object within the videos they had previously watched during live viewing, using the video's timeline.

During the user study, participants' interactions with the 360° video players were tracked using an eye-tracker, which also recorded the duration of participants' gaze fixation on specific areas of a video within the viewport. The entirety of the user study's flow was overseen by the *iMotions*[4] software, known for gathering and analyzing human behavior data via eye-tracking. This software facilitated a presentation-style flow, where slides with text instructions were followed by an automatic opening of a corresponding video. Upon completion of a video and closing of the video player, the software transitioned to the next slide with instructions.

Throughout the experiment, participants filled out a questionnaire using Qualtrics to track their subjective experience with the two 360° video players.

4.1 Restrictions

To maintain the integrity of the study, we implemented some restrictions for participants during both the live viewing and timeline navigation sections.

4.1.1 Live viewing. Participants were restricted from manipulating temporal navigation during the live viewing section. Upon opening the 360° video player, the video played automatically, ensuring seamless transitions between tasks during the user study. This approach standardized viewing time as a controlled variable, preventing participants from rotating around the video before it started, and ensuring each participant started from the same initial perspective. Maintaining time as a controlled variable was crucial; participants were unable to use the timeline to skip ahead or backtrack. This ensured consistent viewing experiences across all participants, minimizing potential confounding variables.

4.1.2 Timeline navigation. The timeline remained visible to participants at all times, serving various functions during both live viewing and timeline navigation sections. In live viewing, although non-interactable, it indicated the progress of the video watched by the participant. During timeline navigation section, additional features were unlocked, allowing participants to utilize play and pause buttons and navigate through the video by clicking on specific

[4]https://imotions.com/

points on the timeline. This feature enabled swift parsing through scenes to facilitate search activities. Additionally, dynamic thumbnails were constantly displayed during timeline navigation tasks, similar to YouTube's behavior of previewing thumbnails when hovering over the timeline.

4.2 Procedure

Participants received an information sheet and a consent form digitally before the experiment, ensuring they were aware of expectations. At the start, they were presented with the same information within the Qualtrics questionnaire and asked to consent. After signing, they completed an initial section providing basic demographic information and their prior experience with 360° videos.

Before starting, the eye-tracker underwent calibration for each participant to ensure maximum accuracy. Participants were then given time to familiarize themselves with our proposed 360° video player with dynamic thumbnails by watching and interacting with a short 360° practice video. During this video, the participants were also given an explanation of the dynamic thumbnails' functionality. Although timeline functions were disabled during the live viewing section of the user study as mentioned in Section 4.1.2, participants could test them during the practice video. The practice video differed from those used in the actual tasks and was primarily used to minimize the impact of novelty on performance.

In the live viewing section of the user study, participants were tasked with watching three different videos, each lasting approximately one minute. To facilitate varied viewing experiences, the videos were strategically divided around the midpoint, resulting in six segments — three for the first half and three for the second half. This segmentation ensured that each segment could be viewed using a different 360° video player. It is important to note that the content between two segments of the same video was attempted to be as similar as possible. Additionally, the viewing of the first half segment always preceded its corresponding second half.

Before the user study, the sequence of video viewing and the choice of the video player for each segment were thoroughly counterbalanced. This careful planning aimed to optimize the reliability, validity, and general applicability of the results. In each instance, participants started viewing with either the conventional or our proposed 360° video player, transitioning to the other video player upon completion of the first half.

After finishing the second half of each video, participants were asked to respond to three statements related to FOMO (RQ1) in the Qualtrics questionnaire, adapted from Yamaguchi et al.'s 7-point Likert scale questionnaire [43], as shown in Table 1. Notably, these statements were initially presented to participants following the practice video viewing. To ensure consistency with the subsequent SUS questionnaire, the Likert scale was adjusted to a 5-point scale (1=strongly disagree, 5=strongly agree) for our user study. These statements collectively aim to gauge user experience, preferences, and potential challenges related to FOMO during interaction with 360° videos.

After the live viewing section of the user study, participants were asked to complete timeline navigation tasks, where they located specific objects within the videos they had previously watched by traversing the video's timeline (RQ2). Unlike the segmented videos used earlier, the videos in this phase were used without segmentation. Each participant performed two navigation tasks: one using a conventional 360° video player and another using our proposed 360° video player. During the timeline navigation section, participants were permitted to use the play/pause button and navigate through the video by clicking specific points on the timeline. Upon finding the target object, participants were instructed to immediately click a green button, located where the lock button had been.

Upon completion of the timeline navigation section, participants were asked to complete the final segment of the Qualtrics questionnaire. This section consisted of ten System Usability Scale (SUS) statements [8] designed to assess the usability of a system (RQ3). Participants rated these statements using a 5-point Likert scale, providing feedback for our proposed 360° video player. Further details on the SUS questionnaire are elaborated in Section 4.4.

Q1. I was afraid that I might miss something in the video.

Q2. I was curious about the parts of the video that were not in my field of view.

Q3. I was able to watch the part of the video that I wanted to watch.

Table 1: FOMO-related statements inspired from Yamaguchi et al.'s work on panoramic thumbnails for 360° videos in VR [43].

4.3 Video choice

The 360° videos used in our study are based on Lin et al.'s classification of 360° videos [24], which categorizes videos based on distinct types of ROIs. This framework was further utilized by Yamaguchi et al. in their work on panoramic thumbnails [43]. Therefore, we opted to use the same database as a benchmark for our study.

- **Type A:** Lin et al. [24] characterize Type A videos as featuring a single static ROI. These videos are commonly found in tour videos, where the guide addresses the audience about the current point of interest. As Yamaguchi et al. [43] used their proprietary footage for Type A analysis, which is unavailable, we opted to use a tour video of Paris from Duanmu et al.'s database [11], as it fits the description.
- **Type B:** Lin et al. [24] characterize Type B videos as featuring a single dynamic ROI that moves within the scene. These videos are often encountered in extreme sports scenarios. We selected the same video used by Yamaguchi et al. [43], titled 'Rain or Shine' by Google Spotlight Stories. This video narrates the story of a girl evading a pursuing rain cloud.
- **Type E:** Type E videos encompass characteristics of both Type C and Type D categories, incorporating multiple ROIs. While Type C videos involve multiple static ROIs, Type D videos feature multiple dynamic ones. Following a methodology similar to Yamaguchi et al. [43], we grouped these videos together, depicting scenes with multiple ROIs simultaneously. We used the same video as their study, which displayed numerous Pokémon scattered throughout the scene.

Video	Video Suffix	Start Time	Mid Point	End Time	Description
Practice	yVLfEHXQk08	00:00	N/A	01:12	Clash of Clans video featuring Hog Riders.
Type A	sJxiPiAaB4k	00:06	00:40	01:16	Guided tour of Paris with a static camera.
Type B	QXF7uGfopnY	00:44	01:24	02:04	Narrative of a girl evading a pursuing rainy cloud.
Type E	pHUVS_GrIeM	00:08	00:48	01:28	Video featuring numerous Pokémon scattered in a real environment.

Table 2: Details of the 360° videos used in our user study, including their respective time segments, midpoints (where the videos divided into two halves), and video suffix identifiers from YouTube links identifying the source. Time format: 'MM:SS'.

The sources for these videos are listed in Table 2, which includes their respective start and end points, along with the midpoint indicating where each video was split into two halves for the live viewing section of the experiment. Moreover, the table also includes information regarding the practice video.

4.4 Measuring usability

The System Usability Scale (SUS), introduced by John Brooke in 1996 [8], is a widely used questionnaire for assessing the perceived usability of a system. It consists of 10 statements with a 5-point Likert scale for responses (see Table 3).

Q1. I think that I would like to use this system frequently.

Q2. I found the system unnecessarily complex.

Q3. I thought the system was easy to use.

Q4. I think that I would need the support of a technical person to be able to use this system.

Q5. I found the various functions in this system were well designed.

Q6. I thought there was too much inconsistency in this system.

Q7. I would imagine that most people would learn to use this system very quickly.

Q8. I found the system very cumbersome to use.

Q9. I felt very confident using the system.

Q10. I needed to learn a lot of things before I could get going with this system.

Table 3: The System Usability Scale (SUS) 10-statement questionnaire [8]. Odd-numbered questions are positively framed, while all even-numbered questions are negatively framed.

The participants are asked to provide their level of agreement or disagreement with the statements. The scores are converted and calculated to provide a usability score that ranges from 0 to 100, with 68 being the average score [6, 9]. Scores above the 68 mark indicate an above-average level of usability, while scores below suggest a below-average user experience. "The simplicity of the SUS statements is intentional as it makes the questionnaire easy to understand, consistent and efficient. The individual statements are not supposed to have diagnostic value in themselves or relate to any specific feature of a particular system" [9]. As a result, as proven throughout the decades of its use, SUS questionnaire is very robust and applicable to varying systems, including newly emerging technology [1].

4.5 Data collection

In addition to the questionnaire outlined in Section 4.2, our 360° video player application recorded several metrics to enhance our understanding of participants' interactions with the 360° videos.

4.5.1 *Live viewing.*

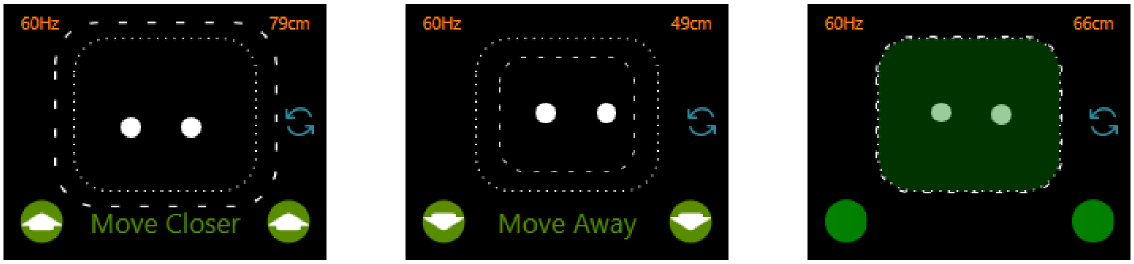
Total effort: This metric quantifies the total rotation (in degrees) that a participant makes while viewing a 360° video, serving as a measure of user engagement and interaction with the content. Comparing the total effort between a conventional and our proposed 360° video player provides valuable insights into whether the dynamic thumbnails can effectively mitigate excessive movement. With the thumbnails displaying the surroundings, participants were expected to be less inclined to actively explore their environment, potentially leading to reduced total effort.

Thumbnail visibility duration: This metric quantifies the duration participants choose to keep the dynamic thumbnails visible. Participants had the freedom to toggle the thumbnails on and off at their discretion using the lock button. Thumbnail visibility duration is expressed as the percentage of total watch time spent with the dynamic thumbnails enabled. Analyzing this metric provides valuable insights into user behavior and the impact of dynamic thumbnail functionality on user engagement.

4.5.2 *Timeline navigation.*

Total effort: For the timeline navigation section, total effort is measured in the same manner as described previously. It quantifies the cumulative rotation performed (in degrees) by a participant during the timeline navigation search task. With thumbnails displaying the surroundings, participants are expected to rotate less as they efficiently scan their environment.

Task completion time: This metric quantifies how quickly a participant can locate a specific object within a previously-watched 360° video (in milliseconds). It provides valuable insights into the efficiency of the participant's navigational behavior and the effectiveness of the interface design in facilitating rapid information retrieval.



(a) (b) (c)

Figure 3: Setup of participants' eye-tracking position during the experiment. The monitor was positioned to ensure participants' eyes were centered on the designated area at the appropriate distance, guided by iMotions feedback (a-c).

4.5.3 Eye-track data.

In addition to the aforementioned metrics, we tracked gaze fixation duration on specific AOIs within the viewport. This duration metric indicates the total amount of time (in milliseconds) that participants spent looking at a specific AOI. This metric was measured for both live viewing and navigation tasks using the iMotions software, which allows us to designate these specific areas. After the experiment, we extracted this data by marking the start and end points of the video within the screen recording. Figure 4 illustrates how the viewport is divided into multiple AOIs.

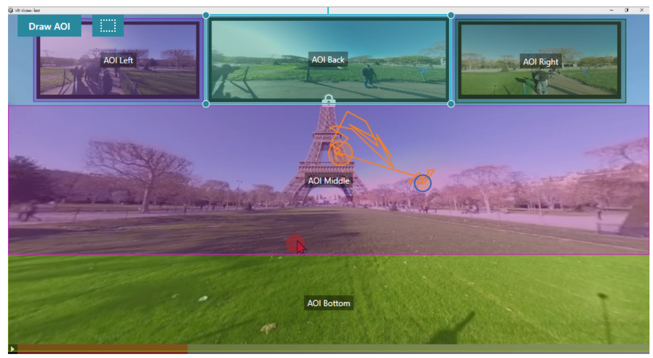


Figure 4: Visualization of the viewport divided into multiple Areas of Interest (AOIs) within iMotions.

4.6 Apparatus

The experiment was conducted using on-site hardware provided by the university, including a designated computer equipped with a mounted and operational eye-tracker. Participants sat in a stationary chair throughout the experiment to ensure a constant distance from the monitor.

4.6.1 Computer specs. The experimental setup utilized a computer equipped with the following specifications:

- Processor: Intel(R) Core(TM) i7-10700K CPU @ 3.80GHz, 8 cores
- RAM: 32GB DDR4 3200 MHz
- Graphics card: NVIDIA Quadro RTX 4000

Additionally, the monitor used for the experiment was an HP E243 with a resolution of 1920x1080 pixels and a refresh rate of 60 Hz.

4.6.2 Eye-tracker. As previously mentioned, participants' behavior was monitored using eye-tracking technology, specifically the AI-X system by Smart Eye. AI-X offers precise measurement and analysis of visual attention. Its sleek design allows it to be discreetly attached to the bottom of the monitor, ensuring that participants are not distracted by its presence and that it does not interfere with the data collection process. Prior to launching the iMotions software, the AI-X itself underwent a calibration. Using Smart Eye's setup tool, we calibrated the AI-X by marking its corners on the desktop monitor to ensure accurate tracking.

We instructed participants to sit comfortably in the stationary chair, as they would normally when using a computer. Following this, we adjusted the positioning of the desktop monitor to ensure that their eyes were centered on the indicated area at the appropriate distance. This procedure helped maintain a consistent viewing experience across all participants, minimizing potential confounding variables related to distance from the monitor and head movement. Additionally, iMotions provided feedback regarding participants' proximity to the monitor, as illustrated in Figure 3, enabling us to make necessary adjustments for optimal eye-tracking accuracy.

4.7 Participants

A total of 17 participants took part in this study, comprising 82.35% male and 17.65% female participants. Participants were not offered any monetary incentives to participate. All participants were university students aged between 21 and 30 years old, with the majority falling within the 21-25 age range (70.59%). The vast majority of participants had little to no prior experience with 360° videos across various devices, except for one participant who reported regular use on both desktop/laptop computers and VR platforms.

Despite the general unfamiliarity with 360° videos, the participants, being young university students, were presumed to be adept at adapting to new technologies. Two participants were excluded from eye-tracking analysis due to pre-existing eye issues. However, this exclusion did not impact the overall study results outside of eye-tracking metrics.

Statement	DT	Type A			Type B			Type E		
		μ	σ	σ^2	μ	σ	σ^2	μ	σ	σ^2
I was afraid that I might miss something in the video.	No	3.35	1.06	1.12	3.47	1.23	1.51	4.41	1.18	1.38
	Yes	1.59	0.62	0.38	1.76	0.83	0.69	2.35	1.11	1.24
I was curious about the parts of the video that were not in my field of view.	No	3.88	1.11	1.24	3.41	1.28	1.63	4.65	0.61	0.37
	Yes	3.24	1.09	1.19	3.06	1.30	1.68	3.53	1.37	1.89
I was able to watch the part of the video that I wanted to watch.	No	3.88	0.78	0.61	4.24	0.75	0.57	3.06	0.90	0.81
	Yes	4.35	0.86	0.74	4.71	0.59	0.35	4.35	0.70	0.49

Table 4: Responses to FOMO-related statements for videos of Type A, B, and E with and without dynamic thumbnails (DT). For each video type, the mean, standard deviation, and variance are presented.

5 RESULTS

In this section, we summarize the results of both the live viewing and the timeline navigation sections of the user study. We also present the findings from the SUS questionnaire and summarize participants' comments on the positive and negative aspects of our proposed 360° video player with dynamic thumbnails. The Shapiro-Wilk test was used to determine the normality of all data groups, ensuring the use of appropriate statistical analyses. All analyses were done using Python[5] (version 3.10).

5.1 Live viewing

To investigate the impact of dynamic thumbnails on FOMO during live viewing section, participants were asked to watch three distinct videos, as described in Section 4.3. Each video was split into two halves at the midpoint, where participants switched to the other 360° video player to continue their viewing experience.

5.1.1 FOMO statements. As mentioned in Section 4.2, participants ranked three FOMO-related statements on a Likert scale upon the completion of each video. The results are shown in Table 4. Given that Likert scale responses cannot be normally distributed, we used the Mann-Whitney U test for pairwise comparisons in the statistical analysis.

For the statement "I was afraid that I might miss something in the video," we performed pairwise comparisons between watching the videos on a conventional player and our proposed 360° video player with dynamic thumbnails for each video type. Significant statistical differences were found for all video types ($U = 259.5$, $p < 0.05$ for Type A; $U = 246$, $p < 0.05$ for Type B; and $U = 252$, $p < 0.05$ for Type E).

For the statement "I was curious about the parts of the video that were not in my field of view," we conducted the same pairwise comparisons. For video Types A and B, our tests indicated no statistically significant differences ($U = 192.5$, $p = 0.091$ and $U = 167$, $p = 0.438$, respectively). However, for Type E, we found a statistically significant difference ($U = 221$, $p = 0.005$).

For the statement "I was able to watch the part of the video that I wanted to watch," we again conducted pairwise comparisons. Our results indicated no statistically significant difference for video Type A ($U = 92.5$, $p = 0.055$). However, we did find statistically significant differences for video Types B and E ($U = 92.5$, $p = 0.043$ and $U = 42$, $p < 0.05$, respectively).

5.1.2 Total effort. We measured the amount of rotation participants exerted with the camera during the live viewing section, referring to it as total effort (see Figure 5). To investigate whether there was a statistically significant difference in total effort between video types, we used the Kruskal-Wallis test. When comparing all video types using a conventional 360° video player, we found a statistically significant difference ($H(2) = 8.284$, $p = 0.016$). However, conducting the same test for videos watched with our proposed 360° video player indicated no statistically significant differences ($H(2) = 4.280$, $p = 0.118$).

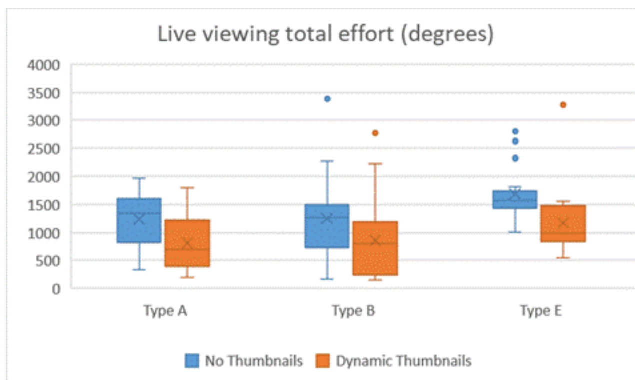


Figure 5: The total effort (total camera rotation) for each video type, with and without dynamic thumbnails, for live viewing section of the user study.

To further analyze the significant differences among video types watched on our proposed 360° video player, we performed a Dunn's test with Bonferroni correction for post hoc analysis. The results indicated no statistically significant difference between videos Type A and Type B ($p = 1$), nor between Type A and Type E ($p = 0.067$). However, we found a statistically significant difference between Type B and Type E ($p = 0.023$).

Afterward, we conducted pairwise comparisons between watching the videos on a conventional player and our proposed 360° video player with dynamic thumbnails to assess the effect for specific

[5] https://www.python.org

video type. For Type A videos, we used an independent samples t-test as both groups were normally distributed. The results showed a statistically significant difference ($t(30) = 2.610$, $p = 0.014$). For Types B and E, we used the Mann-Whitney U test due to the non-normal distribution of the groups. The results indicated no statistically significant difference for Type B ($U = 201$, $p = 0.053$), but a statistically significant difference for Type E ($U = 245$, $p = 0.001$).

5.1.3 Thumbnail visibility duration. Despite being informed of the ability to turn the visibility of thumbnails on and off during the practice video, none of the participants utilized this feature throughout the user study. Therefore, the thumbnail visibility duration for all tests was 100%.

5.1.4 Gaze fixation duration. We tracked participants' gaze fixation duration on specific areas of interest within the viewport (previously depicted in Figure 4) to examine the effect of dynamic thumbnails on their focus during the live viewing section. We specifically analyzed differences in fixation duration at the vertical center of the viewport, as dynamic thumbnails are expected to draw attention away from it. We performed pairwise comparisons on Type A and Type B videos between watching the videos on a conventional player and our proposed 360° video player with dynamic thumbnails, using Mann-Whitney U tests due to non-normal distribution. These comparisons revealed statistically significant differences for both Type A ($U = 204$, $p < 0.05$) and Type B ($U = 213$, $p < 0.05$) videos. For Type E video, which showed normal distribution in both groups, we used an independent samples t-test. This analysis also revealed a statistically significant difference ($t(28) = 8.018$, $p < 0.05$). The differences can be seen in Figure 6 below.

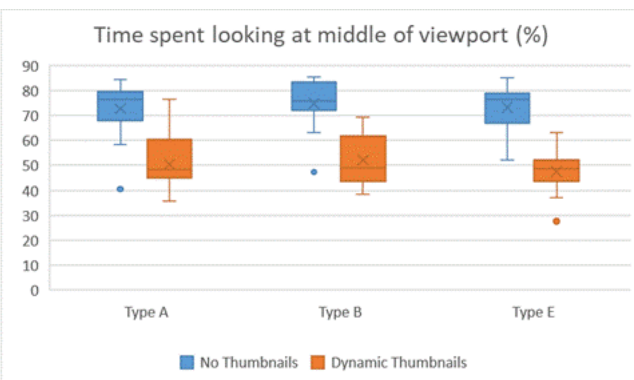


Figure 6: Comparison of participants' gaze fixation duration at the vertical center of the viewport during the live viewing section.

We compared the gaze fixation duration differences between the individual dynamic thumbnails for each video type, as seen in Figure 7. For Type A and Type B videos, which were non-normally distributed, we used the Kruskal-Wallis test, finding statistically significant differences for both Type A ($H(2) = 19.221$, $p < 0.05$) and Type B ($H(2) = 23.545$, $p < 0.05$). Conversely, for Type E videos, which were normally distributed, we used a one-way ANOVA, also revealing statistically significant differences ($F(42) = 59.175$, $p < 0.05$).

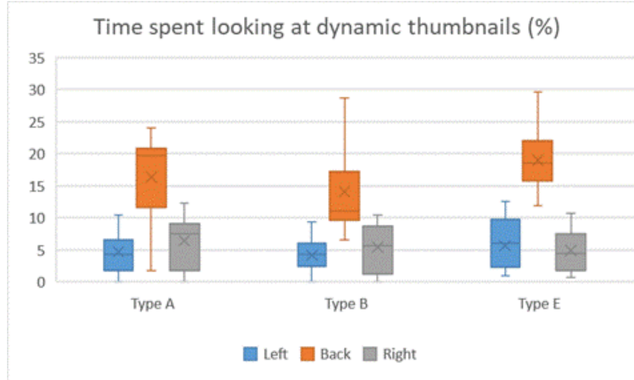


Figure 7: Viewing times of participants for individual dynamic thumbnails for each video type during the live viewing section.

For Type A and Type B videos, we used Dunn's test with Bonferroni correction for post hoc analysis. In both cases, we found statistically significant differences between the top-left side-view and rear-view (both $p < 0.05$), as well as between the top-right side-view and rear-view (both $p < 0.05$). However, no statistically significant difference was found between the top-left and top-right side-views ($p = 0.911$ for Type A and $p = 1$ for Type B). For Type E videos, where we used a one-way ANOVA, Tukey's Honestly Significant Difference (HSD) test was used for post hoc analysis. Similar to Types A and B, we found statistically significant differences between the top-left side-view and rear-view, and between the top-right side-view and rear-view (both $p < 0.05$). However, no statistically significant difference was found between the top-left and top-right side-views ($p = 0.872$).

5.2 Timeline navigation

To investigate the impact of dynamic thumbnails on timeline navigation, participants performed search tasks on videos of Type A and Type B that they had previously watched during the live viewing section of the study. We conducted pairwise comparisons to assess the effects on total effort, task completion time, and gaze fixation duration.

5.2.1 Total effort. Similar to the live viewing section, we measured the amount of rotation participants exerted with their camera during timeline navigation, referring to it as total effort (see Figure 8). An independent samples t-test was conducted to compare the total effort for Type A videos between watching the videos on a conventional player and our proposed 360° video player with dynamic thumbnails. The results indicated no significant difference between the two conditions ($t(15) = -0.907$, $p = 0.379$). A subsequent pairwise comparison using the Mann-Whitney U test, due to non-normal distributions, also indicated no statistically significant difference for Type B videos with and without dynamic thumbnails ($U = 37$, $p = 0.963$).

We then conducted pairwise comparisons between video types. Significant differences were found when comparing Type A videos without thumbnails to Type B videos without thumbnails, as well as Type A videos with dynamic thumbnails to Type B videos with

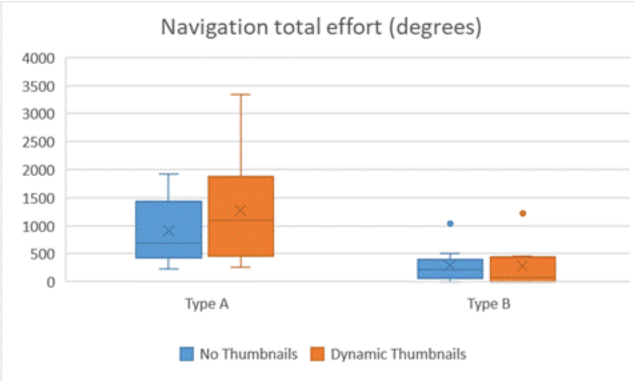


Figure 8: The total effort (total camera rotation) for video Types A and B, with and without dynamic thumbnails, for timeline navigation segment of the user study.

dynamic thumbnails ($U = 9$, $p < 0.05$ and $U = 63$, $p < 0.05$, respectively).

5.2.2 Task completion time. The results for task completion time are shown in Figure 9. Mann-Whitney U tests were used for pairwise comparisons of Type A videos between watching the videos on a conventional player and our proposed 360° video player with dynamic thumbnails, as well as for Type B videos under the same conditions. Both comparisons indicated no statistically significant differences ($U = 17$, $p = 0.075$ and $U = 35$, $p = 0.963$, respectively).

Next, pairwise comparisons were conducted between the two video types. An independent samples t-test compared Type A and Type B videos without dynamic thumbnails, indicating no statistically significant differences ($t(15) = 1.283$, $p = 0.219$). However, when comparing Type A and Type B videos with dynamic thumbnails using the Mann-Whitney U test, a significant difference was found ($U = 62$, $p = 0.01$).

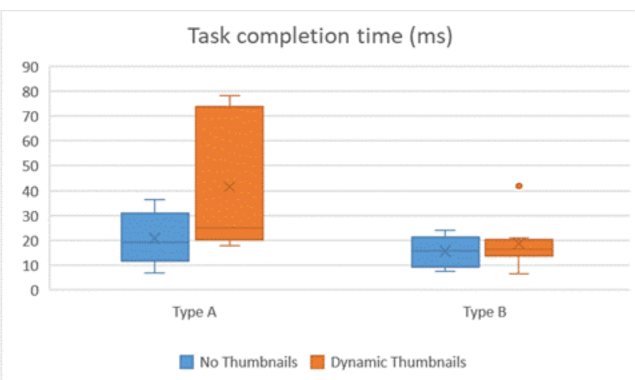


Figure 9: Task completion time for timeline navigation of Type A and Type B videos, with and without dynamic thumbnails.

5.2.3 Gaze fixation duration. Similar to the live viewing section, we tracked participants' gaze fixation duration on the vertical center part of the viewport to examine the effect of dynamic thumbnails on their focus. Independent samples t-tests were used for pairwise comparisons. The comparisons of Type A video between watching on a conventional player and our proposed 360° video player with dynamic thumbnails, as well as Type B video under the same conditions, indicated no statistically significant differences ($t(15) = 0.250$, $p = 0.807$ and $t(15) = 0.611$, $p = 0.552$, respectively).

Next, pairwise comparisons were conducted between the two video types. Comparing Type A with Type B videos without thumbnails, as well as Type A with Type B videos with dynamic thumbnails, also indicated no statistically significant differences ($t(15) = 1.147$, $p = 0.272$ and $t(15) = 1.042$, $p = 0.316$, respectively). The differences can be seen in Figure 10.

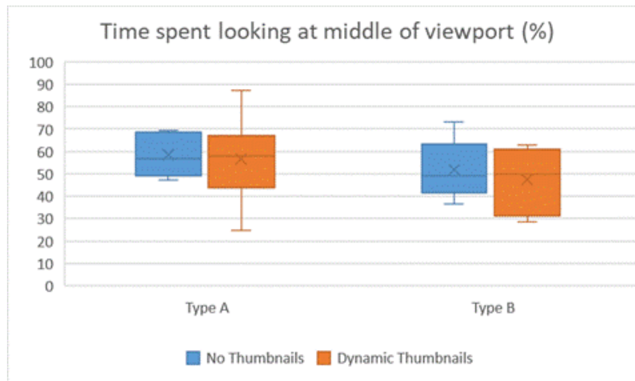


Figure 10: Comparison of participants' fixation at the vertical center of the viewport during timeline navigation.

5.3 Usability

To measure usability, we asked each participant to complete a System Usability Scale questionnaire within our Qualtrics survey. In the SUS, previously shown in Table 3, all odd-numbered questions are positively framed, while all even-numbered questions are negatively framed. Therefore, to calculate the SUS score, a specific formula is used:

$$SUS\ Score = \left(\sum_{i=1,3,5,7,9} (Q_i - 1) + \sum_{j=2,4,6,8,10} (5 - Q_j) \right) \times 2.5 \quad (1)$$

Where Q_i represents the responses to the odd-numbered questions, while Q_j represents the responses to the even-numbered questions. After gathering participants' responses, we computed the individual SUS score for each participant. The average SUS score across all participants was found to be 84.71 ($\sigma = 5.58$). This value serves as a representative measure of usability satisfaction for our proposed 360° video player.

5.4 Participant comments

5.4.1 Positive aspects. The participant comments on the positive aspects of our proposed 360° video player can be categorized into several key points:

- **Enhanced situational awareness**: The dynamic thumbnails allow users to view all the action across different locations simultaneously, reducing the need to focus on a single focal point. Participants found this feature particularly useful when action starts suddenly or when the scene changes. They noted that thumbnails help in quickly identifying salient aspects of the scene, such as a truck approaching video of Type B or a Pokémon appearing in video of Type E, thereby making it clear when something interesting is happening.
- **Improved navigation and comprehension**: The dynamic thumbnails offer a quick way to get an overview of the entire surroundings, making it easier to find points of interest off-screen and understand the situation at a glance. Participants highlighted that thumbnails were especially helpful during the virtual tour, as they provide a quick reference to what is happening outside the main view.
- **Reduced need for camera movement**: By providing a comprehensive view, thumbnails minimize the need for continuous camera rotation, allowing participants to remain oriented and make informed decisions about where to direct their attention next. This was likened to the utility of car mirrors, which provide essential information about the surroundings without requiring constant head movement.

5.4.2 Negative aspects. However, participants also identified several drawbacks to using dynamic thumbnails:

- **Obstruction of screen space**: The dynamic thumbnails occupy screen space, which some participants found obtrusive, particularly in scenes with limited action. This obstruction was seen as diminishing the immersive experience of 360° videos by blocking part of the main view. Specifically, participants noted in video of Type B, where a pursuing cloud appeared in the sky, it occupied the same space as the rear-view thumbnail. This caused participants to either tilt their cameras up or miss seeing the cloud completely.
- **Confusion and disorientation**: Some participants found it easy to get confused about which thumbnail corresponded to which view, leading to occasional disorientation.
- **Perceived redundancy and overstimulation**: In scenes with little happening simultaneously, thumbnails were perceived as unnecessary and sometimes overstimulating. Conversely, in scenes with a lot of activity, managing both the main view and multiple thumbnails was challenging, causing some users to feel they were missing important details.
- **Limited usefulness in certain scenarios**: Participants noted that in scenarios without off-screen points of interest, the thumbnails added little to the experience and could be considered redundant. The side-view thumbnails, in particular, were often deemed less useful compared to the main screen and the rear-view thumbnail.

6 DISCUSSION

6.1 Fear of missing out mitigation

To answer Research Question 1: *To what extent does the proposed interface mitigate users' fear of missing out during the live viewing of 360° videos, as compared to conventional 360° video players?* we look at the results from Section 5.1 and participant comments summarized in Section 5.4.

Our findings indicate that across all video types (Type A, Type B, Type E), the dynamic thumbnails in our proposed 360° video player mitigate FOMO to varying degrees. Positive results were observed for each of the three Likert scale FOMO statements when comparing our proposed 360° video player to a conventional one, with the primary statement, 'I was afraid that I might miss something in the video,' showing significant improvement across all video types. Participant comments echoed these findings, many of which highlighted that the thumbnails allowed them to observe all action in the video and quickly gain awareness of surroundings beyond the main view. Even participants who initially felt the thumbnails were unnecessary appreciated the assurance that they were not missing out on any important events.

These results are further supported by the total effort measure, which measured the amount of camera rotation participants engaged in. Participants rotated their cameras significantly less when using dynamic thumbnails. Although the p-value for Type B videos was not statistically significant, it was very close ($p = 0.054$). This may be attributed to the nature of Type B videos, which typically focus on a single point of interest, resulting in less camera rotation even with conventional video players.

We used gaze fixation duration to assess participants' focus on the vertical center of the viewport while watching 360° videos. Our results indicated that for each video type, participants looked significantly less at the vertical center of the viewport when using our proposed 360° video player, suggesting increased attention towards the dynamic thumbnails. This behavior implies that dynamic thumbnails effectively capture user attention and assist in mitigating FOMO by offering a comprehensive view of the action.

6.2 Enhanced user timeline navigation

To answer Research Question 2: *In what ways does the proposed interface affect user navigation (in terms of task completion time and total rotation exerted) through a video's timeline compared to conventional 360° video players?* we look at the results from Section 5.2.

The analysis of the total effort measure, which measured the amount of camera rotation participants engaged in, revealed no significant difference between our proposed 360° video player featuring dynamic thumbnails and a conventional one. This consistency was observed across both Type A and Type B videos, suggesting that the presence of dynamic thumbnails does not influence the overall rotational effort exerted by users. One possible explanation for this observation is that users may disregard the dynamic thumbnails during timeline navigation search tasks, focusing solely on the main screen rather than processing multiple focal points simultaneously. This is supported by the analysis of gaze fixation duration, which showed no difference in attention towards the vertical center of the viewport during navigation tasks.

The analysis of task completion time also revealed no significant difference when comparing our proposed 360° video player to a conventional one. However, it is noteworthy that the p-value for Type A video comparisons was very close to indicating a statistically significant difference. Given our limited sample size, this result could potentially change with a larger sample. Therefore, further investigation focusing on this aspect may still be worthwhile.

Moreover, a notable observation arises when comparing both the total effort and task completion time between the two video types. This variability suggests that factors beyond the interface design of the video player, such as the content nature and specific characteristics of the search task, likely play a significant role in influencing user navigation behavior. Additional research is needed to isolate these factors and obtain a more comprehensive understanding of their individual impacts on user timeline navigation effort.

6.3 Perceived usability

To answer Research Question 3: *How do users perceive the usability of the proposed interface in contrast to conventional 360° video players?* we look at the results from the SUS questionnaire in Section 5.3 as well as participant comments summarized in Section 5.4.

As mentioned previously in Section 4.4 when introducing the SUS, the SUS score ranges from 0 to 100, with 68 considered the average score and any score above it deemed acceptable [6, 9]. The SUS score for our proposed 360° video player with dynamic thumbnails was 84.71 ($\sigma = 5.58$), calculated as the average of all participant scores. According to Bangor et al.'s interpretation of the SUS scale based on adjectives, this score falls into the 'Good' category, just short of 'Excellent,' which requires a score of 85 [6].

However, even though the SUS is a great indicator of the usability of a system, it lacks context. Therefore, we also look towards participant comments to gain some more information. The participants appreciated how the dynamic thumbnails allowed for a quick overview of the entire 360° scene without having to move the camera around. It enhanced their situational awareness, as anything coming from behind was easily spotted by the dynamic thumbnails. Overall, they had no issues utilizing the dynamic thumbnails as intended.

Despite this, participants were able to point out some drawbacks of the dynamic thumbnails. The most obvious issue was that the dynamic thumbnails would sometimes obstruct points of interest. This was mainly noted in the video of Type B, where the black cloud pursuing the little girl would appear in the sky right above her, where the rear-view dynamic thumbnail would reside when looking directly at the girl. Another issue was that in scenes with little activity, participants found the thumbnails unnecessary. We anticipated these issues and addressed them by including a button that allows participants to toggle the dynamic thumbnails on and off based on their needs, as explained in Section 3.1. However, as results revealed in Section 5.1, participants did not interact with this feature despite being made aware of it. Therefore, it stands to reason that a more obvious or efficient way of toggling dynamic thumbnails would be desirable in the future.

Participants also frequently mentioned that they rarely used the side-view thumbnails, which is supported by the gaze fixation findings in Section 5.1. Several factors might explain this behavior. First, the effort required to look at the sides is much less than what is needed to look at the opposite side of the video. Therefore, the rear-view thumbnail might be more intuitive to check, while users might prefer to rotate the camera slightly to check the sides. Second, participants tend to center their focus on the veritcal center of the viewport for most of the time while watching a 360° video as shown by Duanmu et al.'s work [11] and also now supported by our gaze fixation duration findings. This makes the rear-view thumbnail, positioned closer to their peripheral vision, more noticeable. In contrast, the side-view thumbnails are farther from the center, demanding more effort to focus on them.

6.4 Limitations of the study

While this study provides valuable insights into the design and usability of a new 360° video player with dynamic thumbnails, several limitations should be acknowledged.

Firstly, the study involved a relatively small number of participants, which may limit the generalizability of the findings. The time-intensive nature of each session contributed to this limitation, as the detailed procedures required significant time per participant, thereby restricting the total number of participants.

Additionally, due to time constraints, we limited the number of videos participants watched. For the live viewing section of our user study, we covered Type A, Type B, and Type E videos as defined by Lin et al. [24]. Although we categorized Type C and Type D within the Type E category, future studies could benefit from examining them separately. Furthermore, we restricted our timeline navigation section to Type A and Type B videos to simplify counterbalancing and session durations. Including Type E videos would have complicated the order and extended each session further. Future research should consider including all video types in both sections of the user study.

Secondly, participants had a relatively short acclimatization period with our proposed video player. Despite using a short practice video to introduce them to the 360° video player design, a longer period might have provided a more accurate assessment of the usability and effectiveness of features such as dynamic thumbnails and the lock button. Additionally, exploring the impact of dynamic thumbnails in longer video segments, beyond the one minute clips used in this study, could reveal whether users become more inclined to use the lock button over time.

Future research should also investigate the use of dynamic thumbnails over an extended period to observe how user interactions evolve once the novelty wears off.

Finally, comparing our interface with existing systems, such as Lin et al.'s *Outside-In* [24] or considering Kang et al.'s hybrid system [19] in combination with ours, could have been beneficial for a more comprehensive evaluation. Such comparisons could have highlighted specific strengths and weaknesses of our proposed design relative to current standards in 360° video player technology.

7 CONCLUSION

This study aimed to propose a new 360° video player for 2D displays to enhance the viewer experience. We designed an interface that uses dynamic thumbnails to reveal the majority of the 360° scene to the user and compared it to a conventional 360° video player. To evaluate the effectiveness of our design, we conducted live viewing tasks where users watched 360° videos as they normally would, and timeline navigation tasks that involved a search component. The two video players were evaluated based on the feeling of FOMO experienced by users, the amount of camera movement, and the time taken for the timeline navigation task. Additionally, participants completed a SUS questionnaire to assess the usability of our design.

Results show that users experienced significantly less FOMO during live viewing when using our proposed 360° video player due to better situational awareness of the entire scene. However, there were no statistically significant differences in the time taken to complete the timeline navigation task between the two video players, and users' gaze fixations revealed similar patterns for both players. The results of the SUS questionnaire indicated that the perceived usability of our proposed interface was rated as 'Good,' just short of 'Excellent.'

These findings suggest that dynamic thumbnails can effectively reduce FOMO in 360° video experiences on 2D displays, potentially leading to more satisfying and engaging viewing sessions. While the navigation efficiency was comparable between the two interfaces, the positive usability ratings for our design indicate its potential for wider adoption.

Future research could explore the long-term effects of using dynamic thumbnails on user engagement and satisfaction, as well as their impact in different contexts such as education, entertainment, and virtual tours. Additionally, incorporating user feedback to further refine the interface could yield even more significant improvements in user experience.

REFERENCES

All references can be found at the end of this document.

A IMPLEMENTATION

Our proposed 360° video player interface was developed using the Unity[6] game engine (version 2021.3.17f1). After a thorough review of existing open-source Unity 360° video players, we determined that they did not meet the requirements for this project. The reasons for this conclusion include:

- **Outdated software:** Many of the available open-source 360° video players are no longer maintained and lack support for the latest Unity versions.
- **VR-specific features:** Several of these players are tailored specifically for VR applications, which introduces unnecessary complexity for our use case.
- **Complexity:** The source code of the available open-source 360° video players is overly intricate, resulting in a steep learning curve. It would take more time to understand and modify this existing code than to develop the required features from scratch.

Given these considerations, we decided to build a new 360° video player from the ground up. Unity's `VideoPlayer` API offers a convenient and efficient way to implement the essential features of a 360° video player, thereby simplifying the development process.

A.1 Final design

The figure below shows the final version of our proposed 360° video player. The temporal navigation widgets are located at the bottom of the video player, while the spatial navigation elements, including our dynamic thumbnails, are positioned at the top.

The temporal navigation widgets consist of a conventional timeline for scrolling through video frames and a play/pause button. The spatial navigation includes dynamic thumbnails and a lock button used for controlling the visibility of these thumbnails. The main view has a 95° field of view (FOV). The dynamic side-view thumbnail on the left displays a 65° FOV of the area to the left of the main view, while the side-view thumbnail on the right shows a 65° FOV of the area to the right. The rear-view thumbnail in the center presents a 95° FOV of the area directly behind the main view. Notably, the projection on the rear-view thumbnail is horizontally flipped to mirror what is behind the main view. The rationale for these specific FOVs is discussed in Section 3.3 of the scientific paper.

The lock button is located at the bottom of the rear-view thumbnail. This button enables and disables the thumbnails' visibility when not hovering over the top part of the 360° video player. We positioned it at the bottom of the rear-view thumbnail to ensure it is easily accessible and convenient while remaining unobtrusive. Placing it at the bottom of the 360° video player was avoided to maintain a consistent separation between spatial and temporal navigation.

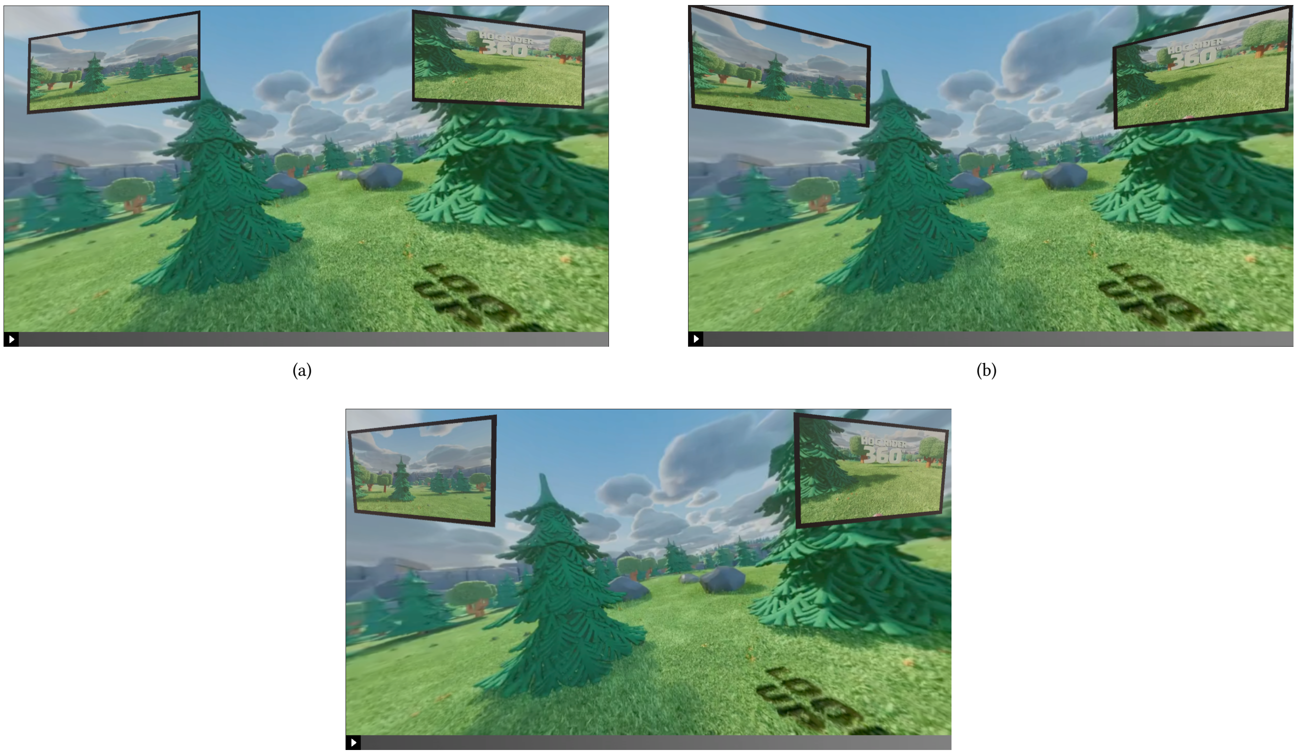
Figure 11: Final version of our proposed 360° video player.

[6]https://unity.com/

A.2 Alternate designs

Due to the way we implemented our proposed 360° video player (by projecting the source 360° video onto a skybox) we experimented with dynamic thumbnails appearing in 3D. Figure 12 below shows some of the design attempts we made for the side-view thumbnails. However, as seen in the figures, tilting the thumbnails created distortions, causing the sides to either shrink or expand. Additionally, the optical flow within the thumbnails did not always align with the way the thumbnails were tilted, which made the viewing experience confusing.

After much deliberation, we decided that for a 360° video player intended for 2D displays, flat 2D thumbnails made the most sense, as seen in the closeup in Figure 13. If we were to implement dynamic thumbnails in VR, revisiting this concept could be interesting.



(a) (b) (c)

Figure 12: Design attempts for side-view thumbnails for our proposed 360° video player.

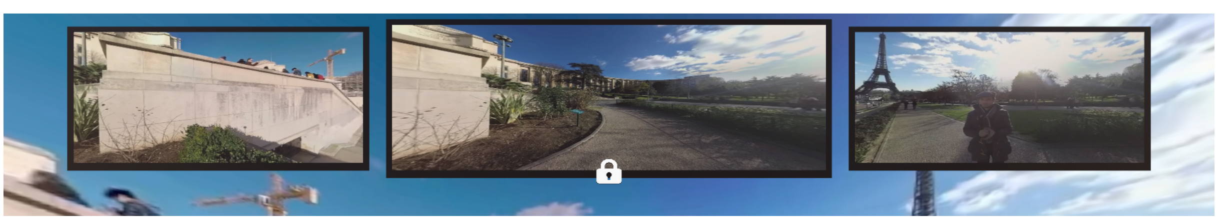


Figure 13: Closeup of the flat 2D thumbnails implemented for our proposed 360° video player.

A.3 Video editing

When downloading videos from YouTube, we discovered that they often use a non-standardized format known as equiangular projection. Since this projection is incompatible with our implementation in Unity, which expects equirectangular projection, we used FFmpeg[7] to convert between these two formats. The conversion was performed using the following command line:

```
ffmpeg -i input.mp4 -vf "v360=eac:equirect" output.mp4
```

Here, 'input.mp4' is the name of the original video file and 'output.mp4' is the name for the converted video file.

Additionally, since we used only specific segments of videos during our user studies, as detailed in Section 4, we required a tool to cut these videos into smaller segments. Online video cutters often caused issues such as corrupting the videos, such as starting with a black screen at the cut point. Consequently, we relied on FFmpeg once again, using the following command line:

```
ffmpeg -i input.mp4 -ss 00:00:03 -t 00:00:08 -async 1 output.mp4
```

Here, 'input.mp4' is the original video file, '-ss 00:00:03' specifies the starting point (3 seconds into the video), '-t 00:00:08' specifies the duration (8 seconds), and output.mp4 is the name of the output video file.

[7] https://ffmpeg.org/

B USER STUDY DESIGN

This appendix provides additional details regarding our user study that were not covered in the scientific paper. These include an overview of the experimental setup, the specific tasks used for the timeline navigation section of the user study, and our counterbalancing method.

B.1 Flowchart

The flowchart below illustrates the entire flow of the user study for each participant. The 360° video player used by participants for both live viewing and timeline navigation tasks varied for each participant and is further detailed in Section B.3, where a Latin square design was used to counterbalance the order of tasks.

At the beginning of the user study, participants were provided with an information sheet and a consent form to review and sign. This is followed by basic demographic questions. Participants were then given time to familiarize themselves with a practice video that introduces our proposed 360° video interface design. Following the practice video, participants were introduced to FOMO related questions in the questionnaire, which were asked after each video type. This step aims to minimize any potential bias for the first live viewing task. As shown in the flowchart, participants alternated between tasks and the Qualtrics questionnaire during the live viewing section of the experiment. After completing both the live viewing and timeline navigation sections, participants were asked to fill out a System Usability Scale questionnaire. Finally, participants were invited to provide general comments on how they felt about our proposed design.

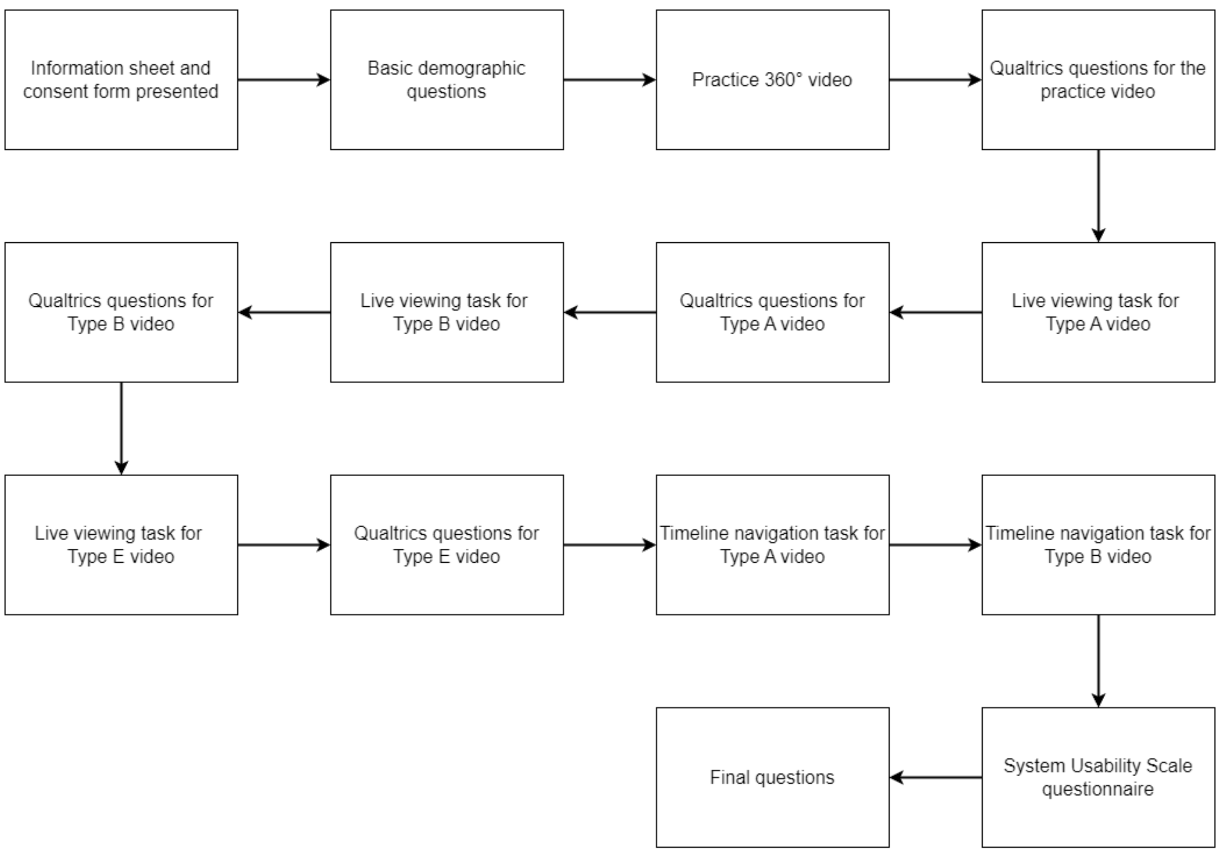
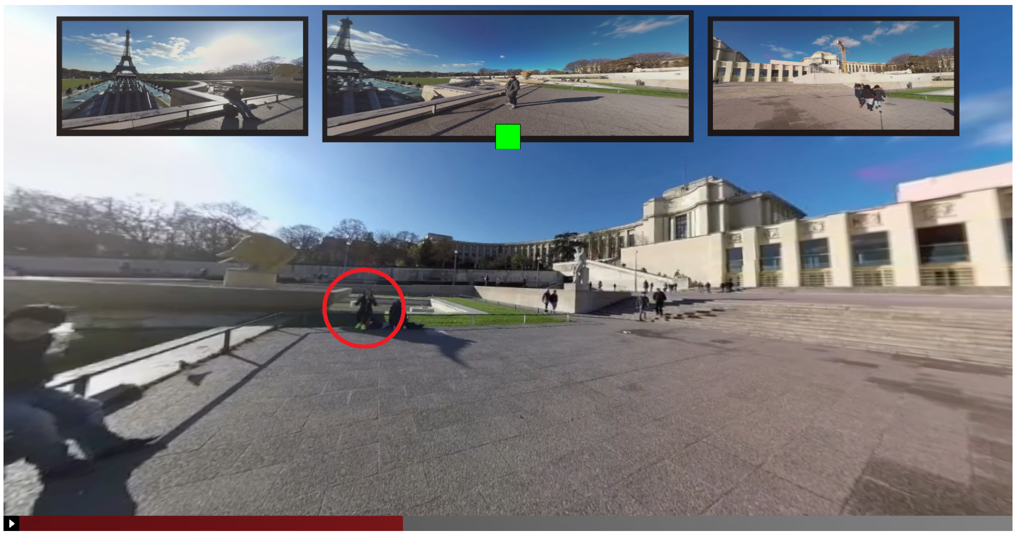


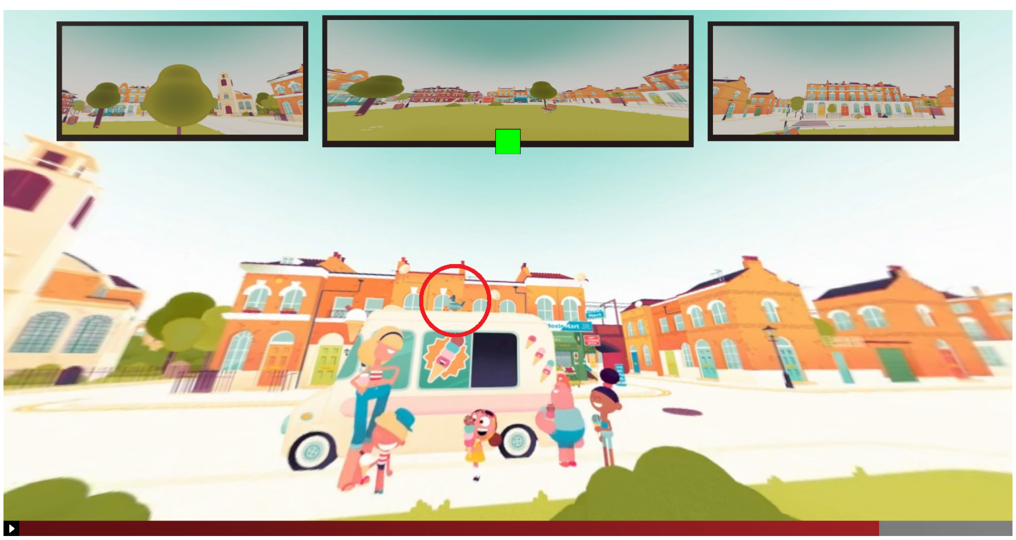
Figure 14: Flowchart illustrating the sequence of tasks in our user study.

B.2 Timeline navigation tasks

This appendix presents the two search tasks for our timeline navigation section of the user study. As shown in the figures below, the lock button used during live viewing tasks was replaced with a green square button. This button was used by participants to click immediately after they found what the task asked them to locate. For Type A video, the task was to find a person wearing yellow shoes, and for Type B video, the task was to find a bird sitting on top of an ice cream van, as highlighted with a red circle in Figure 15a and Figure 15b, respectively.



(a) Search task for Type A video



(b) Search task for Type B video

Figure 15: Search tasks in the timeline navigation section of our user study.

B.3 Latin square design counterbalancing

This appendix presents two tables indicating the order in which each participant watched the videos and the corresponding viewing methods used. A Latin square design was used for counterbalancing.

The Latin square design systematically allocated participants to different video-viewing conditions, ensuring that each video was viewed with each method an equal number of times across the sample. By counterbalancing the order, the study aims to enhance the validity and generalizability of its findings by minimizing order effects and other potential sources of bias.

Live viewing

During the live viewing task, participants were instructed to watch three distinct types of videos: Type A, Type B, and Type E, as described in Section 4.3 of the scientific paper. Each video was divided into two halves at the midpoint, and participants were asked to view each half using a different 360° video player: a conventional 360° video player with no thumbnails and our proposed 360° video player with dynamic thumbnails.

To enhance the validity and generalizability of the study, the order in which participants experienced these videos was carefully balanced. In the table below, participants' viewing sequences are depicted. An orange cell indicates that the first half of the video was viewed with the conventional 360° video player, while a green cell indicates that the first half was viewed with our proposed 360° video player.

Participant	Type A	Type B	Type E
1	No Thumbnails First	With Thumbnails First	No Thumbnails First
2	With Thumbnails First	No Thumbnails First	With Thumbnails First
3	No Thumbnails First	With Thumbnails First	No Thumbnails First
4	With Thumbnails First	No Thumbnails First	With Thumbnails First
5	No Thumbnails First	With Thumbnails First	No Thumbnails First
6	With Thumbnails First	No Thumbnails First	With Thumbnails First
7	No Thumbnails First	With Thumbnails First	No Thumbnails First
8	With Thumbnails First	No Thumbnails First	With Thumbnails First
9	No Thumbnails First	With Thumbnails First	No Thumbnails First
10	With Thumbnails First	No Thumbnails First	With Thumbnails First
11	No Thumbnails First	With Thumbnails First	No Thumbnails First
12	With Thumbnails First	No Thumbnails First	With Thumbnails First
13	No Thumbnails First	With Thumbnails First	No Thumbnails First
14	With Thumbnails First	No Thumbnails First	With Thumbnails First
15	No Thumbnails First	With Thumbnails First	No Thumbnails First
16	With Thumbnails First	No Thumbnails First	With Thumbnails First
17	No Thumbnails First	With Thumbnails First	No Thumbnails First
18	With Thumbnails First	No Thumbnails First	With Thumbnails First
19	No Thumbnails First	With Thumbnails First	No Thumbnails First
20	With Thumbnails First	No Thumbnails First	With Thumbnails First
21	No Thumbnails First	With Thumbnails First	No Thumbnails First
22	With Thumbnails First	No Thumbnails First	With Thumbnails First
23	No Thumbnails First	With Thumbnails First	No Thumbnails First
24	With Thumbnails First	No Thumbnails First	With Thumbnails First

Table 5: Latin square design for 24 participants for live viewing tasks.

Timeline navigation

During the timeline navigation tasks, participants were tasked with finding a specific object within two video types: Type A and Type B. Once again, two different 360° video players were used: a conventional 360° video player without thumbnails and our proposed 360° video player with dynamic thumbnails.

Type E video was excluded from this part of the experiment because we did not use a third 360° video player for comparison. Including Type E would have made balancing awkward with only two video players but three video types. Additionally, due to time constraints in the user study, we did not want to introduce more tasks. Type A and B videos essentially encompass what Type E video represents, just split apart. In future work, incorporating Type E video as a point of comparison would be beneficial.

To enhance the validity and generalizability of the study, the order in which participants experienced these videos was carefully balanced. In the table below, participants' viewing sequences are depicted. An orange cell indicates that the navigation task was done using the conventional 360° video player, while a green cell indicates that the navigation was done with our proposed 360° video player.

Participant	Type A	Type B
1	With Thumbnails	No Thumbnails
2	No Thumbnails	With Thumbnails
3	With Thumbnails	No Thumbnails
4	No Thumbnails	With Thumbnails
5	With Thumbnails	No Thumbnails
6	No Thumbnails	With Thumbnails
7	With Thumbnails	No Thumbnails
8	No Thumbnails	With Thumbnails
9	With Thumbnails	No Thumbnails
10	No Thumbnails	With Thumbnails
11	With Thumbnails	No Thumbnails
12	No Thumbnails	With Thumbnails
13	With Thumbnails	No Thumbnails
14	No Thumbnails	With Thumbnails
15	With Thumbnails	No Thumbnails
16	No Thumbnails	With Thumbnails
17	With Thumbnails	No Thumbnails
18	No Thumbnails	With Thumbnails
19	With Thumbnails	No Thumbnails
20	No Thumbnails	With Thumbnails
21	With Thumbnails	No Thumbnails
22	No Thumbnails	With Thumbnails
23	With Thumbnails	No Thumbnails
24	No Thumbnails	With Thumbnails

Table 6: Latin square design for 24 participants for navigation tasks.

C CONSENT FORM AND QUESTIONNAIRE

This appendix contains the consent form and questionnaire used in our user study. As detailed in the scientific paper, participants received the consent form digitally before arriving at the experiment to ensure they understood what was expected of them. The same consent form was presented again at the start of the study through the Qualtrics questionnaire, giving participants another opportunity to read and consent to it.

The Qualtrics questionnaire was designed to align with the iMotions workflow, facilitating a presentation-style progression between tasks. Within the questionnaire, lines stating "Proceed to the next task." signaled participants to switch from the questionnaire on a laptop to a separate computer running the iMotions experiment. Similarly, slides in the iMotions flow directed participants back to the questionnaire.

A copy of our Qualtrics questionnaire starts on the next page.

Participant Number and Consent Form

Participant Number (filled in by the researcher)

Title of the research: Watch your back!: Dynamic thumbnails for a 360° video player to enhance viewing experience on 2D displays

The goal of the research: This research, conducted by Jakub Kováč, a student at Utrecht Univeristy, aims to investigate the use of dynamic thumbnails in 360° video players for 2D displays. It achieves this by evaluating users' experience of FOMO (fear of missing out) during live viewing of a 360° video, as well as their performance in subsequent tasks involving locating specific points within the video using the video player's timeline.

Potential risks: There are no legal or financial risks associated with participating in this study. You are under no obligation to answer any questions you do not wish to answer. Your participation is entirely voluntary, and you may withdraw from the study at any time. If you choose to withdraw, your data and responses will be promptly deleted and will not be used for the study.

Confidentiality of data: These measures ensure your privacy and the confidentiality of your data throughout the research process.

- No confidential information or personal data will be disclosed or publicized in any way that can be traced back to you.
- We will only ask for information relevant to this research, and no extraneous data will be collected.

- The results of the experiment will be securely stored and documented on a password-protected university computer. Subsequently, they will be transferred to a password-protected laptop accessible only to the researcher via a USB stick. Once the data transfer is complete, it will be promptly deleted from the USB stick to minimize risk.
- Should you wish to review the data we have saved about you, you may request it by emailing the researcher. However, this option will no longer be available once all data has been anonymized. All original data will be destroyed following anonymization.

Voluntary concerns: Your participation in this research is entirely voluntary, and you will not receive any monetary compensation. You have the freedom to withdraw at any time, without providing a reason. In such cases, your results and information will be promptly deleted and will not be used for the study. Opting out or choosing not to participate will have no negative consequences for you whatsoever. However, if you do decide to participate, rest assured that all your data will be handled with the utmost care and protection, as outlined above. If you have any concerns, complaints, or wish to contact the researcher, please reach out using the following information:

Researcher: Jakub Kováč, Email: j.kovac@students.uu.nl

Declaration of consent:

- I understand that by signing this declaration, I am consenting to participate in this study with full knowledge of its purpose, data collection methods, storage, processing procedures, and any associated risks.
- I confirm that I have been adequately informed about the study and understand its objectives. Any questions I had were addressed satisfactorily before signing this form.
- I acknowledge that my participation in this study is voluntary, and there is no pressure, explicit or implicit, to participate.

- I retain the right to refrain from answering certain questions and have the right to discontinue my participation if I feel uncomfortable at any point during the experiment.
- The researcher has explicitly assured me that my data will be anonymized, and I will not be identifiable in any published materials.
- I have carefully read and understood the contents of this form. All my inquiries were addressed to my satisfaction, and I am participating voluntarily.

Do you consent to the terms described above?

- ◯ No, I do not consent
- ◯ Yes, I do consent

How old are you?

- ◯ 18 - 20 years old
- ◯ 21 - 25 years old
- ◯ 26 - 30 years old
- ◯ 31 - 35 years old
- ◯ 36+
- ◯ Prefer not to say

What is your gender?

- ◯ Male
- ◯ Female
- ◯ Non-binary / third gender
- ◯ Prefer not to say

Are you wearing glasses?

- ◯ Yes
- ◯ No

360° videos, also referred to as omnidirectional or panoramic videos, have gained widespread popularity in recent years. They are captured using specialized omnidirectional cameras that, unlike traditional cameras with limited fields of view, are capable of recording from all directions.

Indicate how often you have used each of the following devices for 360° video viewing.

	Never used it	Used it a couple times	Use it quite a bit	Use it regularly	Use it very often
laptop or desktop computer	◯	◯	◯	◯	◯
mobile phone / tablet	◯	◯	◯	◯	◯
virtual reality (VR)	◯	◯	◯	◯	◯

Before proceeding, please start the experiment on the other computer. Throughout the experiment, you will receive instructions on when to complete the next section.

Practice Session

Indicate how you felt watching the practice video with dynamic thumbnails.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I was afraid that I might miss something in the video	○	○	○	○	○
I was curious about the parts of the video that were not in my field of view	○	○	○	○	○
I was able to watch the part of the video that I wanted to watch	○	○	○	○	○

Proceed to the next task.

Main Experiment

Indicate how you felt watching the Paris tour video **with no thumbnails**.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I was afraid that I might miss something in the video	○	○	○	○	○
I was curious about the parts of the video that were not in my field of view	○	○	○	○	○
I was able to watch the part of the video that I wanted to watch	○	○	○	○	○

Indicate how you felt watching the Paris tour video **with dynamic thumbnails**.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I was afraid that I might miss something in the video	○	○	○	○	○

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I was curious about the parts of the video that were not in my field of view	○	○	○	○	○
I was able to watch the part of the video that I wanted to watch	○	○	○	○	○

Proceed to the next task.

Indicate how you felt watching the story about a little girl **with no thumbnails**.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I was afraid that I might miss something in the video	○	○	○	○	○
I was curious about the parts of the video that were not in my field of view	○	○	○	○	○
I was able to watch the part of the video that I wanted to watch	○	○	○	○	○

Indicate how you felt watching the story about a little girl **with dynamic thumbnails**.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I was afraid that I might miss something in the video	○	○	○	○	○
I was curious about the parts of the video that were not in my field of view	○	○	○	○	○

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I was able to watch the part of the video that I wanted to watch	○	○	○	○	○

Proceed to the next task.

Indicate how you felt watching the Pokémon video **with no thumbnails**.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I was afraid that I might miss something in the video	○	○	○	○	○
I was curious about the parts of the video that were not in my field of view	○	○	○	○	○
I was able to watch the part of the video that I wanted to watch	○	○	○	○	○

Indicate how you felt watching the Pokémon video **with dynamic thumbnails**.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I was afraid that I might miss something in the video	○	○	○	○	○
I was curious about the parts of the video that were not in my field of view	○	○	○	○	○
I was able to watch the part of the video that I wanted to watch	○	○	○	○	○

Proceed to the next task.

Final Questions

System Usability Scale

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I think that I would like to use this system frequently.					
I found the system unnecessarily complex.					
I thought the system was easy to use.					
I think that I would need the support of a technical person to be able to use this system.					
I found the various functions in this system were well designed.					
I thought there was too much inconsistency in this system.					
I would imagine that most people would learn to use this system very quickly.					
I found the system very cumbersome to use.					
I felt very confident using the system.					
I needed to learn a lot of things before I could get going with this system.					

Did you find the dynamic thumbnails to be intrusive?

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
Tour video					
Video about a little girl					

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
Pokémon	○	○	○	○	○

Did you find the dynamic thumbnails to be useful?

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
Tour video	○	○	○	○	○
Video about a little girl	○	○	○	○	○
Pokémon	○	○	○	○	○

Can you name some positive aspects/advantages of the dynamic thumbnails?

Can you name some negative aspects/disadvantages of the dynamic thumbnails?

Powered by Qualtrics

D ADDITIONAL RESULTS

This appendix presents supplementary data collected during the experiment, which extends beyond the information provided in Section 5 of the scientific paper.

D.1 General demographics

This section outlines the general demographics of the participants. As illustrated in Figure 16, all participants were students aged between 21 and 30, with the majority falling within the 21-25 age range (70.59%). Figure 17 indicates that the vast majority of participants were male (82.35%). Considering the potential impact of glasses on eye-tracker data, we also inquired about eyewear usage. Figure 18 shows that the number of participants wearing glasses and those not wearing glasses is nearly equal, with a slight majority not wearing glasses (52.94%).

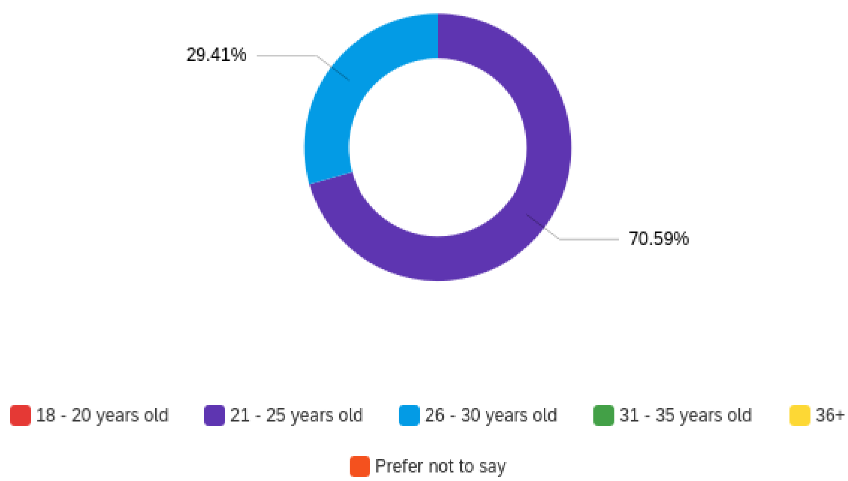


Figure 16: Age distribution of participants.

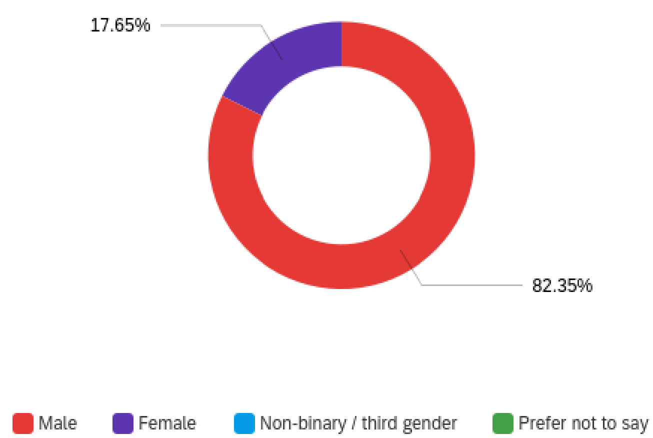


Figure 17: Gender distribution of participants.

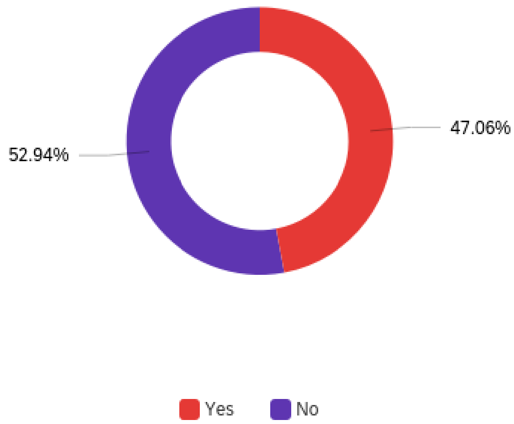


Figure 18: Distribution of eyewear usage among participants.

D.2 Familiarity with 360° videos

The figure below shows the participants' past experience with 360° videos across different devices. As can be seen from the results, the vast majority of participants have very limited or no experience at all. Only one participant reported regularly using 360° videos on a laptop/desktop computer and in VR.

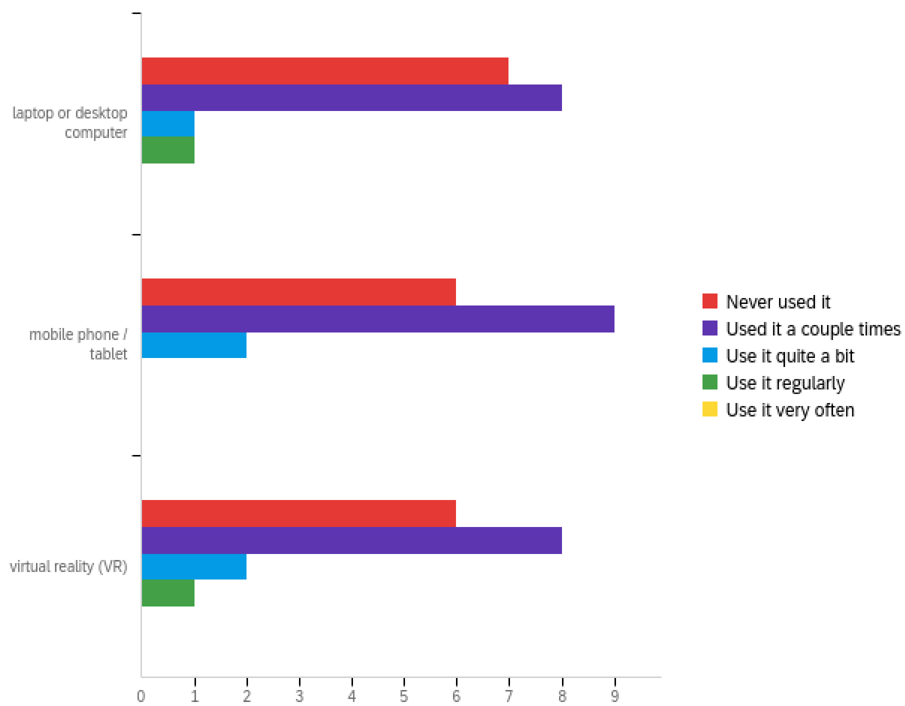


Figure 19: Participants' familiarity with 360° videos across different devices.

D.3 System Usability Scale

This section presents participant responses to the 10 System Usability Scale questions, as shown in the figure below. All odd-numbered questions are positively framed, while all even-numbered questions are negatively framed.

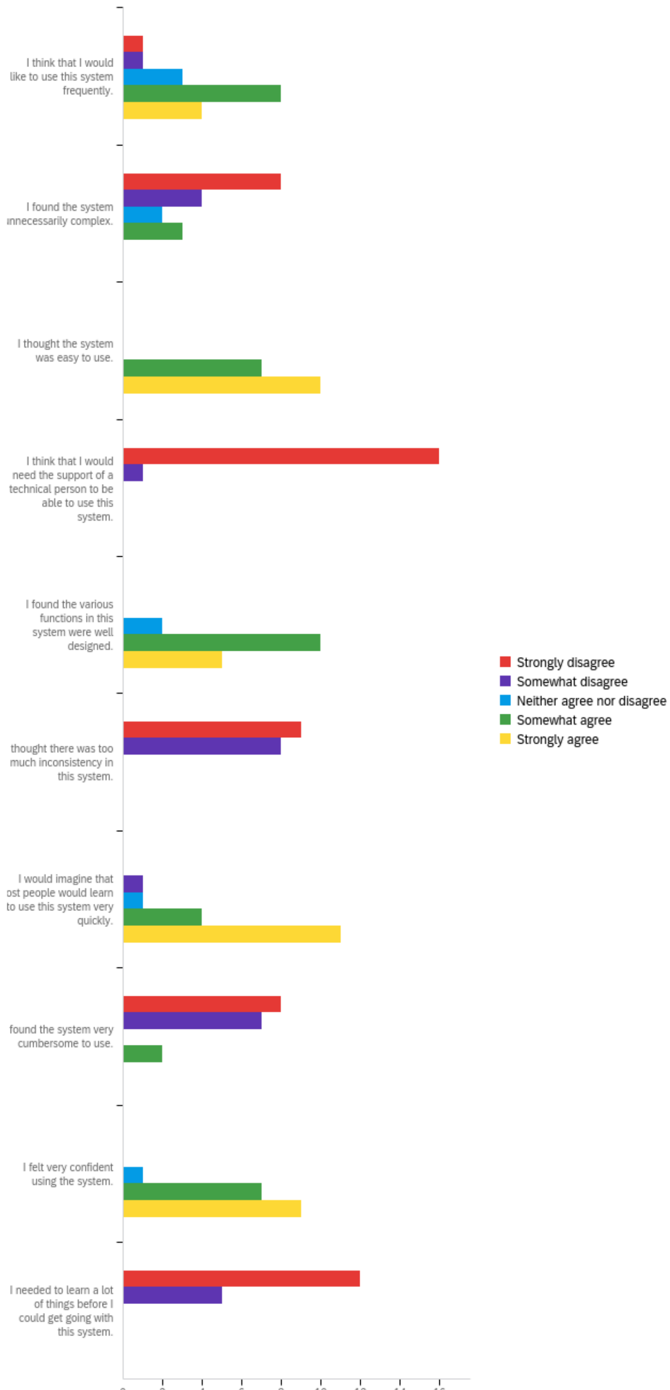


Figure 20: Participant responses to the 10 System Usability Scale questions.

These figures present additional calculations related to Section 5.3 of the scientific paper. Table 7 shows the average SUS score for each statement. Meanwhile, Figure 21 displays the overall SUS score calculated for each participant.

SUS Statement	μ	σ	σ^2
I think that I would like to use this system frequently.	3.76	1.09	1.19
I found the system unnecessarily complex.	2.00	1.17	1.38
I thought the system was easy to use.	4.59	0.51	0.26
I think that I would need the support of a technical person to be able to use this system.	1.06	0.24	0.06
I found the various functions in this system were well designed.	4.18	0.64	0.40
I thought there was too much inconsistency in this system.	1.47	0.51	0.26
I would imagine that most people would learn to use this system very quickly.	4.47	0.87	0.76
I found the system very cumbersome to use.	1.76	0.97	0.94
I felt very confident using the system.	4.47	0.62	0.39
I needed to learn a lot of things before I could get going with this system.	1.29	0.47	0.22

Table 7: Summary of SUS Scores for each statement.

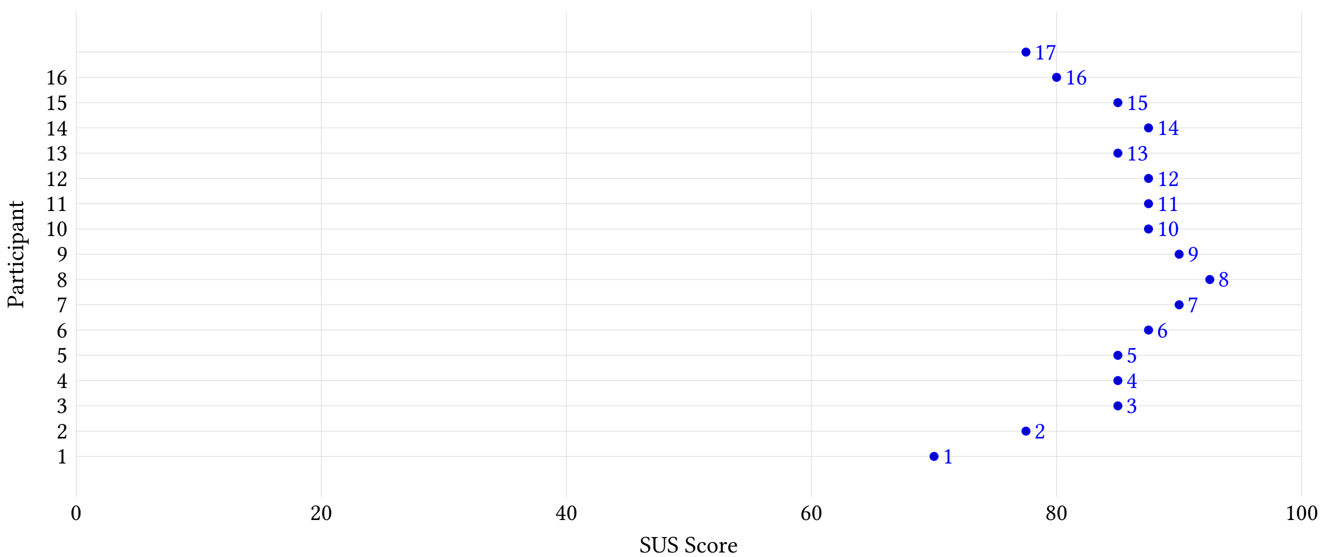


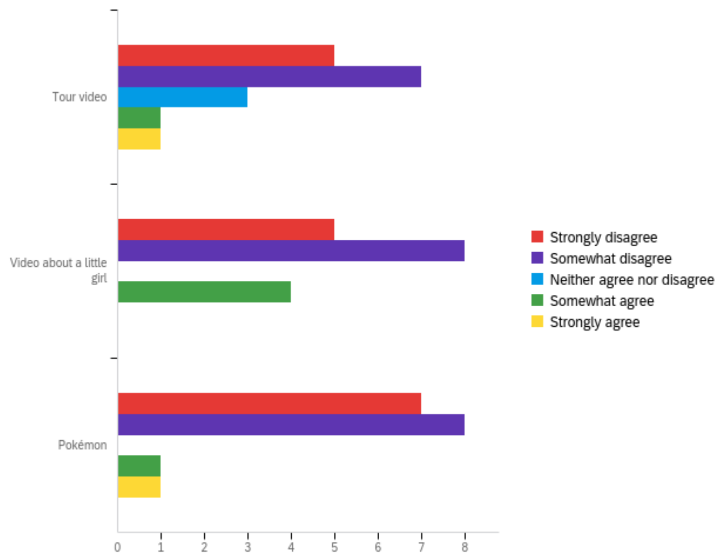
Figure 21: Dot plot of calculated SUS Scores for each participant.

D.4 Perception of dynamic thumbnails

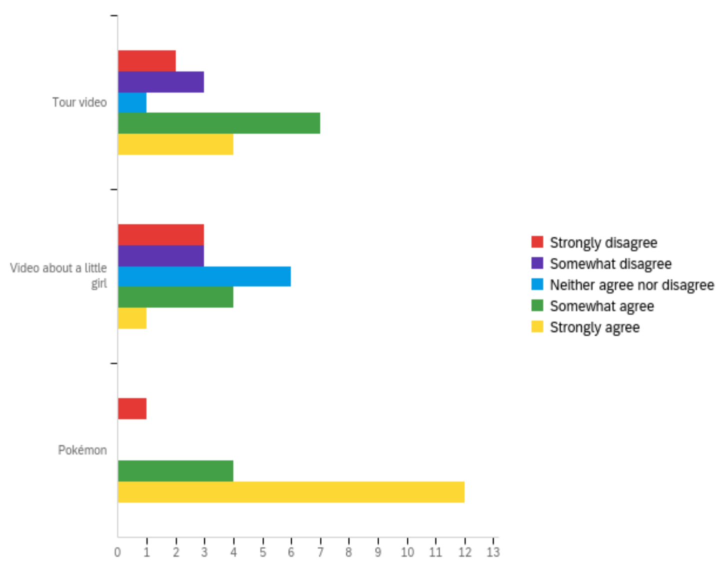
In the figure below, we present participants' perceptions of our dynamic thumbnails. Figure 22a shows responses to the question, "Did you find the thumbnails to be intrusive?" while Figure 22b shows responses to the question, "Did you find the thumbnails to be useful?"

For the question about intrusiveness, the vast majority of participants did not find the dynamic thumbnails intrusive. However, for video Type B, which features a little girl being pursued by a rainy cloud, there is a noticeable increase in participants who found the thumbnails intrusive. This may relate to the cloud often appearing in the same position as the rear-view dynamic thumbnail, as noted in participant comments in Section D.5.2.

For the question about usefulness, participants were generally split for videos of Types A and B. However, for video Type E, which featured multiple Pokémon scattered around a real environment, there was a strong consensus, with participants strongly agreeing about the usefulness of the dynamic thumbnails.



(a) Responses on whether dynamic thumbnails were perceived as intrusive.



(b) Responses on the usefulness of dynamic thumbnails.

Figure 22: Participant responses regarding the perception of dynamic thumbnails for each video type.

D.5 Participant comments

D.5.1 Positive aspects.

- "When there is action in multiple locations in the scene, the thumbnails help me to view all the action instead of having to choose. Additionally, when there is no action yet, the thumbnails help me keep track of where to look once the action starts. (extra useful when a scene changes)"
- "Very easy to use."
- "Especially very salient aspects like the truck coming from the side or certain pokemon are quick to identify with this. It becomes clear when something interesting is happening in the video"
- "you can see the entire view easily, ths is especially helpfull during tour guides"
- "The thumbnails help with seeing when something outside of your view happens, like a pokemon appearing, without having to look around continously."
- "I can be fast at realizing my environment"
- "Its nice to get an overview of the whole surroundings."
- "Able to quickly see what's going on around the main view and not having to turn around all the time to see things. It also makes turning around easier, since I already know where exactly something is that I want to look at."
- "ease of use"
- "wider field of view."
- "-Does help with spotting POI offscreen, if present -Not too intrusive"
- "A lot easier to find things in the scene as it's faster to look aroud"
- "Quickly provides information about every aspect of the video makes it easier to comprehend the situation"
- "It's nice to use in certain settings, like in the experiment, where you're tasked with finding something. It can also be fun when you're exploring someplace without actually being there. And of course, it can be useful in some games where a lot is going on at the same time."
- "You don't need to manually move the camera around as much."
- "It really helped me in spotting things that I otherwise would not have spotted, and all without obstructing too much of the scene."
- "easy to use, easy to understand. Clear overview, less scrolling, like using car mirrors while driving"

D.5.2 Negative aspects.

- "The thumbnails take up some screen space, although this is a very small disadvantage compared to the advantages."
- "A bit easy to lost which one corresponding to which view"
- "Sometimes it feels unncessary when there is not much else to focus on/going on in the video. E.g., in the tour guide video, I was supposed to focus on certain things, and I stayed mainlly there"
- "No one uses 360 video, it obstructs part of the view too"
- "They take up quite a lot of space in scenes where not a lot happens in multiple places at the same time. Like the little girl scene."
- "I can be overstimulated"
- "In situations with a lot going on, i found it harder to focus my attention on specific things (such as finding the tour guide in the middle of all the other tourist)."
- "It takes up a good amount of space in the upper half of the screen, which was especially obtrusive in the video of the little girl, as there was a black cloud on the top."
- "not very useful"
- "was ruining the fun of trying to discover things yourself."
- "-If no POI's offscreen, the system is fairly useless -Side thumbnails dont add too much to the experience"
- "Does obstruct some of the main view"
- "When too much is happening on the scene checking the main view and all thumbnails is harder and made me feel I'm missing something in the video"
- "I felt like having the thumbnails made me more inclined to look around than I usually would be, and it made me too distracted at times."
- "I found myself not using the 2 side thumbnails, focusing more on the main screen and the center thumbnail"
- "In the second video with the girl, the cloud was of some importance, but since it's up in the sky and the thumbnails only show left/right/back, it was sometimes harder to spot."
- "maybe have an L and R above the mirrors/windows. Maybe an upwards view too but that adds something for very specific videos."

D.6 Additional eye-tracking data

This appendix presents additional eye-tracking data gathered by iMotions related to Section 5.1 and Section 5.2 of the scientific paper.

Timeline navigation

The figure below shows the amount of time participants spent looking at the individual dynamic thumbnails during each video type during timeline navigation section of the user study. As discussed in Section 5.2, participants used the rear-view (labeled "Back") thumbnail significantly more than the side-view (labeled "Left" and "Right") thumbnails. Despite this, the average duration of usage, particularly for the side-view thumbnails, is extremely close to 0%.

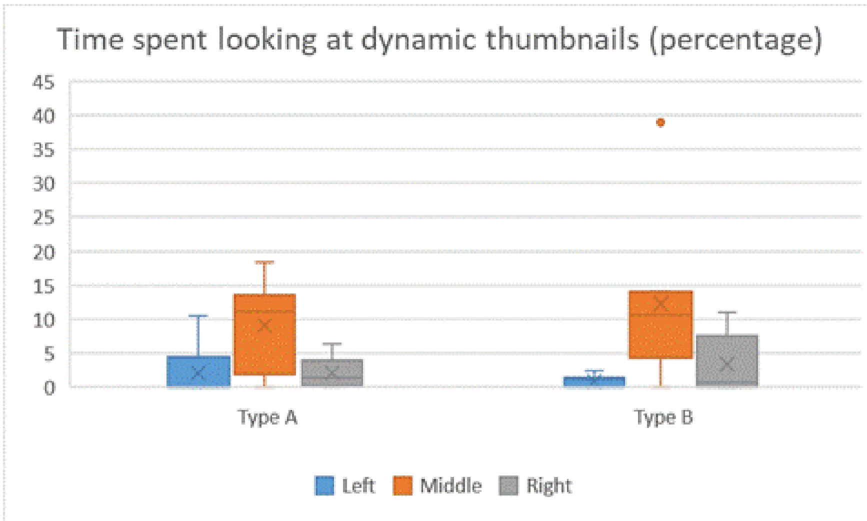


Figure 23: Viewing times of participants for individual dynamic thumbnails for each video type during timeline navigation.

D.7 Video type comparisons

In this appendix, we present additional statistical comparisons between video types (Type A, Type B, and Type E, as described in Section 4.3 of the paper). As with the main results, all data groups first underwent the Shapiro-Wilk test to assess normality. Based on the results of this test, we applied the appropriate statistical analyses.

D.7.1 Live viewing center of viewport. We compared the gaze fixation duration on center of the viewport between all video types for videos watched on a conventional 360° video player without dynamic thumbnails using the Kruskal-Wallis test, as some groups were not normally distributed. The results indicated no statistically significant difference between these groups ($H(2) = 0.210$, $p = 0.901$). We performed the same comparison for all video types watched on a 360° video player with dynamic thumbnails. Since these groups were normally distributed, we used a one-way ANOVA, which also indicated no statistically significant differences ($F(42) = 0.847$, $p = 0.436$).

The interesting takeaway from the main results in Section 5 of the paper is that pairwise comparisons for each video type, with and without dynamic thumbnails, indicated statistically significant differences. This suggests that while there are differences between video players, users exhibit similar gaze fixation duration patterns regardless of video type on a specific 360° video player. This highlights the impact of the video player on user engagement, rather than the content itself.

D.7.2 Live viewing specific dynamic thumbnails. We also compared the gaze fixation duration per specific dynamic thumbnail between all video types. For both the top-left and top-right side-view thumbnails, we used one-way ANOVA tests as both groups were normally distributed. The results indicated no statistically significant differences ($F(2, 27) = 0.735$, $p = 0.486$ and $F(2, 27) = 0.744$, $p = 0.482$).

We then performed statistical tests comparing the rear-view dynamic thumbnails for each video type using the Kruskal-Wallis test, as some groups were not normally distributed. This test indicated statistically significant differences ($H(2) = 6.134$, $p = 0.047$). We conducted Dunn's test with Bonferroni correction for post hoc analysis. The results showed no statistically significant difference between Type A and Type B videos ($p = 0.359$), and between Type A and Type E videos ($p = 1$). However, there was a statistically significant difference between Type B and Type E videos ($p = 0.043$).

The implications are that for all video types, as reported previously in Section D.6, participants showed limited usage of side-view dynamic thumbnails, with no significant differences in usage. However, for the rear-view dynamic thumbnail, there is a clear distinction between Type B and Type E videos. In Type B, participants followed one dynamic region of interest, while Type E featured multiple static and dynamic regions of interest. It makes sense that participants would be more inclined to view rear-view thumbnails in Type E videos ($\mu = 19\%$) compared to Type B ($\mu = 14.09\%$).

D.7.3 Timeline navigation dynamic thumbnails. We compared the time participants spent fixating their gaze on each of the dynamic thumbnails. We used the Kruskal-Wallis test for this analysis, as some groups were not normally distributed. For Type A video, we found a significant difference ($H(2) = 7.419$, $p = 0.025$). Similarly, for Type B video, we also found a significant difference ($H(2) = 6.615$, $p = 0.037$). Subsequently, we used Dunn's test with Bonferroni correction for post hoc analysis.For Type A video, there was a statistically significant difference between the top-left side-view thumbnail and the rear-view thumbnail ($p = 0.023$). However, no statistically significant difference was found between the top-right side-view thumbnail and the rear-view thumbnail ($p = 0.207$), nor between the top-left side-view thumbnail and the top-right side-view thumbnail ($p = 1$). We obtained similar results for Type B video: a statistically significant difference between the top-left side-view thumbnail and the rear-view thumbnail ($p = 0.042$), but no statistically significant difference between the top-right side-view thumbnail and the rear-view thumbnail ($p = 0.178$), nor between the top-left side-view thumbnail and the top-right side-view thumbnail ($p = 1$). There are results for Figure 23.

We did not include this part in the scientific paper because the results of gaze fixation duration for the vertical center of the viewport during the timeline navigation task (see Section 5.2.3) showed no statistically significant difference. Therefore, further analysis of a feature that participants did not engage with would not add value.

E LITERATURE REVIEW

This chapter discusses the preceding literature to a master's thesis project. The thesis is based on 360° videos, specifically viewed from 2D displays, such as computer monitors or mobile devices. While the research on 360° videos has been growing over the last few years, it has been very limited in the context of 2D displays. This research tries to propose a novel interface design that will enhance the user interaction, navigation, and experience while watching 360° videos.

Firstly, the review takes a closer look into what 360° video is, how its used and highlights its challenges. Afterwards, the concept of fear of missing out is briefly introduced along with its negative effects on human well-being, aiming to better understand the issues it presents. This phenomenon is further explored in the context of 360° videos, with an analysis of participants' experiences as they interact with such videos. Subsequently, we will examine various systems developed to either mitigate FOMO, guide user attention, or generally enhance the viewing experience of 360° videos. Moving forward, we delve into the challenge of navigating 360° videos using the timeline (also occasionally referred to as the seek bar). We scrutinize the existing approaches adopted by popular online video sharing platforms and examine proposed enhancements, considering both conventional and 360° videos.

Finally, in the process of creating a new user interface, it is imperative to give consideration to user experience. We will explore important design considerations that emerge when implementing user interface enhancements for a video player. Furthermore, we will discuss methodologies for evaluating the effectiveness of a newly proposed system.

E.1 Introduction to 360° videos

360° videos, also referred to as omnidirectional or panoramic videos, have gained widespread popularity in commercial applications in recent years. Major video platforms such as YouTube and Facebook have played a pivotal role in championing their utilization as they are trying to take advantage of the new medium. Initially, 360° videos found primary use in gaming and entertainment content. However, their applications have expanded significantly, with a increasing number of uses in diverse fields such as education, tourism [22, 41], and sports, among many others. They are captured using specialized omnidirectional cameras that, unlike traditional cameras with limited fields of view, are capable of recording light from all directions.

One might naturally infer that 360° videos were primarily designed for VR head-mounted displays (HMDs) like HTC Vive or Oculus Rift [20], affordability of which is increasing for the average consumer. Despite this, HMDs are not always available and the predominant consumption of 360° videos still occurs on conventional 2D displays, such as those on desktop computers or mobile devices. Additionally, even though HMDs offer viewers a more immersive experience, it often comes at the expense of heightened cognitive burden, motion sickness and physical discomfort [7]. This poses a potentially significant issue during prolonged video viewing, as highlighted by participants in the study conducted by Passmore et al. [30] on the viewer experience of watching 360° videos. Consequently, given the choice, viewers might still lean towards watching 360° videos on 2D displays.

HMDs grant viewers the freedom to explore their surroundings by simply moving their heads, whereas viewers watching 360° videos on 2D displays navigate through a 2D normal field of view (NFoV) using a mouse or touchscreen. This poses a challenge as viewers need to constantly focus and refocus on their intended targets [23]. Alternatively, 360° videos can be displayed in their entirety using equirectangular projection [29] but the unfamiliar format and geometrical distortions can make the video difficult to watch [19, 22].

E.2 Fear of missing out

Fear of missing out (FOMO) is a phenomenon primarily researched in the context of social media [2, 28, 33]. It is defined as *"a pervasive apprehension that others might be having rewarding experiences from which one is absent"* and characterized by *"the desire to stay continually connected with what others are doing"* [33]. This phenomenon was first recognized and contextualized in field of marketing by Dan Herman, in *Journal of Brand Management*, attributing it to consumers as a compelling motivator for making purchases [13]. However, the term gained significant recognition in 2004 when Patrick J. McGinnis coined it in the Harvard Business School magazine, *Harbus*.

Simpler access to information online via technology tends to motivate individuals to easily compare their own lives to the lives of individuals on social media, as noted by Abel et al. [2]. This tendency contributes to a diminished sense of satisfaction with one's own life and behaviors. For instance, on social media platforms, people often share their experiences, like attending parties or traveling. This can lead their followers to compare their own lives and, in turn, foster a sense of being left out or a desire to do more with their own lives.

In their paper, Przybylski et al. [33] revealed that individuals high in FOMO reported lower general mood and reduced life satisfaction. Following this, Baker et al. [5] explored the connection between FOMO and its influence on both physical and depressive symptoms. Their investigation extended to examining its association with mindful attention and awareness. The findings revealed that individuals with a greater fear of missing out experienced more physical symptoms, more depressive symptoms and less mindful attention. These outcomes imply a correlation with poorer physical, emotional, and cognitive health. These results find further support in work by Milyavskaya et al. [28] who investigated the FOMO among college freshmen. Their study reported fatigue, stress, physical symptoms, and decreased sleep.

While FOMO is typically scrutinized in context of negative effects and regarded as a negative phenomenon, certain studies propose that it may not be entirely negative [3, 28]. Some research suggests that exploring the potentially positive aspects of FOMO warrants further investigation in the future.

E.3 FOMO in 360° videos

Aitamurto et al. [3] conducted a thorough examination of FOMO within the context of 360° videos. They define FOMO as fear of missing out on parts of a video outside of viewer's field of view (FOV) that contribute to a narrative occurring. It is an inevitable phenomenon as the FOV a viewer looks through is limited, typically varying from 60° to 110° [21, 34, 42, 43]. Hence, when a viewer opts to watch a specific segment of a 360° video, they are intentionally forgoing the opportunity to witness events unfolding beyond the confines of their current FOV.

In their experiment, 10 minute 360° film was used, which was vertically divided into two 180° spheres, that follow the first-person point of view of two distinct characters. Consequently, two narrative threads run concurrently, presenting different sides of the story, allowing viewers to choose which perspective to follow. Moreover, the study involved participants viewing the video under four distinct conditions:

- Free choice: the viewer can choose any FOV in the 360° sphere
- 30-second timed rotation: every 30 seconds the FOV switches to the other character's point of view
- 90-second timed rotation: every 90 seconds the FOV switches to the other character's point of view
- 180-degree view: the viewer sees point of view of only one character throughout the entire film

Additionally, participants were instructed to view the film a second time, starting from the alternate FOV. The participants were not told that a second viewing of the same film would commence, as they did not want it to affect the results of the first viewing session. To assess the FOMO, they devised a 5-item Likert scale designed to gauge feelings and emotions associated to FOMO, encompassing feelings of distraction, frustration, apprehension about missing unfolding scenes on the opposite side of the 360° sphere, and regret over choices made.

Their findings show that the main contributing factor to FOMO in 360° videos is the viewer's awareness of events occurring in parallel. It causes anxiety, frustration and stress as on any given moment the viewer would need to adjust their FOV to keep track of important events while at the same time getting distracted by looking around the environment or things happening in other sections of the video. These finding are further supported by other studies on 360° videos [4, 12, 30, 34, 40, 44]:

- Passmore et al. [30] reported participants getting distracted by environments and not being sure where to look, consequently making them confused and fear that they are missing out on events happening on the other side of the video.
- Fonseca et al. [12] had participants report that they were too distracted with their surroundings and visuals in immersive videos and missed out on some parts of narration (such as voice over).
- Sarker et al. [34] reported that overstimulating participants in 360° videos with audio-visual cues results in lack of guidance and evokes feelings of frustration as they feel like they are not in control. Consequently creating feeling of stress thinking they might be missing something important as too many things are happening at the same time.
- According to Tse et al. [40], 35% of the participants reported some level of FOMO, as revealed in their interview data. Additionally, 33% of the participants found 360° videos to be potentially distracting as a medium, diverting attention from the narrative occurring.
- In research done by Zoric et al. [44], participants expressed concern over stress with dealing with too many options when viewer is their own producer in panoramic videos. *"If you are your own producer you know that you will miss something and when you have the choice of being very individual in picking some specific scene then I think it's easy for you to feel worried, like 'am I missing something? Is everyone else watching something else?"*.

Furthermore, findings by Aitamurto et al. suggest that some participants may experience the feeling of excitement due to their freedom of choice with abundance of options. They refer to this phenomenon as joy of missing out (JOMO) and suggest viewers can feel FOMO and JOMO simultaneously. A viewer might feel stress and anxiety when trying to capture the entire narrative occurring, while at the same time feel excitement and joy from having freedom of choice. While this concept is intriguing and warrants further research, it is not the focus of this thesis.

Finally, in a recent study, Uslu et al. [41] have examined the influence of FOMO in the context of 360° tour videos, specifically museums. They suggest that viewers who value continual learning experience anxiety about missing on information relevant to their interest. They also mention how users' intent matters when it comes to watching a video, whether watching casually or trying to view it as a learning experience, heavily influences the intensity of FOMO.

E.4 360° video viewing techniques

This section examines and discusses different approaches researchers have used to address challenges like FOMO, steering user attention, and providing guidance during the viewing of 360° videos. These approaches include implementations in both VR and conventional 2D displays, as insights gained from both efforts may not be exclusive.

Lin et al. [24] introduced a visualization technique called *Outside-In*, which incorporates off-screen regions of interest (ROIs) onto the main screen through picture-in-picture (PIP) thumbnails. The concept primarily draws inspiration from perspective projection, defined as the altering the shapes of objects based on their distance from the viewer. In their study, Lin et al. classify videos into five distinct types [24]:

- Type A: commonly seen in tour-guided videos, where a guide tells the audience about the concurrent target.
- Type B: found in extreme sports videos, for example, one athlete moves fast in the scene
- Type C: commonly seen in videos of entertainment shows, such as a group of performers playing music or dancing around the camera
- Type D: includes films in which characters move and chase each other
- Type E: a mixture of Types C and D, in that the characters sometimes moves around and sometimes stay still

Types A and B include videos featuring only a single ROI, while Types C, D, and E involve multiple ROIs. In their study, they exclusively focused on the latter. Their results demonstrate that, overall, the *Outside-In* technique helps users in swiftly comprehending the storyline and the spatial relationship of ROIs. It enables users to focus more on the main storyline while easily perceiving information that is out of sight. Nevertheless, there is a noteworthy shortcoming should be considered. Some users expressed concerns about how occluded screen can get when there were numerous PIP thumbnails at once. Despite this, the findings indicate no significant difference in distraction levels when compared to the use of arrows to point towards a ROI.

In their study, Pavel et al. [31] adopted a different approach. Rather than bringing the ROI to the user, they reposition user's perspective towards the ROI. Their *viewport-oriented* technique reorients the shot at each cut, ensuring that the most important ROI aligns with the user's FOV. With their *active reorientation* technique, users can press a button to promptly realign the shot so that the most important ROI lies in their FOV. The system relies on labels to mark the positions of important points in the video. These labels can be generated either automatically or manually. Their findings indicate that, when compared to conventional cuts, users spend more time looking at ROIs. Users also exhibit a preference to these two proposed techniques (compared to conventional cuts). Noteworthy, users who expressed dissatisfaction with the *active reorientation* technique primarily mentioned the inconvenience of having to press a button.

Yamaguchi et al. [43] directly tackled the challenge of alleviating FOMO during the viewing of 360° videos, specifically within the realm of the VR experience. They introduced a novel solution in the form of a panoramic thumbnail situated at the tip of the VR controller. However, it is not always visible. Users have the freedom to choose to display the thumbnail at their convenience and also hide it if desired. The panoramic thumbnail utilizes equirectangular projection. As explained earlier in the introduction, this projection method creates a flat image representation of the entire video frame. Hence, users have an on-demand overview of the entire video at a point in time, allowing them to observe events occurring behind them. The videos selected for their experiment were drawn based on Lin et al.'s [24] classification, specifically Types A, B, and E. Their findings indicate that this overview enhances the users' viewing experience and can contribute to mitigating FOMO. Notably, they observed that *"many participants looked at the panoramic thumbnail immediately after the video started"* [43]. This observation aligns with the research of Duanmu et al. [11], who investigated user navigation behaviors in 360° videos on 2D displays. Duanmu et al.'s data suggests that users tend to extensively explore their surroundings at the start of videos, after which their attention shifts towards the main content. Teodosio et al. [38] referenced Julian Hochberg's work *In the Mind's Eye* [14], where Hochberg, back in 1978, articulated the concept of the *mind's eye* - the human ability to visualize or imagine things mentally. They describe how the human visual system enables rapid exploration of a space, seamlessly integrating information gathered from successive glances into a unified and coherent spatial model within the *mind's eye*. This may suggest the significance of providing assistance for user orientation and navigation at the start of a video to ensure their familiarity with the overall setting.

Su et al. [37] introduced *Pano2Vid* and its companion algorithm, *AutoCam*. The primary function of *Pano2Vid* is to extract "natural-looking" NFOV videos, defined by a 65.5° horizontal angle. The objective is to create videos that simulate the appearance of content captured by a conventional NFOV camera. Subsequently, *AutoCam* assesses the "capture-worthiness" of the scenes, essentially gauging the human-like quality. The system then selects trajectories based on the "capture-worthiness" score and the degree of human-like camera movements. These selections are determined through an unsupervised learning model trained on video data sourced from YouTube. Subsequent to this, Su et al. [36] further refine the *Pano2Vid* technique by incorporating the ability to modify the FOV. This enhancement aims to simulate a more human-like quality, considering that zooming in and out is a common practice when capturing content through NFOV cameras.

In their study, Lin et al. [23] center their attention on the challenge of users needing to focus and refocus on their intended target when viewing 360° videos. To tackle this challenge, they devised two assistance techniques: *Auto Pilot (AP)*, which guides the viewer directly to the target, and *Visual Guidance (VG)*, which indicates the location of the target. They emphasize the importance of viewers swiftly navigating to their intended target, as it is often the main focus of the video. The experiment included two types of video: a sports video (Type B) and a tour video (Type A). In the sports video, users showed a significant preference for *AP* as it automatically tracked the skateboarder without requiring user intervention. On the other hand, in the tour video, the results were more evenly distributed. Those who favored *AP* appreciated the focus on the tour guide, who consistently highlighted essential buildings. In contrast, individuals who preferred *VG* expressed enjoyment in the freedom to explore their surroundings. Lin et al. suggest that *"the varied preferences and experiences seemed to be linked not only to individual differences but also to the participants' goal of watching the video"* [23]. This aligns with assertions made by Uslu et al. [41], who also discuss how the intent behind watching a museum tour video affects the extent of FOMO.

Kang et al. [19] condense and classify the aforementioned implementations into two distinct categories: automatic navigation [15, 21, 23, 36, 37] and interactive navigation [23, 24, 31, 43]. The former involves the creation of standard non-360° from 360° videos, guided by specific criteria. For instance Lin et al.'s *AP* follows the main target of a video. On the other hand, interactive navigation allows users to freely interact with the 360° video while receiving additional assistance to ensure they do not miss out on important events. The main drawback identified by Kang et al. regarding automatic navigation is that users forfeit the ability to interact with 360° videos, which is intended to be their primary function. In their solution, they address the issue by creating a hybrid automatic and interactive navigation system. The system identifies the most salient areas in the input video, creating an autopilot based on these scores. Users retain the ability to change their viewing direction, and the system adapts accordingly, generating a new path based on the users' intent.

Finally, Wang et al. [42] presented *Transitioning360*, a system closely resembling the one proposed by Kang et al.'s [19]. It generates multiple paths, the number of which is specified by an input. The generated paths are crafted based on content awareness and path diversity. Content awareness aims to focus on the most interesting segments of the video, while diversity tries to ensure a broad range of perspectives without excessive overlap. To present the available paths to users, they project thumbnails onto the screen. The default layout, referred to as the *horizontal layout*, consists of a strip positioned horizontally at the bottom of the video, featuring all the thumbnails. Additionally, they also incorporate Lin et al.'s *Outside-In* PIP thumbnails [24], utilizing perspective projection. In summary, participants exhibited a strong preference for watching 360° videos using the *Transitioning360* system with the *Outside-In's* visualization method. Their own proposed *horizontal layout* of thumbnails was perceived as confusing by some users, as they struggled to discern their location within the video. In contrast, the *Outside-In* approach provided a clearer representation of the video's context.

In addition to all this, alternative methods for manipulating user attention have been explored, including the utilization of audio and visual cues [10] or social cues conveyed through the actions of a storyteller, without the incorporation of supplementary interface add-ons [39]. However, these approaches are considered irrelevant to the scope of this thesis.

E.5 Timeline navigation

Video navigation on timelines is an issue that has been researched thoroughly in the recent years on conventional videos. An action, referred to as scrubbing, is when user moves the slider along a timeline of a video and the current frame updates based on it. This is a technique that is commonly used when user either wants to skim through the video to get a quick overview of its content or when they want to rewind to some particular scene when re-watching. Video streaming sites such as YouTube and Netflix[8] have also implemented a small thumbnail on their video players that appears when scrubbing through a video timeline.

E.5.1 Conventional videos. Schoeffmann et al. [35] scrutinized various techniques aimed at enhancing video content navigation. As they point out, a significant number of these focus on improving timeline behaviour. In the realm of navigating timelines in conventional videos, a considerable amount of research is dedicated particularly to challenges inherent in longer videos, where even slight movements can lead to flashing and abrupt scene changes. This issue arises when the number of frames in a video exceeds the number of pixels on the timeline. Pongnumkul et al. [32] proposed a solution known as the *content-aware dynamic timeline* to tackle this challenge. This approach draws inspiration from work done by Hürst et al. [18], who utilized elastic graphical interfaces. In these interfaces, the playback speed undergoes dynamic adjustment as a nonlinear function of the distance between the handle and the mouse pointer. Elastic graphical interfaces were first introduced by Masui et al. [25] in 1996.

On the other hand, Matejka et al. [26, 27] opted for a different approach instead of manipulating the timeline directly; they chose to alter the presentation of visual content while scrubbing through a timeline. As previously mentioned, popular video sharing platforms like YouTube and Netflix provide users with only a small preview of the frame. In their implementation of *Swift* [26], they specifically addressed the latency issue associated with loading these previews on longer videos. *Swift* reduces the video resolution while the timeline is being dragged, and upon releasing the mouse, it returns to its original resolution. Matejka et al. then extended their idea with the introduction of *Swifter* [27]. *Swifter* incorporates a grid of thumbnails around the active timeline location. This grid enables users to select a specific frame they wish to jump to. In tasks involving scene location, *Swifter* demonstrated a 48% improvement in speed compared to existing commercial techniques.

E.5.2 360° videos. Li et al. [22] introduced the concept of *Route Tapestries*, which involves a seamless orthographic-perspective mapping of scenes along predefined camera routes. Their approach is specifically tailored for video players intended for 2D screens. Li et al. highlight that existing systems utilize planar thumbnails to provide an overview of an entire frame for temporal navigation. These thumbnails are generated by transforming spherical 360° images into 2D visualizations; for example, YouTube employs the equirectangular projection method. As mentioned in the introduction, equirectangular projections exhibit geometric distortions. While they may offer users an overview of the current frame, navigating through them can be difficult. According to their user study, participants generally expressed a preference for *Route Tapestries* when given the choice, presumably for the reasons outlined earlier.

E.6 Design considerations

E.6.1 Design implications. Zoric et al. [44] highlight some design implications in their research on panoramic videos, which are less free-form than 360° videos but still share numerous similarities, like being able to pan around and selectively focus your attention on sections of a video. They emphasise the importance of balance between active and passive viewing. Even though having full control over a video can be exciting, sometimes users just want to sit back and let the creator take the control. Having constant control over a video can be too much effort, cause exhaustion, and even stress, especially over a longer period of time.

In the realm of 360° videos, it is imperative that the experience of free-form viewing and exploration remains intuitive for users. Navigating within the environment should be seamless, ensuring that viewers do not feel disoriented or lost amidst the dynamic scenes. Zoric et al. suggest that facilitating this can be achieved through the implementation of enhanced overviews of the entire scene, allowing users to maintain a sense of spatial awareness. Yamaguchi et al. [43] exemplified this by representing the entire current video frame on panoramic overview thumbnail. Additionally, Zoric et al. suggest that providing supplementary detailed views within the video enhances the viewer's ability to delve into specific elements of interest (ROIs). Lin et al. [24] exemplified this concept through their implementation of *Outside-In's* PIP thumbnails. Wang et al.'s [42] demonstrated through their utilization of horizontal thumbnails that merely placing them sequentially on a strip without any context leads to confusion.

[8] https://www.netflix.com/

Furthermore, Duanmu et al. [11] investigated the viewing behaviors of individuals watching 360° videos on 2D displays. They covered twelve diverse video types, ranging from a roller coaster ride to a city tour of Amsterdam. The study revealed intriguing implications for the implementation of new interface add-ons. Notably, they observed that at the beginning of the videos, users tended to explore their surroundings extensively. As discussed in Section E.4, this phenomenon could be attributed to human tendency to visually scan their environment, contributing to the mental construction of a spatial representation within the individual's *mind's eye*. Subsequently, the viewer's focus shifts to the video's main content. Duanmu et al. demonstrate this by visualizing a heatmap that indicates users' focal points across the duration of the video. During the roller coaster ride, users predominantly fixate their view on the track, while in the city tour of Amsterdam, they look at buildings, pedestrians, and vehicles due to the absence of a dominant focal point. The findings also indicated that most exploration happened along the horizontal axis at the center of the video, with the top and bottom receiving minimal attention.

Finally, as previously discussed in Section E.4, participants in Pavel et al.'s research noted that they *"didn't want to have to click a button to see the important point"* [31]. This suggests that introducing new keybindings as an additional means of interacting with a video player could be perceived as more of a nuisance and, therefore, should not be considered when designing an implementation.

E.6.2 Accessibility. When designing new systems and their interfaces it is important to keep accessibility in mind. Not only does everyone deserve opportunity to receive information, but it also helps publishers to reach wider audiences. This principle holds true for 360° video players as well. Hughes et al. [17] conducted a comprehensive survey in their paper, examining various 360° video players that are currently being used. They cite the World Wide Web Consortium (W3C)[9] as a source for the *Web Content Accessibility Guidelines*, which outline four essential features: transcripts, subtitles, audio descriptions, and sign language. When developing a new video player add-ons, it is crucial to make diligent efforts to avoid compromising the usability of any of these features.

E.6.3 Usability. The System Usability Scale (SUS), introduced by John Brooke in 1996 [8], is a widely used questionnaire for assessing the perceived usability of a system. It consists of 10 statements with a 5-point Likert scale for responses:

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well designed.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

The participants are asked to provide their level of agreement or disagreement with the statements. The scores are converted and calculated to provide a usability score that ranges from 0 to 100, with 68 being the average score [9]. Scores above the 68 mark indicate an above-average level of usability, while scores below suggest a below-average user experience. *"The simplicity of the SUS statements is intentional as it makes the questionnaire easy to understand, consistent and efficient. The individual statements are not supposed to have diagnostic value in themselves or relate to any specific feature of a particular system"* [9]. As a result, as proven throughout the decades of its use, SUS questionnaire is very robust and applicable to varying systems, including newly emerging technology [1]. In his 2013 retrospective on the SUS questionnaire [9], Brooke reflected on the enduring relevance and applicability of his scale to emerging technologies. He emphasizes that it will be a considerable time before a technology that does not conform to the SUS questionnaire emerges. To demonstrate the viability of the SUS questionnaire, Huang [16] applied it to assess the usability of riding a scooter in a 360° VR video. Additionally, Broeck et al. [7] incorporated the questionnaire into their study, where they conducted a comparative analysis of the viewing experience of 360° videos on different devices.

[9]https://www.w3.org/

REFERENCES

[1] Aaron Bangor, P. T. K. and Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human–Computer Interaction*, 24(6):574–594.

[2] Abel, J., Buff, C., and Burr, S. (2016). Social media and the fear of missing out: Scale development and assessment. *Journal of Business Economics Research (JBER)*, 14:33.

[3] Aitamurto, T., Won, A. S., Sakshuwong, S., Kim, B., Sadeghi, Y., Stein, K., Royal, P. G., and Kircos, C. L. (2021). From fomo to jomo: Examining the fear and joy of missing out and presence in a 360° video viewing experience. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA. Association for Computing Machinery.

[4] Aitamurto, T., Zhou, S., Sakshuwong, S., Saldivar, J., Sadeghi, Y., and Tran, A. (2018). Sense of presence, attitude change, perspective-taking and usability in first-person split-sphere 360° video. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–12, New York, NY, USA. Association for Computing Machinery.

[5] Baker, Z. G., Krieger, H., and LeRoy, A. S. (2016). Fear of missing out: Relationships with depression, mindfulness, and physical symptoms. *Translational Issues in Psychological Science*, 2(3):275.

[6] Bangor, A., Kortum, P., and Miller, J. (2009). Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3):114–123.

[7] Broeck, M. V. d., Kawsar, F., and Schöning, J. (2017). It's all around you: Exploring 360° video viewing experiences on mobile devices. In *Proceedings of the 25th ACM International Conference on Multimedia*, MM '17, page 762–768, New York, NY, USA. Association for Computing Machinery.

[8] Brooke, J. (1996). Sus: a 'quick and dirty' usability scale. *Usability evaluation in industry*, 189(3):189–194.

[9] Brooke, J. (2013). Sus: a retrospective. *Journal of usability studies*, 8(2):29–40.

[10] Brown, A., Sheikh, A., Evans, M., and Watson, Z. (2016). Directing attention in 360-degree video. pages 29 (9 .)–29 (9 .).

[11] Duanmu, F., Mao, Y., Liu, S., Srinivasan, S., and Wang, Y. (2018). A subjective study of viewer navigation behaviors when watching 360-degree videos on computers. In *2018 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6.

[12] Fonseca, D. and Kraus, M. (2016). A comparison of head-mounted and hand-held displays for 360° videos with focus on attitude and behavior change. In *Proceedings of the 20th International Academic Mindtrek Conference*, AcademicMindtrek '16, page 287–296, New York, NY, USA. Association for Computing Machinery.

[13] Herman, D. (2000). Introducing short-term brands: A new branding tool for a new consumer reality. *Journal of Brand Management*, 7.

[14] Hochberg, J. and Peterson, M. (2007). In the mind's eye. *In the mind's eye: Julian Hochberg on the Perception of Pictures, Films, and the World*, pages 70–99.

[15] Hu, H.-N., Lin, Y.-C., Liu, M.-Y., Cheng, H.-T., Chang, Y.-J., and Sun, M. (2017). Deep 360 pilot: Learning a deep agent for piloting through 360deg sports videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[16] Huang, F.-H. (2020). Perceived usability evaluation of 360° immersive video service: Empirical evaluation of the system usability scale. In Stephanidis, C. and Antona, M., editors, *HCI International 2020 - Posters*, pages 438–444, Cham. Springer International Publishing.

[17] Hughes, C. J. and Montagud, M. (2021). Accessibility in 360 video players. *Multimedia tools and applications*, 80(20):30993–31020.

[18] Hürst, W., Götz, G., and Jarvers, P. (2004). Advanced user interfaces for dynamic video browsing. In *Proceedings of the 12th Annual ACM International Conference on Multimedia*, MULTIMEDIA '04, page 742–743, New York, NY, USA. Association for Computing Machinery.

[19] Kang, K. and Cho, S. (2019). Interactive and automatic navigation for 360° video playback. *ACM Trans. Graph.*, 38(4).

[20] Koulieris, G. A., Akşit, K., Stengel, M., Mantiuk, R. K., Mania, K., and Richardt, C. (2019). Near-eye display and tracking technologies for virtual and augmented reality. *Computer Graphics Forum*, 38(2):493–519.

[21] Lai, W.-S., Huang, Y., Joshi, N., Buehler, C., Yang, M.-H., and Kang, S. B. (2017). Semantic-driven generation of hyperlapse from 360 degree video. *IEEE Transactions on Visualization and Computer Graphics*, 24(9):2610–2621.

[22] Li, J., Lyu, J., Sousa, M., Balakrishnan, R., Tang, A., and Grossman, T. (2021). Route tapestries: Navigating 360° virtual tour videos using slit-scan visualizations. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, UIST '21, page 223–238, New York, NY, USA. Association for Computing Machinery.

[23] Lin, Y.-C., Chang, Y.-J., Hu, H.-N., Cheng, H.-T., Huang, C.-W., and Sun, M. (2017a). Tell me where to look: Investigating ways for assisting focus in 360° video. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 2535–2545, New York, NY, USA. Association for Computing Machinery.

[24] Lin, Y.-T., Liao, Y.-C., Teng, S.-Y., Chung, Y.-J., Chan, L., and Chen, B.-Y. (2017b). Outside-in: Visualizing out-of-sight regions-of-interest in a 360° video using spatial picture-in-picture previews. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, page 255–265, New York, NY, USA. Association for Computing Machinery.

[25] Masui, T., Kashiwagi, K., and Iv, G. (1996). Elastic graphical interfaces for precise data manipulation.

[26] Matejka, J., Grossman, T., and Fitzmaurice, G. (2012). Swift: reducing the effects of latency in online video scrubbing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 637–646.

[27] Matejka, J., Grossman, T., and Fitzmaurice, G. (2013). Swifter: Improved online video scrubbing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, page 1159–1168, New York, NY, USA. Association for Computing Machinery.

[28] Milyavskaya, M., Saffran, M., Hope, N., and Koestner, R. (2018). Fear of missing out: prevalence, dynamics, and consequences of experiencing fomo. *Motivation and emotion*, 42(5):725–737.

[29] Neng, L. A. R. and Chambel, T. (2010). Get around 360° hypervideo. In *Proceedings of the 14th International Academic MindTrek Conference: Envisioning Future Media Environments*, MindTrek '10, page 119–122, New York, NY, USA. Association for Computing Machinery.

[30] Passmore, P., Glancy, M., Philpot, A., Roscoe, A., Wood, A., and Fields, B. (2016). Effects of viewing condition on user experience of panoramic video.

[31] Pavel, A., Hartmann, B., and Agrawala, M. (2017). Shot orientation controls for interactive cinematography with 360 video. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pages 289–297.

[32] Pongnumkul, S., Wang, J., Ramos, G., and Cohen, M. (2010). Content-aware dynamic timeline for video browsing. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, page 139–142, New York, NY, USA. Association for Computing Machinery.

[33] Przybylski, A. K., Murayama, K., DeHaan, C. R., and Gladwell, V. (2013). Motivational, emotional, and behavioral correlates of fear of missing out. *Computers in Human Behavior*, 29(4):1841–1848.

[34] Sarker, B. (2017). Decoding the user experience in mobile virtual reality narratives. In Lackey, S. and Chen, J., editors, *Virtual, Augmented and Mixed Reality*, pages 437–452, Cham. Springer International Publishing.

[35] Schoeffmann, K., Hudelist, M. A., and Huber, J. (2015). Video interaction tools: A survey of recent work. *ACM Comput. Surv.*, 48(1).

[36] Su, Y.-C. and Grauman, K. (2017). Making 360° video watchable in 2d: Learning videography for click free viewing. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1368–1376.

[37] Su, Y.-C., Jayaraman, D., and Grauman, K. (2017). Pano2vid: Automatic cinematography for watching 360° videos. In Lai, S.-H., Lepetit, V., Nishino, K., and Sato, Y., editors, *Computer Vision – ACCV 2016*, pages 154–171, Cham. Springer International Publishing.

[38] Teodosio, L. A. and Mills, M. (1993). Panoramic overviews for navigating real-world scenes. In *Proceedings of the first ACM international conference on Multimedia*, pages 359–364.

[39] Tong, L., Jung, S., and Lindeman, R. W. (2019). Action units: Directing user attention in 360-degree video based vr. In *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology*, VRST '19, New York, NY, USA. Association for Computing Machinery.

[40] Tse, A., Jennett, C., Moore, J., Watson, Z., Rigby, J., and Cox, A. L. (2017). Was i there? impact of platform and headphones on 360 video immersion. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, page 2967–2974, New York, NY, USA. Association for Computing Machinery.

[41] Uslu, A. and Tosun, P. (0). Examining the impact of the fear of missing out on museum visit intentions. *Journal of Hospitality & Tourism Research*, 0(0):10963480231168608.

[42] Wang, M., Li, Y.-J., Zhang, W.-X., Richardt, C., and Hu, S.-M. (2020). Transitioning360: Content-aware nfov virtual camera paths for 360° video playback. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 185–194.

[43] Yamaguchi, S., Ogawa, N., and Narumi, T. (2021). Now i'm not afraid: Reducing fear of missing out in 360° videos on a head-mounted display using a panoramic thumbnail. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 176–183.

[44] Zoric, G., Barkhuus, L., Engström, A., and Önnevall, E. (2013). Panoramic video: Design challenges and implications for content interaction. In *Proceedings of the 11th European Conference on Interactive TV and Video*, EuroITV '13, page 153–162, New York, NY, USA. Association for Computing Machinery.