MASTER THESIS

Effectively browsing 360-degree videos

Student:
Peter de Keijzer
ICA-3859436

Supervisor:
Dr. Wolfgang Hürst
Prof. Lynda Hardman

Department of Information and Computing Sciences

May 23, 2019

Universiteit Utrecht

1 Abstract

360-degree videos allow for immersive viewing experiences when watched using a head-mounted display (HMD). However, they also introduce new problems not present in "regular" 2D videos. For example, it is exceedingly more difficult to quickly browse a 360-degree video to get a quick glimpse of the content. This 360-degree video browsing problem can be divided into two smaller problems. One such problem is allowing the user to get a good overview of the content while browsing the video and another is being able to easily find scenes containing interesting objects or areas in the video. The HMD's limited field of view allows them to see only a small part of the video at a time which proves problematic when trying to get a good overview, or when trying to scan the virtual environment for a particular object or area of interest.

We introduce a number of techniques that address both of these problems, allowing the user to efficiently browse through a 360-degree video. We further demonstrate that these techniques cause no motion sickness, which is a common problem with head-mounted VR.

2 Acknowledgements

Throughout the writing of this master thesis I received a great amount of support and assistance from Dr. Wolfgang Hürst, who agreed to be my supervisor despite his very busy schedule. I would like to thank him for his expertise and knowledge. I would also like to thank Prof. Lynda Hardman who agreed to be my second supervisor.

Furthermore, I would like to thank my brother who helped me in the early stages of the user testing phase of this study and provided me with valuable counsel in deliberations over problems and findings.

Contents

3 Scientific Paper

This section contains the scientific paper that was written as a part of this master's thesis.

Effectively browsing 360-degree videos

Peter de Keijzer
Utrecht University

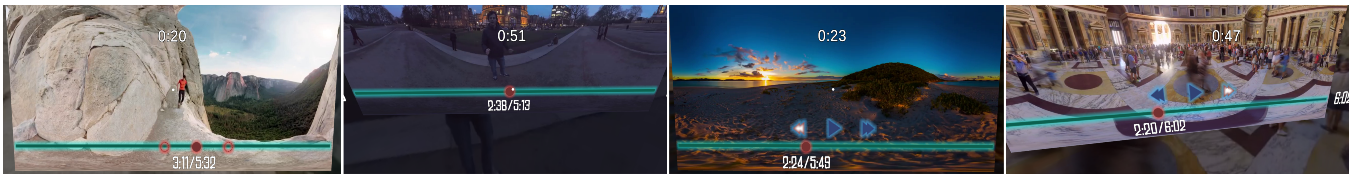

Figure 1: The techniques designed to address the 360-degree video browsing problem

ABSTRACT

360-degree videos allow for immersive viewing experiences when watched using a head-mounted display (HMD). However, they also introduce new problems not present in "regular" 2D videos. For example, it is exceedingly more difficult to quickly browse a 360-degree video to get a quick glimpse of the content. This 360-degree video browsing problem can be divided into two smaller problems. One such problem is allowing the user to get a good overview of the content while browsing the video and another is being able to easily find scenes containing interesting objects or areas in the video. The HMD's limited field of view allows them to see only a small part of the video at a time which proves problematic when trying to get a good overview, or when trying to scan the virtual environment for a particular object or area of interest.

We introduce a number of techniques that address both of these problems, allowing the user to efficiently browse through a 360-degree video. We further demonstrate that these techniques cause no motion sickness, which is a common problem with head-mounted VR.

KEYWORDS

360-degree video, panoramic video, video browsing, video overview, finding interesting content

1 INTRODUCTION

Panoramic cameras have gained increasing attention since the start of the 21st century. While the technology was in its infancy by the end of the 90's, allowing only for the exploration of a panorama of stitched still images, it has since evolved to allow consumers to enjoy full-fledged 360-degree videos. With the introduction of consumer-level virtual reality headsets in recent years such as the Oculus Rift, the HTC Vive and the Google Cardboard, as well as support for streaming 360-degree videos on video sharing platforms such as YouTube, panoramic videos have become readily available for consumption by the general public. However, little research has been done when it comes to the most effective or desired ways to interact with this new media form.

One example of such interactions is quickly browsing a 360-degree video to get a general overview of the video's content. In "regular" 2D video, this is easy: the user simply clicks on a point on the timeline and the video advances to that point. Alternatively, the user can set the playback speed to a higher value and advance through the video more quickly that way, or even press the left and right arrow keys to jump back and forth by a certain amount of seconds. However, in 360-degree video, how does one navigate to different points on the video's timeline while still getting to see all of the content? The user is restricted to seeing the content within the field of view (FOV) of the device they are using and will generally not have enough time to look around the entire environment while browsing.

Another example is finding scenes containing objects or areas of interest within the 360-degree video. We again compare to 2D video, where quickly finding a particular scene is as easy as looking at the screen while jumping through the timeline or playing the video back at a higher speed, until the desired scene is found. In 360-degree video, on the other hand, how does the user know they're looking at the right part of the video? The relevant part of the video that they're trying to find might be in the exact opposite direction of where they're looking.

In this study, we will introduce four different techniques that allow the user to get a good overview of a 360-degree video's content, as well easily finding scenes of interest. We will evaluate and compare their performance at these two tasks based on an analysis of quantitative and qualitative data that we have collected in a user test. The quantitative data is a combination of objective data collected from the user's performance at several tasks using the different techniques, as well as subjective data collected from a questionnaire. The qualitative data was collected by eliciting user feedback during the tests as well as a short debriefing interview after the experiment.

We aim to address the following research question:

- Is it possible to develop a technique that allows users to efficiently browse through 360-degree videos?

2 RELATED WORK

This research is related to studies on how to guide the user's attention in 360-degree video content, as we had to consider how to allow the user to see the most important parts of the video to get the best general overview of a video's content while quickly browsing through the video. Another related area is amplified head rotation, as that allows the user to more easily and efficiently navigate through the virtual environment with reduced head motion (and therefore see more of the video's content). Futhermore, it is important to consider research in motion sickness caused by VR headsets to make sure that our techniques do not cause a significant negative effect on the user's comfort.

Guiding User Attention

The main difference between 360-degree videos and "regular" 2D video is the added dimension of orientation. This introduces a number of problems related to guiding the user through the video content. One such problem is guiding the user's attention in the virtual environment. Since users are free to look wherever they want, it is important for producers of 360-degree videos to be able to to direct the user's attention towards important areas or objects in the video.

Nielsen et al. [1] performed a user study that evaluates the effects of implicit diegetic (part of the virtual environment) cues, as well as explicit non-diegetic cues (not part of the virtual environment) on guiding the user's attention. The results suggest that subtle diegetic cues may be more effective when it comes to guiding the user's attention while maintaining a sense of presence in the VE than explicit (non-)diegetic cues.

Grogorick, Stengel and Magnor [2] further investigate the effectiveness of subtle diegetic cues on attention direction in their study. They make the distinction between static virtual environments that do not require many head movements from users, and dynamic environments that do require a significant number of head movements. Results indicate that gaze guidance in the form of a subtle visual cue is especially effective in scenarios where the object to find is not immediately distinguishable from its direct surroundings.

Rothe, Huβmann and Allary [3] also conducted research on the effectiveness of diegetic cues for gaze guidance. These cues include stationary lights, moving lights, and sounds, as well as objects appearing out of nowhere. Results show that moving lights and objects were effective at garnering the user's attention, whereas non-moving lights/objects had no effect at all. Furthermore, the experiment showed that guiding attention at the start of a new scene was very difficult. A likely explanation for this is that the user wants to take in the new environment before they focus their attention on any moving objects.

Lin et al. [4] propose a visualization technique designed the guide the user's attention to points of interest (POI) that are outside of the FOV. The basic idea is to generate a small picture preview in the peripheral area of the screen in the direction of the POI. Results show that the this visualization technique outperforms a standard pointing arrow technique when it comes to giving the user spatial awareness of the POI's location, and there is no noticeable difference in the amount of interference (in terms of sense of presence in the virtual environment) experienced between the two techniques.

Sassatelli et al. [5] take gaze guidance research in a different direction by proposing a technique called Snap-Changes that moves the user's orientation directly onto a POI if the angle between the POI and user gaze is large enough, effectively "snapping" it into position. Results show that users spent less time moving their heads with the Snap-Changes technique than without. Re-positioning the user in front of the POI seemed to incite them to keep their focus on it, which leads to a significantly higher amount of time spent looking at these POIs.

While not specifically aimed at 360-degree videos, Waldin, Waldner and Viola [6] propose a method for gaze guidance in images using high-frequency flicker. The idea behind it is that the flicker is not noticeable when looking directly at it if operating at a frequency between 60 and 72 Hz, but it will draw the user's attention if located in the peripheral part of the FOV. Results showed that users can be effectively guided towards POIs in the virtual environment using the flicker technique.

Amplified Head Rotation

When watching 360-degree videos while seated on a fixed chair or couch, it may be preferable to manipulate the field of view in the virtual environment using reduced body and head motion to allow the user to see more of the virtual environment with less effort. One way to achieve this is to amplify the head rotation in the real world to produce a larger rotation in the virtual world. This is almost exclusively done for rotations in the yaw-axis, therefore, any mention of an augmented rotation in this section pertains to augmented rotations in the yaw-axis unless stated otherwise.

Sargunam, Moghadam, Suhail and Ragan [7] conducted an experiment where they evaluated amplified head rotation and compared them to regular head-tracked viewing. The implementation uses a scaling amplification factor. The angle between the direction in which the user is looking and the predetermined "forward direction" is multiplied by this amplification factor to determine the angle of rotation in the virtual world. A user study found that while the objective test results on spatial orientation did not indicate significantly worse performance for the amplified head rotation technique, the subjective user ratings indicated that the users experienced more difficulty navigating and orientating themselves with the amplified head rotation. The study found no significant effect on motion sickness experienced by gamers, however, there was a significant effect on the sickness experienced by non-gamers.

Ragan, Scerbo, Bacim and Bowman [8] performed a similar experiment with amplified head rotation that evaluated performance on 3D search tasks, spatial orientation and motion sickness. Results show that differences in display type, amplification factor and visible range may influence the effectiveness of the amplified head rotation technique. Contrary to [7], no evidence of a noticeable increase in motion sickness was found.

Hong and Kim [9] further affirm that there is no significant increase in motion sickness when using either constant amplification or dynamic (scaling) amplification. The user study found that users did not notice any significant increase in the rotation for either constant amplification or dynamic amplification, but test results did show a noticeable decrease in the amount of movement from the test participants' heads, indicating that if given an appropriate amplification level, the amplified head rotation technique succeeds in creating a lower-effort interaction without sacrificing comfort or accuracy.

Sellén [10] evaluated the effectiveness of amplified head rotation in regard to object selection at different distances and angles (from the "forward direction") and at different scaling factors for the head rotation amplification. The study found that a moderate amplification factor (2x) was preferable to both no amplification (too slow) and higher amplification factor (3x), which felt uncomfortable for the object selection task, according to test participants.

Kopper, Stinson and Bowman [11] ran an experiment similar to aforementioned articles, aimed at understanding the effects of amplified rotation on visual scanning tasks as well as counting tasks, but their contribution differentiates itself by the fact that they evaluated the effects of varying display characteristics; specifically, a varying FOV. Test results revealed that with a FOV of 102 degrees, no amplification factor (1x, 2x, or 3x) made a significant impact on the accuracy for the visual scanning task. With a small FOV of 30 degrees, accuracy increased with increased amplification. For a FOV of 52 degrees, moderate amplification (2x) increased accuracy, while high amplification (3x) decreased it. For the counting task, test results showed that moderate amplification performed about as well as the 1-to-1 view, whereas high amplification decreased performance significantly.

An interesting experiment related to [11] was performed by Bolte, Bruder, Steinicke, Hinrichs and Lappe [12], who conducted a psychological experiment to identify how much amplification can be applied to rotations, but in the pitch- and roll-axii rather than the usual yaw-axis, as well as the effect of different sizes for the FOV. They found that moderate augmentation (amplification factor of 1.3x for pitch and 1.44x for roll) will go unnoticed by users. This is consistent with the findings of the aforementioned papers for yaw-axis rotation amplification, where users reported high amplification as feeling uncomfortable and moderate amplification as natural.

Motion Sickness

Research has shown that in some cases, head-mounted VR can cause discomfort in the form of nausea and dizziness, particularly if the difference between the movement in the virtual environment and the real world is significant [13][14]. This can be the case with amplified head rotations, however, as shown by [9], [10] and [11], this is negligible if the amplification is kept at a moderate level.

The findings in these works were all taken into account when designing the techniques discussed in Section 4.

3 PROBLEM

In traditional 2D video, if a user wants to get a quick glimpse of what the video has to offer in terms of content, they can achieve this quite easily. They can simply hover the mouse over different parts of the video timeline and click to jump to those specific timestamps. Alternatively, they can press the left and right arrow keys to jump back and forth by a set amount of seconds (usually 5 or 10 depending on the video platform). In most cases, they can even set the video playback to a higher speed to move through the video quickly. All of these things can be done while the user keeps their eyes on the video itself. Finding a specific scene of interest is also easily done in traditional video: the user simply moves through the video quickly using any of the aforementioned techniques until the scene appears on the screen. We define being able to do both of these things efficiently as being able to efficiently browse a video.

In the case of 360-degree video, it is not so simple. Instead of having the entire video in front of them, the user is limited to seeing only a small part of the virtual environment; they are restricted by the FOV of the device that they are using. How does one skip or fast-forward through a 360-degree video quickly while still getting to see most or all of the content (that is, while still getting a "good overview" of the video)? We will call this problem "the overview problem".

Also, how does one find a scene containing an object or point of interest (POI) quickly within the 360-degree video? The POI could be outside of the FOV while skipping through the video, causing the user to completely miss it. This makes the problem significantly more complex than in the 2D case. We will refer to this problem as "the scene-finding problem".

By our definition of "efficiently browsing videos" (which can be extended to the 360-degree case in an arbitrary manner), it follows that solving these two problems allows the user to efficiently browse 360-degree videos. Therefore, we will focus on addressing these two problems to address our research question.

In this study, we will focus specifically on 360-degree video watched through an HMD, though we believe the results are relevant for other types of VR as well (such as mobile VR). Our goal for this study is to develop a technique that solves both of the problems mentioned above, while having no negative effect on the user's discomfort (in terms of motion sickness). The design of the techniques is discussed in the next section.

4 SOLUTION

To tackle the overview problem and the scene-finding problem described in the previous section, we came up with a number of requirements for the design of the techniques. They are as follows:

(1) The interface must allow the user to maintain a good overview of the video content.
(2) The interface must allow the user to quickly move to different parts of the video.
(3) The interface must be easy and intuitive to use.

We settled on four different techniques that we believe fulfill these requirements. Before going into a detailed description of each of them, we will explain how we decided on these techniques.

To fulfill requirement (1), we initially wanted to use some kind of gaze guidance technique as described in the Related Works section to guide the user's attention to important areas of the video (POIs). While designing a technique similar to the one proposed in [4], we discovered that mapping the entire virtual environment of a 360-degree video to a large rectangle that fits in the user's FOV (as shown in Figure 1, effectively creating an "animated panoramic image") works very well for allowing the user to get a good overview of the entire video. No longer is the user restricted to seeing only a small part of the virtual environment that is within the FOV, rather we've brought the entire virtual environment into the FOV. We refer to these mapped rectangles as "Large Thumbnails", as they resemble the thumbnails one sees when hovering the mouse over the timeline of a YouTube video. We decided to go with these Large Thumbnails instead of using gaze guidance.

We had also developed an amplified head rotation technique that amplifies the head rotation in the virtual environment by two times the head rotation in the real world, meaning that the user only had to look 90 degrees to the left or right to achieve a 180 degree rotation in either direction in the virtual world. This allowed the user to easily observe the entire virtual environment by looking left and right. At first, we wanted to compare the amplified head rotation technique to the Large Thumbnail technique described above to see which allowed the user to get a better overview of 360-degree video, but after getting very positive feedback from users on the Large Thumbnail technique during preliminary user tests, we chose to just use the Large Thumbnail technique and leave amplified head rotation out of the equation.

To fulfill requirement (2), we came up with two different modes for skipping through the video: the Timeline mode that allows the user to jump to different points on the timeline, and the Fast-Forward mode that allows the user to fast-forward (or rewind) through the video.

For requirement (3), we settled on two different ways to control the interface: the Controller mode, where the user interacts with the interface through a wireless controller, and the Gaze mode, where the user simply uses gaze interaction to interact with the interface. The controller we used in the user test is the 8bitdo SN30PRO (depicted in Figure 2).

Figure 2: 8bitdo SN30Pro wireless gamepad

Ultimately we ended up with 4 different techniques. They are the Timeline/Fast-Forward modes combined with Controller/Gaze modes. All of them were also combined with the Large Thumbnail technique. In the analysis later on, we will evaluate how well each technique addresses the overview problem and the scene-finding problem, and also how easy they are to use. We will now describe each of them individually.

Timeline Controller (TC)

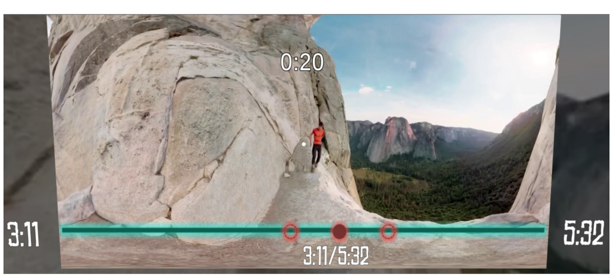

Figure 3: Timeline Controller technique (Large Thumbnail depicted as well, which has the entire 360-degree video mapped to it. Note that we left in the original virtual environment in the background with a darker hue for a nice visual effect)

The Timeline Controller technique (Figure 3) allows the user to manipulate the current position on the video timeline by using the wireless controller. The current time in the video is depicted by the red filled circle on the timeline with the timestamp underneath it.

Two empty circles are depicted on the timeline as well, to the left and right of the filled circle. We will call these the "jump indicators". The purpose of these jump indicators is to indicate the position to which the player can jump by pressing either the left or the right directional buttons on the wireless controller (the black buttons on the left side of the gamepad depicted in Figure 2. Pressing the right directional button will move the red circle (and therefore the current time in the video) to the position of the right jump indicator. Pressing the left directional button will move the video to the position of the left indicator.

Users can also change the distance of the jump indicators to the red circle. Pressing the up directional button will move the jump indicators away from the red circle. Pressing the down directional button will move the jump indicators toward the red circle. This way the user can easily decide if they want to make big or small jumps through the video when pressing the left and right directional buttons.

Timeline Gaze (TG)

Figure 4: Timeline Gaze technique

The Timeline Gaze technique allows the user to manipulate the current position on the video timeline by simply gazing at the timeline. The forward gaze is indicated by the white dot in Figure 4. The red circle indicating the current time in the video will immediately jump to the white dot if placed over the timeline. This technique is the main reason why we decided to place the video timeline slightly higher up on the bottom side of the Large Thumbnail instead of directly below it. The reasoning is that having to look down towards the timeline to jump to a different time in the video causes a large part of the Large Thumbnail to move out of the FOV at the top, which means the player will not be able to see that part of the 360-degree video. We tried to alleviate this by moving the timeline up slightly.

Fast-Forward Controller (FC)

The Fast-Forward Controller technique allows the user to fast-forward and rewind through the video at a much higher speed than the regular playback speed. We call it "Fast-Forward" for short, but it should be noted that it also allows for rewinding (which is essentially fast-forwarding in reverse). Three buttons are depicted above the timeline as shown in Figure 5: the regular playback button in the center, the fast-forward button on the right and the rewind button on the left.

Figure 5: Fast-Forward Controller technique in rewind mode

If the user is in normal playback mode (indicated by the middle button being white), the user can activate the fast-forward mode by pressing the right shoulder button on the controller (the top right white button on the back of the gamepad depicted in Figure 2). Pressing the left shoulder button in fast-forward mode will return the video to normal playback mode. Pressing the left shoulder button in normal playback mode will activate rewind mode. Pressing the right shoulder button in rewind mode will return the video to normal playback mode.

The playback speed for both fast-forward mode and rewind mode is the same and has been calculated as being the speed that allows the user to move through the entire video within 30 seconds. That means the user can essentially fast-forward through the entire video and then rewind back through the entire video in 1 minute (which is the time limit for the tasks in our experiments). This translates to a playback speed of roughly 12x normal playback speed for video around 6 minutes in length (which is about the running time of the videos used in our experiments). We felt this was an appropriate speed that allows the user to get a decent overview of the video.

Fast-Forward Gaze (FG)

Figure 6: Fast-Forward Gaze technique in fast-forward mode

The Fast-Forward Gaze technique allows the user to fast-forward and rewind through the video at a much higher speed than the regular speed. It is much the same as FC except that the user activates the different modes simply by gazing at the buttons (depicted in Figure 6). The user can take their gaze off the buttons after activating them.

5 USER STUDY

To evaluate the effectiveness of the techniques at addressing the overview and scene-finding problems, we conducted a user study where test participants were given an "overview task" as well as a "scene-finding task", for each technique. After each task, we collected quantitative data by recording data from the participants' performance on the tasks, as well as through a questionnaire. Qualitative data was collected at the end of each experiment, through a short debriefing interview.

Test Participants

We recruited 32 participants of which 16 were male and 16 were female, aged 18 to 30 years old. These participants were all recruited at Utrecht University. Almost all of them had little to no experience with head-mounted VR. We split the 32 participants into two groups of 16, where one group always started with the two Timeline techniques, and the other group always started with the two Fast-Forwarding techniques. This way we could incorporate one tutorial for each pair of techniques instead of heaving to explain each technique individually, as would have been the case had we decided not to segregate the two types of techniques. We further subdivided each of the two groups into 4 groups of 4 participants that shared the same unique technique order.

We used four different videos (we chose to test only the visual part of the videos, so we disabled the audio) to test the techniques on. These were all some kind of "virtual tour" video: a city tour in London, a city tour in Rome, an island tour and a mountain climbing tour. A Latin Square assignment was used on each group of 4 participants for the order of the videos to ensure that no two participants had the same combination of video and technique orders. Finally, the order of participants was also randomized. The final experiment order can be seen in in Figure 7.

	TIMELINE TECHNIQUES FIRST					FAST-FORWARD TECHNIQUES FIRST			
	TC	TG	FC	FG		FC	FG	TC	TG
Participant 1	Video B	Video C	Video D	Video A	Participant 16	Video D	Video C	Video B	Video A
Participant 12	Video D	Video A	Video C	Video B	Participant 25	Video C	Video D	Video A	Video B
Participant 28	Video C	Video B	Video A	Video D	Participant 8	Video B	Video A	Video D	Video C
Participant 24	Video A	Video D	Video B	Video C	Participant 32	Video A	Video B	Video C	Video D
	TG	TC	FG	FC		FG	FC	TG	TC
Participant 19	Video D	Video A	Video B	Video C	Participant 21	Video B	Video C	Video A	Video D
Participant 30	Video B	Video C	Video D	Video A	Participant 10	Video A	Video D	Video B	Video C
Participant 27	Video C	Video D	Video A	Video B	Participant 14	Video C	Video B	Video D	Video A
Participant 22	Video A	Video B	Video C	Video D	Participant 26	Video D	Video A	Video C	Video B
	TG	TC	FC	FG		FC	FG	TG	TC
Participant 29	Video D	Video C	Video A	Video B	Participant 23	Video A	Video B	Video C	Video D
Participant 20	Video C	Video B	Video D	Video A	Participant 13	Video D	Video A	Video B	Video C
Participant 4	Video A	Video D	Video B	Video C	Participant 3	Video C	Video D	Video A	Video B
Participant 7	Video B	Video A	Video C	Video D	Participant 11	Video B	Video C	Video D	Video A
	TC	TG	FG	FC		FG	FC	TC	TG
Participant 2	Video A	Video C	Video B	Video D	Participant 31	Video C	Video A	Video D	Video B
Participant 9	Video D	Video A	Video C	Video B	Participant 17	Video D	Video B	Video A	Video C
Participant 15	Video C	Video B	Video D	Video A	Participant 5	Video A	Video C	Video B	Video D
Participant 18	Video B	Video D	Video A	Video C	Participant 6	Video B	Video D	Video C	Video A
TC = Timeline Controller					Video A = Rome				
TG = Timeline Gaze					Video B = FreeSolo				
FC = Fast-Forward Controller					Video C = London				
FG = Fast-Forward Gaze					Video D = Underwater				

Figure 7: Experiment order

Experiment Protocol

At the start of the experiment, the test participant was asked to read and sign a consent form. After signing their consent, they were asked to fill in some basic information about their background/experience with head-mounted VR. Then, they were briefed about the experiment and the tasks to come and asked to put on the HMD. Depending on which pair of techniques they would start

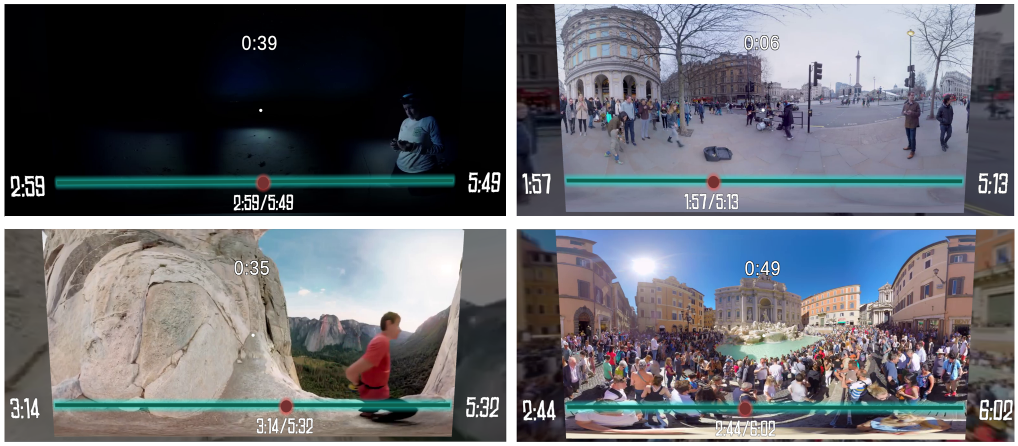

Figure 8: The scenes to find for the scene-finding task. Clockwise from top-left - (1) island tour: turtles on beach (2) London tour: street band (3) mountain climbing tour: red glowing mountain in background (4) Rome tour: Trevi Fountain

with (Timeline or Fast-Forward), they would be walked through a short tutorial on the controls for those techniques, before starting with the first task for the first technique.

The first task was the overview task. For this task, participants were instructed to try to get the best overview possible of the content for a 5-6 minute video, by applying the technique that was currently being evaluated. They were told that they would be queried about what they've seen afterwards. They had one minute to complete this task. After the task, they were asked if they remembered a specific scene from the video that contained an object that would be easily noticeable. These scenes are displayed in Figure 8. The rationale behind this is that if the participant got a good overview of the video, then they would remember such a scene and we could use this data to evaluate how well the technique allows the user to get a general overview of the video.

The second task was the scene-finding task. For this task, we instructed the user to find back in the video the scene that we asked them about in the overview task (regardless if they remembered it or not). They also had one minute to complete this task. We measured the time it took them to find the scene to evaluate how well the technique allows them to find specific content in the video.

These two tasks were then repeated for the second technique and video. After the second technique/video, a second tutorial was given to explain the controls of the remaining two techniques (Fast-Forward techniques if the first two techniques were Timeline and vice versa). After this tutorial, the last two techniques were tested in the same manner. After testing the fourth technique, the user was asked to take off the HMD and a short debriefing interview was conducted to get qualitative feedback on the techniques.

6 RESULTS & ANALYSIS

We wish to answer the following research question:

- Is it possible to develop a technique that allows users to efficiently browse through 360-degree videos?

To answer this research question, we propose the following two hypotheses to validate:

(1) Both Fast-Forward Controller and Fast-Forward Gaze techniques allow the user to keep their eyes on the 360-degree video while moving through it quickly and will therefore outperform and be preferred over the two Timeline techniques for getting a general overview of the content.

(2) Both Timeline Controller and Timeline Gaze techniques allow the user to quickly jump to specific timestamps in the 360-degree video and will therefore outperform and be preferred over the two Fast-Forward techniques for finding specific scenes.

If both these hypotheses can be validated, it follows that a combination of both Timeline and Fast-Forward would be ideal for addressing the overview problem and the scene-finding problem, and would therefore allow for efficient browsing of 360-degree videos.

We will first discuss the analysis of the objective data we collected and then move on to the analysis of the subjective data collected from the questionnaires.

Objective data collected from overview task

For the overview task, we asked the user if they remember seeing a specific scene in the video (Figure 8, scenes were never more than 30 seconds away from the halfway point in the video to ensure that it will be roughly equally challenging to find the scene in any of the four videos).

We wanted to measure the number of remembered scenes proportional to the total number of scenes, to evaluate how well the user got a general overview of the video using the technique. The numbers are as follows:

	TC	FC	TG	FG
REMEMBERED SCENE	38%	57%	65%	65%

We then tested for statistically significant differences between these groups. We found two pairs that showed significant differences: *TC/TG* and *TC/FG*. An example of the process is shown in Figure 9.

TC/TG

Calculate sample proportions: $p_1 = 10 / 26 = 0{,}38$
$p_2 = 17 / 26 = 0{,}65$

Take difference: $p_2 - p_1 = 0{,}65 - 0{,}38 = 0{,}27$

Calculate overall sample proportion: $p = \frac{10+17}{26+26} = 0{,}52$

Calculate standard error: $\sqrt{0{,}52(1 - 0{,}52)(\frac{1}{26} + \frac{1}{26})} = 0{,}14$

Calculate test statistic: 0,27 / 0,14 = 1,93

From z-table it follows that p = 1 - 0,9713 = 0,0287 which is less than 0,05 so there is a statistically significant difference

Figure 9: Testing for significant differences in number of remembered scenes for TC/TG

These results are unexpected, as we hypothesized that both Fast-Forward techniques would outperform the Timeline techniques, yet only one of them does, and one of the Timeline techniques outperforms the other.

Objective data collected from scene-finding task

For the scene-finding task, we measured the amount of time it took the participant to find a specific scene (unique for each video). Figure 10 shows the average completion time per technique for this task. The initial impression from this graph is that TG outperforms the other techniques by a significant margin.

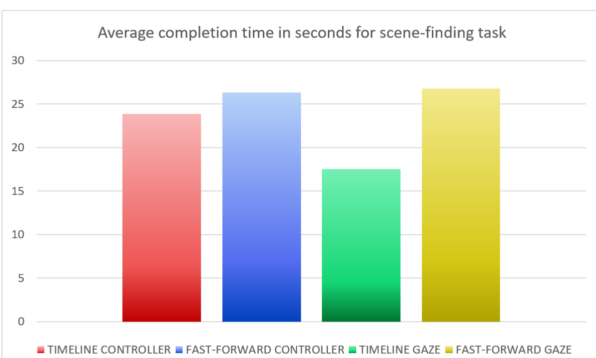

Figure 10: The average completion time in seconds per technique for the scene-finding task

Figure 11 displays four histograms that show the frequency of completion times per technique for the scene-finding task. The histograms show that for both Timeline techniques, higher frequencies are concentrated more towards the left side of the graph, indicating generally faster task completion times. For the Fast-Forward techniques, the higher frequencies are concentrated more towards the center, indicating generally poorer performance.

To prove that there are significant differences between the means of these four groups, we performed a Kruskal-Wallis test on the four data groups. The results show that there are indeed significant differences between the group means. To find where the differences in the groups reside, we then performed a Conover test as a post-hoc, which tests the difference in the means between all pairs of groups. We decided on using Kruskal-Wallis-Conover because they are better equipped to deal with uneven sample sizes than a standard ANOVA and Tukey-Kramer test [15][16]. The results are shown in Figure 12.

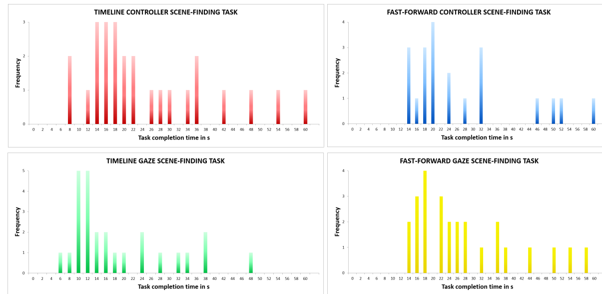

Figure 11: Histograms of completion time in seconds for the scene-finding task

group 1	group 2	R-mean	std err	t-stat	R-crit	p-value
TC	FC	7,224358974	8,421578384	0,857839071	16,7189	0,39314
TC	TG	15,67307692	7,961029101	1,968724988	15,8046	0,0519
TC	FG	10,05769231	7,961029101	1,263365851	15,8046	0,20955
FC	TG	22,8974359	8,421578384	2,718900763	16,7189	0,00779
FC	FG	2,833333333	8,421578384	0,336437328	16,7189	0,73728
TG	FG	25,73076923	7,961029101	3,23209084	15,8046	0,00169

Figure 12: The results from the Kruskal-Wallis test followed by a Conover test

Two of the pairs show a significant difference in the means; the TG/FG pair and the TG/FC pair. The remaining four pairs show no significant differences in the means. The results indicate that TG outperforms both Fast-Forward techniques, as the average completion time for that technique is significantly lower. However, TC does not outperform any of the other techniques.

User scores

Other than the objective data that we showed in the previous subsections, we also collected data using a questionnaire. The questionnaire results can be seen in Figure 13.

		TC	FC	TG	FG
AVERAGE	OVERVIEW SCORE	3,59	3,97	3,47	3,97
	OVERVIEW EASE	4,16	4,50	3,56	4,16
	SCENE-FINDING EASE	4,28	4,28	4,31	3,88
	DISCOMFORT SCORE	1,03	1,09	1,09	1,06

Figure 13: The results from the questionnaire

These are the average scores assigned by our test participants. The questionnaires are based on a 5-point Likert scale. The questions associated with these scores are respectively:

(1) How confident are you that you got a good overview of the video content (5 is very confident)?

(2) How easy did you find it to use this technique to get an overview of the video (5 is very easy)?
(3) How easy did you find it to use this technique to find a specific scene in the video (5 is very easy)?
(4) Did you experience any nausea or dizziness (5 is the highest level of discomfort)?

We have performed Kruskal-Wallis tests on all of the scores to test for significant differences.

Overview score. At a first glance, we can already tell that both Timeline techniques have been rated lower for getting a good overview than the Fast-Forward techniques. This would support H1. TG also seems to be rated significantly lower when it comes to ease of use for getting a good overview. A Kruskal-Wallis test on the overview scores reveals no significant differences between the four groups, however.

Overview ease. Kruskal-Wallis showed significant differences exist between the groups for the overview ease score. The results of the Conover are shown in Figure 14. The results show that FC has been rated significantly higher than TG for ease of use for getting a good overview of the video content. There are no significant differences between the other groups, however.

group 1	*group 2*	*R-mean*	*std err*	*t-stat*	*R-crit*	*p-value*
TC	FC	11,875	9,270514729	1,280942898	18,34894547	0,202604271
TC	TG	17,96875	9,270514729	1,938268858	18,34894547	0,054863108
TC	FG	0,53125	9,270514729	0,05730534	18,34894547	0,954394207
FC	TG	29,84375	9,270514729	3,219211756	18,34894547	0,001640981
FC	FG	12,40625	9,270514729	1,338248238	18,34894547	0,183264306
TG	FG	17,4375	9,270514729	1,880963518	18,34894547	0,062322138

Figure 14: Results of Kruskal-Wallis-Conover tests on the overview ease score

Scene-finding ease. For finding specific scenes, FG has been rated much lower than the other techniques. FC has been rated just as high as the two Timeline techniques though, which is unexpected. Kruskal-Wallis-Conover shows no significant differences between the four groups for this score, however.

Discomfort score. The discomfort was almost unanimously rated as 1 for all techniques, with the occasional 2 for a very small number of test participants. For completeness, we performed a Kruskal-Wallis on the discomfort scores and as expected, the results show no significant difference in the means.

7 DISCUSSION

While the quantitative data seems to suggest that the Timeline techniques perform better at the scene-finding task and the Fast-Forward techniques are better for getting a general overview of the video content, the results from the statistical analysis show that the differences between the four techniques in these regards are not significant enough to be able to validate either hypothesis.

Hypothesis 1 states that both Fast-Forward Controller (FC) and Fast-Forward Gaze (FG) should show significantly better performance than both Timeline Controller (TC) and Timeline Gaze (TG) for the overview task. However, analysis on the objective data (number of remembered scenes vs total scenes) for this task shows that only FG significantly outperforms TC and somehow TG (and not FC) outperforms TC. It is unclear why TG performed so well in this task. TG requires the user to constantly look at the timeline to be able to jump to different points and should therefore perform worse because it takes attention away from the video. This is also what test participants said about TG. This might simply be a contingency, in which case a follow-up experiment would be appropriate to prove it as such.

Analysis on the OVERVIEW SCORE from the questionnaire shows no significant difference, indicating that participants felt that none of the techniques allowed them to get a significantly better overview of the video content. Analysis on the OVERVIEW EASE scores shows shows no significant differences between any of the techniques either, indicating that none of the techniques are significantly easier to use for the overview task. All of the techniques were highly rated, however (3.47 or higher), for both OVERVIEW SCORE and OVERVIEW EASE showing that they are all generally easy to use for this task and allow users to get a good overview.

Hypothesis 2 states that both TC and TG should show significantly better performance than both FC and FG for the scene-finding task. However, analysis on the objective data (task completion times) for this task indicates that only TG outperforms both FC and FG. No significant differences were found between TC and FC/FG. This is most likely because the TC technique is considerably more complex than the other techniques, requiring multiple inputs to move around the timeline, whereas the others are mainly single-input techniques.

Analysis on the SCENE-FINDING EASE questionnaire scores show no significant differences between any of the techniques, indicating that none of the techniques are significantly easier to use for the scene-finding task. All of the techniques were highly rated, however (3.88 or higher), showing that they all work pretty well for this task.

DISCOMFORT was rated as extremely low for all techniques (no higher than 1.09 on a 5-point scale). No significant differences were found between the means for this score. We can thus conclude that there is a negligible effect on the user's discomfort in terms of general nausea or dizziness, confirming that the techniques do not cause motion sickness.

While the objective data and user scores cannot confirm the hypotheses, qualitative feedback suggests that participants generally agree with them. The consensus among participants is that the Timeline techniques work best for the scene-finding task while the Fast-Forward techniques work best for getting a general overview of the content. A number of them suggested that a combination of both types of techniques would work best to tackle the 360-degree video browsing problem. Below, we provide some examples of comments from test participants. Note that "video" in the following comments refers to the Large Thumbnail directly in front of the user.

Participant 22 said about Timeline Controller: "It's easier to find a scene because you can jump to specific points on the timeline but I'm afraid of skipping parts of the video so this is effective for finding scenes but maybe not so much for getting an overview."

Participant 7 said about Timeline Gaze: "Works great for jumping back to a particular scene, but trying to get a good overview is less appealing because you have to look at the timeline which means you can't look at the video."

Participant 24 said about Timeline Controller: "Not having to look away from the video is better for overview, also wireless controller allows for better control over fast forwarding speed."

Participant 29 said about Fast-Forward Gaze: "In general I found fast-forwarding/rewinding easier because you got to see the entire video instead of jumping only to fragments."

Most participants preferred the techniques involving the wireless controller over the gaze interaction techniques (Figure 15). This is because the controller allowed the user to manipulate the current time in the video while still being able to look at the Large Thumbnail, whereas the gaze interaction forced them to look down at either the timeline or the Fast-Forward buttons.

Participant 4 said about the controller: "You could keep your eyes on the screen, you did not have to look away, so it was the best to get a complete overview."

Participant 27 said: "The gaze interaction requires your head to move around, which is not always comfortable or efficient."

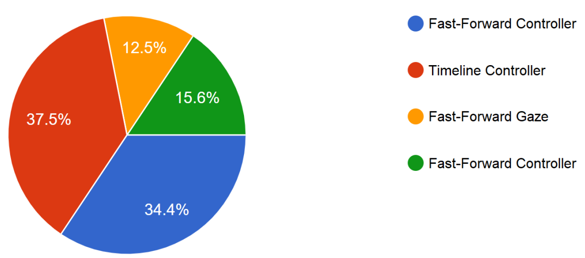

Figure 15: Pie chart of technique preference

8 CONCLUSION

We have brought to light two problems that occur while quickly browsing 360-degree videos that are not found in traditional 2D videos: the problem of maintaining a good overview of the content and the problem of trying to find a scene containing a specific object or area of interest in the video. Both problems exist because the FOV limits the user to only seeing a small part of the video at a time, causing them to overlook large parts of the video during browsing.

We introduced four techniques designed to deal with these two problems and we performed a user study to demonstrate that these techniques are successful in solving these problems. We distinguish between Timeline techniques and Fast-Forward techniques, where the Timeline techniques allow the user to select a point on the video timeline to jump to and the Fast-Forward techniques allow the user to fast-forward and rewind in the video. Both types of techniques have a wireless controller mode and a gaze interaction mode, resulting in 4 total techniques. All of them are used in conjunction with so called "Large Thumbnails" which are rectangular screens that fit in the user's FOV to which the 360-degree video is mapped, allowing them to see the video in its entirety.

While quantitative analysis cannot validate the hypotheses that we posed in this study, qualitative feedback shows that users agree that the Timeline techniques work better for finding specific points of interest in the video and the Fast-Forward techniques work better for getting a general overview of the video. All techniques were highly rated for both the overview and scene-finding tasks. Both interaction modes (controller and gaze) were highly rated, though most participants preferred the wireless controller. Multiple participants suggested after the experiment that using both Timeline and Fast-Forward techniques at the same time could be ideal for solving both problems. Therefore, we believe this study is a step in the right direction towards solving the 360-degree video browsing problem.

Future Work

For future work, it would be interesting to see how these techniques perform on longer videos. In this study, we used exclusively 5-6 minute videos, which works well for the Fast-Forward techniques, but we suspect it would get increasingly harder to see what's going on in the video when the playback speed in Fast-Forward/Rewind mode increases.

Furthermore, we used very similar videos for the user tests. All of them were some variation of a virtual tour (city tour of Rome, city tour of London, island tour and mountain climbing experience). In future work, we would like to see different genres in the selection of videos.

To ensure the interfaces were as simple and intuitive as possible, we made the controls very limited, especially for the Fast-Forward techniques (where there are only 3 modes: normal playback speed, fast-forward mode and rewind mode). It would be interesting to experiment with different speed settings for this mode. This was also a recurring suggestion from our test participants.

Perhaps we could have obtained more conclusive results if we had made a different kind of comparison. Our current results show that the techniques' performances are too close to one another to show any significant differences, but we think comparing the four techniques + Large Thumbnails to the four techniques without Large Thumbnails (which would then be considered the "normal case") would show significant differences in terms of performance for both overview task and scene-finding task.

We chose to disable the audio since we argued that it would add little value to the experiment as it is mostly unintelligible since participants are constantly cutting off the audio by jumping from point to point with the Timeline techniques and are also moving too fast through the videos with the Fast-Forward techniques to be able to properly process and understand the audio. Still, it could be worthwhile to consider adding audio in future research on 360-degree video browsing to see how that influences being able to get a good overview or finding specific scenes.

REFERENCES

[1] Lasse T. Nielsen, Matias B. Møller, Sune D. Hartmeyer, Troels C. M. Ljung, Niels C. Nilsson, Rolf Nordahl, and Stefania Serafin. Missing the point: An exploration of how to guide users' attention during cinematic virtual reality. *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, 2016.

[2] Steve Grogorick, Michael Stengel, and Marcus Magnor. Subtle gaze guidance for immersive environments. *Proceedings of SAP*, 2017.

[3] Sylvia Rothe, Heinrich Huβmann, and Allary Mathias. Diegetic cues for guiding the viewer in cinematic virtual reality. *Proceedings of the Second International Symposium of Chinese CHI*, 2014.

[4] Yung-Ta Lin, Yi-Chi Liao, Shan-Yuan Teng, Yi-Ju Chung, Liwei Chan, and Bing-Yu Chen. Outside-in: Visualizing out-of-sight regions-of-interest in a 360 video using spatial picture-in-picture previews. *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, 2017.

[5] Lucile Sassatelli, Anne-Marie Pinna-Déry, Marco Winckler, Savino Dambra, Giuseppe Samela, Romaric Pighetti, and Ramon Aparicio-Pardo. Snap-changes: a dynamic editing strategy for directing viewer's attention in streaming virtual reality videos. *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*, 2018.

[6] N. Waldin, M. Waldner, and I. Viola. Flicker observer effect: Guiding attention through high frequency flicker in images. *Computer Graphics Forum*, 2017.

[7] Shyam Prathish Sargunam, Kasra Rahimi Moghadam, Mohamed Suhail, and Eric D. Ragan. Guided head rotation and amplified head rotation: Evaluating semi-natural travel and viewing techniques in virtual reality. *IEEE Virtual Reality (VR)*, 2017.

[8] Eric D. Ragan, Siroberto Scerbo, Felipe Bacim, and Doug A. Bowman. Amplified head rotation in virtual reality and the effects on 3d search, training transfer, and spatial orientation. *IEEE Transactions on Visualization and Computer Graphics*, 2017.

[9] Seokjun Hong and Gerard J. Kim. Accelerated viewpoint panning with rotational gain in 360 degree videos. *VRST '16*, 2016.

[10] Jonna Karlsson Sellén. Performance of amplified head rotation in virtual reality under different object distances. 2017.

[11] Regis Kopper, Cheryl Stinson, and Doug A. Bowman. Towards an understanding of the effects of amplified head rotations. 2012.

[12] Benjamin Bolte, Gerd Bruder, Frank Steinicke, Klaus Hinrichs, and Markus Lappe. Augmentation techniques for efficient exploration in head-mounted display environments. 2010.

[13] Sue VG Cobb, Sarah Nichols, Amanda Ramsey, and John R Wilson. Virtual reality-induced symptoms and effects (vrise). *Presence: Teleoperators & Virtual Environments*, 8(2):169–186, 1999.

[14] Emilee Patrick, Dennis Cosgrove, Aleksandra Slavkovic, Jennifer A Rode, Thom Verratti, and Greg Chiselko. Using a large projection screen as an alternative to head-mounted displays for virtual environments. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 478–485. ACM, 2000.

[15] Gregory W Corder and Dale I Foreman. Nonparametric statistics for non-statisticians, 2011.

[16] William Jay Conover and Ronald L Iman. On multiple-comparisons procedures. *NM, USA: Los Alamos Scientific Lab*, pages 1–17, 1979.

[17] A london city guided tour (360 vr video).

[18] Rome: A guided city tour (360 vr video).

[19] Free solo 360 | national geographic.

[20] 360 underwater national park | national geographic.

4 Literature Review

This section contains the literature review that was written as a part of this master's thesis. It should be noted that the subject matter of the study conducted for this master's thesis has changed considerably since compiling the literature review and therefore some parts of it are no longer relevant. A large part of it has still been incorporated into the scientific paper, however.

Literature Review

Peter de Keijzer

May 13, 2019

Abstract

360-degree videos are a new type of media that have gained increasing attention in recent years. This study will provide research into the different possibilities for interacting with 360-degree videos and aims to uncover the most desired and effective ways for users to interact with this new media form. It will serve as a basis for a contribution in the form of a scientific survey paper that evaluates different aspects such as user experience, task performance, comfort and self-orientation.

1 Introduction

Panoramic cameras have gained increasing attention since the start of the 21st century. While the technology was in its infancy by the end of the 90's, allowing only for the exploration of a panorama of stitched still images, it has since evolved to allow consumers to enjoy full-fledged 360-degree videos. With the introduction of entry-level virtual reality headsets in recent years such as the Oculus Rift, the HTC Vive and the Google Cardboard, as well as the support for streaming 360-degree videos on video sharing platforms such as YouTube, panoramic videos have become readily available for consumption by the general public. However, little research has been done when it comes to the most effective or desired ways to interact with this new media form.

One example of such interactions is fast-forwarding or skipping through a 360-degree video. How does one choose a point on the timeline to look at when there is a full 360-degree field of view to explore at all times? This also begs the question: how does one create a summary (e.g. a movie trailer) for 360-degree videos? The added dimension of movement makes it hard to present the videos' most interesting parts to the user.

On a related note, highlighting or directing attention towards the interesting points in a 360-degree video is a matter entirely of its own. Given the user's freedom to look in any direction they want, how does the director draw the user's attention towards an important point in the video without intruding on the user's sense of presence?

Lastly, while not necessarily an issue for virtual reality games where the user is expected to move around, users of 360-degree videos, and especially those seated on a fixed chair or sofa, could prefer interacting with and navigating in the virtual environment in ways that reduce the amount of required body motion and head rotation for increased comfort.

This study will provide research into the different possibilities for interacting with 360-degree videos and aims to discover the most desired and effective ways for users to interact with this new media form. It will serve as a basis for a contribution in the form of a scientific paper that evaluates different aspects such as user experience, task performance, comfort and self-orientation.

2 Literature review

The literature review is structured as follows: it is divided into the three types of interaction that we want to research, with each section containing an analysis of relevant papers for that particular area, followed by a summary and any conclusions that might be drawn from the individual literature reviews.

2.1 Interactively manipulating field of view

When watching 360-degree videos while seated on a fixed chair or couch, it may be preferable to manipulate the field of view in the virtual environment using reduced body and head motion to ensure low-effort interaction and increased comfort, while making sure that any potential motion sickness remains at an acceptable level. One way to achieve this is to amplify the head rotation in the real world to produce a larger rotation in the virtual world. This is almost exclusively done for rotations in the yaw-axis, therefore, any mention of an augmented rotation in this section pertains to augmented rotations in the yaw-axis unless stated otherwise.

Sargunam, Moghadam, Suhail and Ragan (1) conducted an experiment where they evaluated amplified head rotation and compared them to regular head-tracked viewing. The implementation uses a scaling amplification factor. The angle between the direction in which the user is looking and the predetermined "forward direction" is multiplied by this amplification factor to determine the angle of rotation in the virtual world. The amplification factor is chosen so that the rotation is small when the user is close to the forward direction and increases the farther the user's gaze moves away. A user study found that while the objective test results on spatial orientation did not indicate significantly worse performance for the amplified head rotation technique, the subjective user ratings indicated that the users experienced more difficulty navigating and orientating themselves with the amplified head rotation. The study found no significant effect on the sickness experienced by gamers, however, there was a significant effect on the sickness experienced by non-gamers. Furthermore, users indicated that they preferred standard head rotation to amplified head rotation because having to turn their head without moving their body caused their necks to strain. The option to change the forward direction would alleviate this issue and may influence the user's preference. The paper makes no mention of having tested this, however.

Ragan, Scerbo, Bacim and Bowman (2) performed a similar experiment with amplified head rotation that evaluated 3D search, spatial orientation and motion sickness, and reported similar results. However, they did not test specifically for seated usage of an HMD (head-mounted display) and opted instead to test for standing HMD usage as well as a CAVE display system. The HMD display was modified by adding padding to the outer range of the FOV (field of view) to simulate the walls of the CAVE system. They used a fixed scalar for the amplification factor with a different value depending on the test case. Three different test scenarios were created; one with no amplification on the head rotation, one with a multiplication factor of 1.5 over a physical range of 270 degrees, and one with a multiplication factor of 4.0 over a range of 90 degrees. The user study found that differences in display type, amplification factor and visible range may influence the effectiveness of the amplified head rotation technique, but the study was limited in that the authors modified the HMD to more resemble the CAVE use case which begs the question whether test results for the HMD would have been different if used in its original state. Furthermore, the study states in its conclusion that amplified head rotation reduces "realism" but no evidence is provided for this. Contrary to (1), no evidence of a noticeable increase in motion sickness was found. However, it is not stated how much experience the test participants had with VR (Virtual Reality) or even gaming and thus if the lack of motion sickness could be attributed to this. It is also unclear if and how the fact that the test participants were standing and were able to move their whole body rather than just their head plays into this.

Hong and Kim (3) further affirm that there is no significant increase in motion sickness when using either constant amplification or dynamic amplification, but like Ragan et al., they make no mention of the participants' experience with VR headsets. Their experiment evaluated the effectiveness of both constant amplification and dynamic amplification while watching a game of ping pong. Test participants were asked to look at the left and right players, as well as the moving ball at various points during the test. The user study found that users did not notice any significant increase in the rotation for either constant amplification or dynamic amplification, but test results did show a noticeable decrease in the amount of movement from the test participants' heads, indicating that if given an appropriate amplification level, the amplified head rotation technique succeeds in creating a lower-effort interaction without sacrificing comfort or accuracy. Contrary to (1), spatial orientation and navigation were not evaluated in this study, which may explain why users did not experience the same detrimental effects.

Sellén (4) evaluated the effectiveness of amplified head rotation in regard to object selection at different distances and angles (from the "forward direction") and at different scaling factors for the head rotation amplification. The study found that the object distance was a greater influence on the object selection accuracy than the amplification factor. Different amplification factors were preferable at different distance/angle combinations, however, the one-to-one view (no amplification) came out on top in every test scenario when it comes to accuracy. Regarding ease of use, the study found that a moderate amplification factor (2x) was preferable to both no amplification (too slow) and higher amplification factor (3x), which felt uncomfortable for the object selection task, according to test participants. It should be noted that novice VR users as well as more experienced VR users were lumped together in the study, rather than separating the two groups to test for differences which may influence the effectiveness of the technique.

Kopper, Stinson and Bowman (5) ran an experiment similar to aforementioned articles, aimed at understanding the effects of amplified rotation on visual scanning tasks as well as counting tasks, but their contribution differentiates itself by the fact that they evaluated the effects of varying display characteristics; specifically, a varying FOV. Test results revealed that with a FOV of 102 degrees, no amplification factor (1x, 2x, or 3x) made a significant impact on the accuracy for the visual scanning task. With a small FOV of 30 degrees, accuracy increased with increased amplification. For a FOV of 52 degrees, moderate amplification (2x) increased accuracy, while high amplification (3x) decreased it. For the counting task, test results showed that moderate amplification performed about as well as the 1-to-1 view, whereas high amplification decreased performance significantly.

An interesting experiment related to (5) was performed by Bolte, Bruder, Steinicke, Hinrichs and Lappe (6), who conducted a psychological experiment to identify how much amplification can be applied to rotations, but in the pitch- and roll-axii rather than the usual yaw-axis, as well as the effect of different sizes for the FOV. They found that moderate augmentation (amplification factor of 1.3x for pitch and 1.44x for roll) will go unnoticed by users. This is consistent with the findings of the aforementioned papers for yaw-axis rotation amplification, where users reported high amplification as feeling uncomfortable and moderate amplification as natural. On the other hand, users estimated slightly amplified rotations in the pitch- and roll-axii as matching the real-world rotations, whereas for yaw-axis rotations, users estimate slightly condensed rotations as matching real-world rotations.

Amplified head rotations is a topic that has been studied in different contexts and for different purposes. However, more research is needed to better understand the effects of amplified head rotation, specifically on self-orientation, task performance and comfort (e.g. nausea and potential neck strain), as the literature studied above gives inconclusive evidence on some aspects and conflicting evidence on others (motion sickness). Care should be taken to ensure that user experience with VR headsets is accounted for during user tests.

2.2 Summarizing 360-degree videos

With the increasing popularity of VR, more and more 360-degree videos have been released to the public, whether it be for educational purposes or for entertainment. While they allow users to feel much more "present" in the environment of the content, they also bring new issues to light that have yet to be effectively addressed. One such issue: allowing users to fast-forward through a 360-degree video, either to skip unwanted scenes or to get a quick glimpse of the overall content. In the case of regular videos, short trailers are effective at summarizing the content of the full-length picture. But the added dimension of movement in 360-degree videos make showing the interesting snapshots a non-trivial issue. Should users be allowed to look around freely during video trailers? Or should their gaze be automatically fixed on the areas that are deemed most interesting by the content creators? There is currently very little literature focused specifically on summarizing 360-degree videos, or VR content in general. Therefore, to address the problem of summarizing

360-degree videos, we delve into studies that are concerned with cinematic content for VR, and more specifically, how to edit this type of content.

Kjaer et al. (7) published research on the effects of the frequency of cinematic cuts on the user's self-orientation and ability to follow the story in cinematic VR. They conducted two separate experiments comparing the same videos but with a different number of cuts, and using a within-subjects design for the first experiment, and a between-subjects design for the second one, to compensate for the limitations of the first study. Test results reveal that an increased number of cuts did not have a significant effect on the subject's ability to self-orientate in the virtual environment. This indicates that cutting to different parts of a video is very feasible, which is a prevalent aspect in video trailers.

Gödde, Gabler, Sigmund and Braun (8), however, assert that cuts have a negative affect on the user's sense of presence in the virtual environment, and instead suggest the use of crossfades and fades to black when changing scenes. Furthermore, for effective gaze guidance during scene transitions, they suggest overlapping the POI in the initial scene with the POI in the transitioning scene to make sure the player is looking at it after the transition has happened. This is important because users generally need a moment to re-orientate themselves in the new environment after a transition. After orientating themselves, their attention will return to the first POI that their eyes fell-on after the transition. When cutting a trailer into several scenes, these are important pointers to keep in mind.

Elmezeny, Edenhofer and Wimmer (9) make the distinction between technical immersion, which is akin to the feeling of being present in the virutal environment, and narrative immersion, which is the influence of the story in the virtual environment on the user. Even though they are separate aspects, they help to strengthen each other and create a stronger sense of belonging to the virtual environment. They conducted a user study and verified that the interplay between technical immersion and narrative aspects such as spatio-temporal immersion and emotional immersion can trigger a greater response from users than traditional video. Therefore, when creating a trailer for 360-degree video content, it is important to properly set the story, as well as ensure that POIs and characters that illicit emotions are brought to the limelight (i.e. through gaze guidance).

Literature has been published on the subject of automatic trailer generation for traditional video through machine-learning ((10), (11) and (12)), however, this is out of the scope of this study as we are not looking to learn how to teach computers to recognize the most interesting scenes in traditional video, but we want to learn the most effective way to summarize 360-degree videos while maintaining the user's sense of presence in the virtual environment as well as their ability to self-orientate and their level of comfort. As mentioned before, literature in this area is very limited and to compensate this, literature related to editing for cinematic VR experiences was studied instead to find guidelines that will prove useful in developing a method to accomplish this. It's safe to say that this study will be one of the first to study this specific interaction aspect of 360-degree video content.

2.3 Directing attention in 360-degree videos

While 360-degree videos are a great medium that offer a level of immersion and presence that is not found in regular motion pictures, they introduce a host of new problems as well. One such difficulty is the problem of drawing the user's attention to the right places at the right times during playback. With the user's ability to look at any point in the video, how do video directors attract the user's attention to points in the video that they deem important and/or interesting? While multiple senses can be stimulated to capture the user's attention (e.g. auditory or even haptic), for this study, we will focus on visual cues.

Nielsen et al. (13) performed a user study that evaluates the effects of implicit diegetic (part of the virtual environment) cues, as well as explicit non-diegetic cues (not part of the virtual environment) on guiding the user's attention, as well as the effect on the sense of presence in the VE. For the diegetic cue, a firefly was used that is meant to indicate to the user where the interesting points in the virtual environment are. For the non-diegetic cues, the user's forward direction was forced into the direction of the interesting point, but the user was still allowed to look around freely. Participants' sense of presence and the influence felt from the (non-)diegetic cues were measured afterwards through a questionnaire survey. While test results showed that the forced direction cue did result in a lower sense of presence, the difference was not statistically significant compared to the firefly cue and no cue at all. The firefly cue was perceived as noticeably more helpful when it comes to guiding attention. These findings suggest that subtle diegetic cues may be more effective when it comes to guiding the user's attention while maintaining a sense of presence in the VE than explicit (non-)diegetic cues.

Grogorick, Stengel and Magnor (14) further investigate the effectiveness of subtle cues on attention direction in their study. They made the distinction between static virtual environments that do not require many head movements from users, and dynamic environments that do require a significant number of head movements. Users were asked to complete a hidden object search task, namely finding a sphere hidden behind an object in a virtual environments filled with polyhedrons of various shapes and colors, with and without gaze guidance. In this case, gaze guidance consisted of a subtle circular discoloration on the polyhedron covering the sphere. Test results showed that the visual search task was completed significantly quicker with gaze guidance (more than two standard deviations from the mean) than without. Performance on a general search task that did not include hidden objects was also evaluated, but the results did not show significantly improved performance. This indicates that gaze guidance in the form of a subtle visual cue is especially effective in scenarios where the object to find is not immediately distinguishable from its direct surroundings. This would translate well to virtual tour videos for example, where a subtle highlight or partial discoloration in the vicinity of the object of interest (say, a building) may help distinguish it from its surroundings.

Rothe, Hußmann and Allary (15) conducted research on the effectiveness of diegetic cues for gaze guidance. These cues include stationary lights, moving lights, and sounds, as well as objects appearing out of nowhere. A user study was conducted that was aimed at finding the most effective way to capture a user's attention using these cues. Results show that moving lights and objects were effective at garnering the user's attention, whereas non-moving lights/objects had no effect at all. Furthermore, the experiment showed that guiding attention at the start of a new scene was very difficult. A likely explanation for this is that the user wants to take in the new environment before they focus their attention on any moving objects. While these three studies show that diegetic cues are good tools for gaze guidance, not all VR content allows for the creation of diegetic cues (for example, 360-degree videos where the imagery is prerecorded). It is therefore important to investigate different kinds of cues that can be incorporated into 360-degree videos easily.

So far, we have studied papers that evaluated different ways to guide the user's gaze towards an on-screen object. However, there will certainly be cases where the object or area of interest is off-screen (outside the FOV of the HMD). Lin et al. (16) propose a visualization technique designed the tackle this particular problem. The basic concept is introducing off-screen points of interest by generating a small picture preview in the peripheral area of the screen in the direction of the POI. An invisible line runs through the center-of-view and the picture preview tilts the larger the distance between it and the viewport. The tilting also creates a greater degree of immersion, as the tilting effectively gives the preview diegetic properties and allows it to blend into the environment more than a flat picture preview would. A user study was conducted comparing the Outside-In technique with a standard pointing arrow based interface for gaze guidance. Users filled in a post-study questionnaire to rate the amount of interference experienced with each technique and the level of spatial understanding of the POI in the virtual environment, among other aspects. Ratings show that the Outside-In technique outperforms the pointing arrow technique when it comes to giving the user spatial awareness of the POI's location in the virtual environment, and there is no noticeable difference in the amount of interference experienced between the two techniques. The logical explanation is that the tilting of the picture previews has effectively made them diegetic visual cues that respond to the changing of the virtual environment, whereas the arrows, while smaller, remain non-diegetic and therefore create a lesser sense of presence.

Sassatelli et al. (17) take gaze guidance research in a different direction by proposing a technique called Snap-Changes that moves the user's orientation directly onto a POI if the angle between the POI and user gaze is large enough, effectively "snapping" it into position. While this sounds like it would be very intrusive and by extension would interfere with the user's sense of presence, the authors delve into the concept of video-editing to explain that fast-cuts (or snap-changes) do not interfere with the user's vestibular system and thus cause no motion sickness and are therefore a viable option for gaze guidance. Unfortunately, effects on the sense of presence were not tested. A user study was conducted to evaluate the technique's effectiveness at reducing head motion, as well as allowing users to more easily notice the predetermined POIs in the virtual environment. Test results show that users spent less time moving their heads with the Snap-Changes technique than without. Re-positioning the user in front of the POI seemed to incite them to keep their focus on it, which leads to a significantly higher amount of time spent looking at these POIs, confirming that Snap-Changes is successful in gaze guidance.

While not specifically aimed at 360-degree videos, Waldin, Waldner and Viola (18) propose a method for gaze guidance in images using high-frequency flicker. The idea behind it is that the flicker is not noticeable when looking directly at it if operating at a frequency between 60 and 72 Hz, but it will draw the user's attention if located in the peripheral part of the FOV. The results of a user experiment showed that users were clearly aware of the flicker in the peripheral vision, but rarely noticed it in the foveal vision. The results further showed that users could be effectively guided towards POIs in the virtual environment. Users reported negligible discomfort when looking directly at the POIs highlighted by the flicker. It would be interesting to evaluate the effectiveness of this technique in a VR setting with the user wearing an HMD, though this technique may be limited by the HMD's FOV. High-end VR headsets such as the HTC Vive and the Oculus Rift may benefit more from the technique due to their high FOV compared to some the cheaper HMDs. It is unclear how the flicker's location within the peripheral vision affects its perceivability.

In this section, we've seen various methods to guide the user's attention towards POIs in the virtual environment, from explicit non-diegetic cues to implicit diegetic cues, each effective in their own right. For this study, we are especially interested in subtle cues that offer the user as much freedom of movement in the virtual environment and that do not intrude (too much) on the user's sense of presence. Initially, we assumed that non-diegetic cues would be largely ineffective because they are not experienced as "part of the environment". However, as seen in (16), non-diegetic cues can be given diegetic properties such as tilting in relation to the angle between the user's forward gaze and the POI. Further research in this particular area could deliver very interesting results.

3 Summary and conclusions

Virtual reality is an upcoming medium that offers high levels of immersion and presence in virtual environments. With the introduction of high-end, but affordable HMDs such as the Oculus Rift, the HTC Vive and the Google Cardboard, VR has come into the limelight in recent years and many interested parties are releasing more and more VR content, and a large number of institutes and universities are already conducting research related to VR.

VR and especially 360-degree videos are a relatively new medium, however, and therefore there are still issues that have not yet been well researched. One such topic is user interaction. User interaction can be decomposed into many separate interaction aspects, such as interactive FOV manipulation, but also guiding the user's attention in the virtual environment towards points of interest and even summarizing a 360-degree video to quickly glean what kind of content it has to offer. This literature study has focused on these three aspects specifically.

We discussed several studies that focus on interactively manipulating the user's FOV by amplifying rotations in the virtual environment. All of the papers that we reviewed implement similar amplified rotations in the yaw-axis. While the direction of research for this type of interaction seems clear, there are still many ways to implement the amplified rotations, by using a static multiplication factor based on the head movement, or a dynamic one, for example and so far there hasn't been a method that is clearly superior to others. Furthermore, many of these papers report conflicting test results, especially when it comes to motion sickness experienced by the user.

When it comes to summarizing 360-degree videos, very little literature has been released on the subject. So little, in fact, that it was necessary to go into the direction of cinematic VR and editing to find literature that could serve as a basis for a study.

Guiding the user's attention in VR is a subject that has been much more thoroughly studied. Many different methods have been proposed, be it either explicit non-diegetic visual cues, or implicit diegetic cues, or even less obvious approaches such as audio cues and high-frequency flickers that are designed to be nearly invisible when looked at directly, but quite intrusive when appearing in the peripheral vision. We are predominantly interested in gaze guidance for 360-degree video and the solutions for this specific medium are less abundant. This is because it is hardly possible to create diegetic cues for pre-recorded footage, therefore, we are limited mostly to non-diegetic cues for gaze guidance. There may be effective ways to give non-diegetic cues a more diegetic nature by transforming them according to the user's position and rotation relative to the virtual environment and thus making them less intrusive on the user's sense of presence. This is an interesting direction to take for further research.

References

[1] S. P. Sargunam, K. R. Moghadam, M. Suhail, and E. D. Ragan, "Guided head rotation and amplified head rotation: Evaluating semi-natural travel and viewing techniques in virtual reality," *IEEE Virtual Reality (VR)*, 2017.

[2] E. D. Ragan, S. Scerbo, F. Bacim, and D. A. Bowman, "Amplified head rotation in virtual reality and the effects on 3d search, training transfer, and spatial orientation," *IEEE Transactions on Visualization and Computer Graphics*, 2017.

[3] S. Hong and G. J. Kim, "Accelerated viewpoint panning with rotational gain in 360 degree videos," *VRST '16*, 2016.

[4] J. Karlsson Sellén, "Performance of amplified head rotation in virtual reality under different object distances," 2017.

[5] R. Kopper, C. Stinson, and D. A. Bowman, "Towards an understanding of the effects of amplified head rotations," 2012.

[6] B. Bolte, G. Bruder, F. Steinicke, K. Hinrichs, and M. Lappe, "Augmentation techniques for efficient exploration in head-mounted display environments," 2010.

[7] T. Kjær, C. B. Lillelund, M. Moth-Poulsen, N. C. Nilsson, R. Nordahl, and S. Serafin, "Can you cut it?: An exploration of the effects of editing in cinematic virtual reality," in *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, ser. VRST '17. New York, NY, USA: ACM, 2017, pp. 4:1–4:4.

[8] M. Gödde, F. Gabler, D. Siegmund, and A. Braun, "Cinematic narration in vr – rethinking film conventions for 360 degrees," 2018.

[9] E. Ahmed, N. Edenhofer, and J. Wimmer, "Immersive storytelling in 360-degree videos: An analysis of interplay between narrative and technical immersion," *Journal For Virtual Worlds Research*, vol. 11, no. 1, 2018.

[10] J. R. Smith, D. Joshi, B. Huet, W. Hsu, and J. Cota, "Harnessing a.i. for augmenting creativity: Application to movie trailer creation," in *Proceedings of the 25th ACM International Conference on Multimedia*, ser. MM '17. New York, NY, USA: ACM, 2017, pp. 1799–1808.

[11] G. Irie, T. Satou, A. Kojima, T. Yamasaki, and K. Aizawa, "Automatic trailer generation," in *Proceedings of the 18th ACM International Conference on Multimedia*, ser. MM '10. New York, NY, USA: ACM, 2010, pp. 839–842.

[12] E. Amer and A. Nabil, "A framework to automate the generation of movies' trailers using only subtitles," in *Proceedings of the 7th International Conference on Software and Information Engineering*, ser. ICSIE '18. New York, NY, USA: ACM, 2018, pp. 126–130.

[13] L. T. Nielsen, M. B. Møller, S. D. Hartmeyer, T. C. M. Ljung, N. C. Nilsson, R. Nordahl, and S. Serafin, "Missing the point: An exploration of how to guide users' attention during cinematic virtual reality," *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, 2016.

[14] S. Grogorick, M. Stengel, and M. Magnor, "Subtle gaze guidance for immersive environments," *Proceedings of SAP*, 2017.

[15] S. Rothe, H. Hußmann, and A. Mathias, "Diegetic cues for guiding the viewer in cinematic virtual reality," *Proceedings of the Second International Symposium of Chinese CHI*, 2014.

[16] Y.-T. Lin, Y.-C. Liao, S.-Y. Teng, Y.-J. Chung, L. Chan, and B.-Y. Chen, "Outside-in: Visualizing out-of-sight regions-of-interest in a 360 video using spatial picture-in-picture previews," *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, 2017.

[17] L. Sassatelli, A.-M. Pinna-Déry, M. Winckler, S. Dambra, G. Samela, R. Pighetti, and R. Aparicio-Pardo, "Snap-changes: a dynamic editing strategy for directing viewer's attention in streaming virtual reality videos," *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*, 2018.

[18] N. Waldin, M. Waldner, and I. Viola, "Flicker observer effect: Guiding attention through high frequency flicker in images," *Computer Graphics Forum*, 2017.

5 Appendix

5.1 Instrumentation

We developed a VR app for Android phones using Unity3D Game Engine and the Google Cardboard SDK. A different version of the app (with a unique combination of video order + technique order) was developed for each user test (32 apps in total). This study was originally a continuation of Anvar Arashov's master thesis project [1] and therefore his Unity3D project files were used as a starting point for the development of the apps.

We used a VR headset for mobile phones of unknown brand (Figure 1).

Figure 1: The HMD used in the user tests

We used a Samsung Galaxy S9 with 4GB of RAM and a 10nm 64-bit Octa-Core Processor 2.8GHz + 1.7GHz (Maximum Clock Speed, Performance Core + Efficiency Core) to run the VR app on.

Two of the four techniques required the usage of a wireless controller. We used the 8bitdo SN30PRO wireless gamepad for these techniques.

Figure 2: The 8bitdo SN30PRO wireless gamepad

To monitor the participants' performance, we connected the Galaxy S9 to a laptop with a USB cable and used the Vysor app for real-time screen mirroring.

5.2 Questionnaire

This section contains the questionnaire used in the user test. Note that the first three post-task survey questions were asked after each overview task and the remaining three questions were asked after each scene-finding task. ”Relevant content” in the question ”How easy did you find it to find relevant content using this technique” refers to the scene we asked the participant to find in the scene-finding task.

360-degree video browsing experiment

In this study, we evaluate different ways to browse 360-degree video using a virtual reality headset. We want to test which technique is best to get a quick overview of a video's content, and which is best to find a particular scene in the video.

You will be wearing a VR headset and will be instructed by me. We will be testing 4 different techniques. There will be two tasks per technique (one to get an overview of the video's content, one to find a particular scene in it).

There will be no monetary compensation for the experiment. It is completely voluntary and you are under no obligation to complete it.

Using a VR headset may cause motion sickness/nausea. If this happens, you can request a break at any point during the experiment. You may also stop the experiment at any point, for any reason. Should you choose to do so, any data you may have provided will be deleted and not used for the research.

Before we start the experiment, I will ask you to fill in a pre-experiment survey where you answer some questions regarding your background and experience with VR. During the experiment, I will ask you for feedback on the techniques. After the experiment, there will be an informal interview. Any information you provide is completely anonymous and cannot be traced back to your name.

If you have any questions beforehand or during the experiment, please feel free to ask them.

By clicking on the button below, you confirm that you have read and understand the information above and that you are willing to participate in this study.

* Required

Pre-experiment survey

Please answer some questions regarding your background and experience with VR and 360-degree videos.

1. **Age:** *

   ____

2. **Sex:** *

   Mark only one oval.

   - ( ) Male
   - ( ) Female
   - ( ) Other

3. **Experience with mounted headset VR?** *

   Mark only one oval.

   - ( ) None
   - ( ) Used once or twice
   - ( ) Regular user

4. **Please only answer this question if you selected "regular user" in the question above**
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
I rarely use VR	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	I use VR very frequently

5. **Experience with 360-degree video using head-mounted display** *
Mark only one oval.

- ⬭ None
- ⬭ Used once or twice
- ⬭ Regular user

6. **Please only answer this question if you selected "regular user" in the question above**
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
I rarely watch 360-degree video using HMD	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	I watch 360-degree video using HMD very frequently

Experiment 1

Experiment 1 will now begin. Please continue to the next section after completing the experiment.

Post-task survey 1

7. **How confident are you that you got a good overview of the video content?** *
Mark only one oval.

	1	2	3	4	5	
Not at all	⬭	⬭	⬭	⬭	⬭	Very confident

8. **How easy did you find it to get an overview using this technique?** *
Mark only one oval.

	1	2	3	4	5	
Very hard	⬭	⬭	⬭	⬭	⬭	Very easy

9. **Do you remember this specific object?** *
Mark only one oval.

- ⬭ Yes
- ⬭ No

10. **How easy did you find it to find relevant content using this technique?** *
Mark only one oval.

	1	2	3	4	5	
Very difficult	⬭	⬭	⬭	⬭	⬭	Very easy

11. **Did you experience any discomfort (nausea, dizziness)?** *
Mark only one oval.

	1	2	3	4	5	
Not at all	⬭	⬭	⬭	⬭	⬭	Excessively

12. **Do you have any feedback for this technique?**

Experiment 2

Experiment 2 will now begin. Please continue to the next section after completing the experiment.

Post-task survey 2

13. **How confident are you that you got a good overview of the video content?** *
Mark only one oval.

	1	2	3	4	5	
Not at all	⬭	⬭	⬭	⬭	⬭	Very confident

14. **How easy did you find it to get an overview using this technique?** *
Mark only one oval.

	1	2	3	4	5	
Very hard	⬭	⬭	⬭	⬭	⬭	Very easy

15. **Do you remember this specific object?** *
Mark only one oval.

- ⬭ Yes
- ⬭ No

16. **How easy did you find it to find relevant content using this technique?** *
Mark only one oval.

	1	2	3	4	5	
Very difficult	⬭	⬭	⬭	⬭	⬭	Very easy

17. **Did you experience any discomfort (nausea, dizziness)?** *
Mark only one oval.

	1	2	3	4	5	
Not at all	⬭	⬭	⬭	⬭	⬭	Excessively

18. **Do you have any feedback for this technique?**

Experiment 3

Experiment 3 will now begin. Please continue to the next section after completing the experiment.

Post-task survey 3

19. **How confident are you that you got a good overview of the video content?** *
Mark only one oval.

	1	2	3	4	5	
Not at all	⬭	⬭	⬭	⬭	⬭	Very confident

20. **How easy did you find it to get an overview using this technique?** *
Mark only one oval.

	1	2	3	4	5	
Very hard	⬭	⬭	⬭	⬭	⬭	Very easy

21. **Do you remember this specific object?** *
Mark only one oval.

- ⬭ Yes
- ⬭ No

22. **How easy did you find it to find relevant content using this technique?** *
Mark only one oval.

	1	2	3	4	5	
Very difficult	◯	◯	◯	◯	◯	Very easy

23. **Did you experience any discomfort (nausea, dizziness)?** *
Mark only one oval.

	1	2	3	4	5	
Not at all	◯	◯	◯	◯	◯	Excessively

24. **Do you have any feedback for this technique?**

Experiment 4

Experiment 4 will now begin. Please continue to the next section after completing the experiment.

Post-task survey 4

25. **How confident are you that you got a good overview of the video content?** *
Mark only one oval.

	1	2	3	4	5	
Not at all	◯	◯	◯	◯	◯	Very confident

26. **How easy did you find it to get an overview using this technique?** *
Mark only one oval.

	1	2	3	4	5	
Very hard	◯	◯	◯	◯	◯	Very easy

27. **Do you remember this specific object?** *
Mark only one oval.

◯ Yes
◯ No

28. **How easy did you find it to find relevant content using this technique?** *
Mark only one oval.

	1	2	3	4	5	
Very difficult	◯	◯	◯	◯	◯	Very easy

29. **Did you experience any discomfort (nausea, dizziness)?** *
Mark only one oval.

	1	2	3	4	5	
Not at all	◯	◯	◯	◯	◯	Excessively

30. **Do you have any feedback for this technique?**

____
____
____
____
____

Post-experiment interview

31. **Which technique did you prefer over all others?** *
Mark only one oval.

- ◯ 1. Fast-forwarding/rewinding with game controller
- ◯ 2. Jumping on timeline with game controller
- ◯ 3. Fast-forwarding/rewinding with gaze interaction
- ◯ 4. Jumping on timeline with gaze interaction

32. **Why did you prefer this technique over the others?** *

____
____
____
____
____

33. **Please rank the techniques above from best to worst (i.e. 3, 2, 4, 1)** *

____

5.3 Data

This section contains all data collected for this study. First we present the experiment testing order, followed by the objective data collected from the experiments and then finally the subjective data collected from the questionnaire.

5.3.1 Participant Testing Order

	TIMELINE TECHNIQUES FIRST			
	TC	**TG**	**FC**	**FG**
Participant 1	Video B	Video C	Video D	Video A
Participant 12	Video D	Video A	Video C	Video B
Participant 28	Video C	Video B	Video A	Video D
Participant 24	Video A	Video D	Video B	Video C
	TG	**TC**	**FG**	**FC**
Participant 19	Video D	Video A	Video B	Video C
Participant 30	Video B	Video C	Video D	Video A
Participant 27	Video C	Video D	Video A	Video B
Participant 22	Video A	Video B	Video C	Video D
	FC	**FG**	**TC**	**TG**
Participant 16	Video D	Video C	Video B	Video A
Participant 25	Video C	Video D	Video A	Video B
Participant 8	Video B	Video A	Video D	Video C
Participant 32	Video A	Video B	Video C	Video D
	FG	**FC**	**TG**	**TC**
Participant 21	Video B	Video C	Video A	Video D
Participant 10	Video A	Video D	Video B	Video C
Participant 14	Video C	Video B	Video D	Video A
Participant 26	Video D	Video A	Video C	Video B

	FAST-FORWARD TECHNIQUES FIRST			
	TG	**TC**	**FC**	**FG**
Participant 29	Video D	Video C	Video A	Video B
Participant 20	Video C	Video B	Video D	Video A
Participant 4	Video A	Video D	Video B	Video C
Participant 7	Video B	Video A	Video C	Video D
	TC	**TG**	**FG**	**FC**
Participant 2	Video A	Video C	Video B	Video D
Participant 9	Video D	Video A	Video C	Video B
Participant 15	Video C	Video B	Video D	Video A
Participant 18	Video B	Video D	Video A	Video C
	FC	**FG**	**TG**	**TC**
Participant 23	Video A	Video B	Video C	Video D
Participant 13	Video D	Video A	Video B	Video C
Participant 3	Video C	Video D	Video A	Video B
Participant 11	Video B	Video C	Video D	Video A
	FG	**FC**	**TC**	**TG**
Participant 31	Video C	Video A	Video D	Video B
Participant 17	Video D	Video B	Video A	Video C
Participant 5	Video A	Video C	Video B	Video D
Participant 6	Video B	Video D	Video C	Video A

TC = Timeline Controller
TG = Timeline Gaze
FC = Fast-Forward Controller
FG = Fast-Forward Gaze

Video A = Rome
Video B = FreeSolo
Video C = London
Video D = Underwater

5.3.2 Scene-Finding Task Completion Times

This section contains the completion time in seconds for the scene-finding task, per technique. The average completion time per technique is displayed at the bottom.

	Timeline Controller	Fast-Forward Controller	Timeline Gaze	Fast-Forward Gaze
Participant 1		13,58	37,56	16,23
Participant 2	7,24	13,12	4,16	53,3
Participant 3	59,76	18,61	23,28	24,57
Participant 4	11,92	19,54	8,07	13,13
Participant 5	29,51	27,13	13,54	23,74
Participant 6	12,81	16,4		
Participant 7	13,12		6,79	25,48
Participant 8	26,34	17,32		57,98
Participant 9	12,04	19,86	47,81	14,91
Participant 10	16,32		8,17	
Participant 11	21,1	22,66	36,54	14,62
Participant 12	16,28	22,12	8,41	34,81
Participant 13	20,16	31,26		23,23
Participant 14	32,24	50,21	30,42	27,84
Participant 15				
Participant 16	52,54	59,47		17,16
Participant 17	34,28	45,85	10,37	
Participant 18			8,47	48,7
Participant 19			16,44	35,53
Participant 20	47,53	14,15	9,52	17,41
Participant 21	16,4	12,81		
Participant 22	35,2	17,37	12,93	13,68
Participant 23	15,93		22,84	21,16
Participant 24			26,64	14,43
Participant 25	41,87		10,46	31,33
Participant 26	15,45	18,87	14,48	37,64
Participant 27	24,13	50	32,63	17,27
Participant 28		30,43	10,13	20,03
Participant 29	18,75		11,95	43,21
Participant 30	7,47		10,85	27,22
Participant 31	18,38	31,87	18,41	21,39
Participant 32	14,35		15,36	
AVERAGE	23,89	26,32	17,55	26,77

5.3.3 Remembered Scenes

In the overview task we asked the participant if they remember seeing a specific scene in the video. The green fields indicate that they remembered. The red fields indicates that they didn't. The white fields indicate that the participant was unable to complete the task within the time limit of 1 minute. We did not consider these timeouts in the statistical analysis of this data (which is why the blank fields contain no data).

	Timeline Controller	Fast-Forward Controller	Timeline Gaze	Fast-Forward Gaze
Participant 1		Yes	No	Yes
Participant 2	Yes	Yes	Yes	Yes
Participant 3	No	Yes	Yes	Yes
Participant 4	No	Yes	Yes	Yes
Participant 5	No	Yes	No	Yes
Participant 6	Yes	Yes		
Participant 7	No		Yes	No
Participant 8	No	No		No
Participant 9	No	No	No	No
Participant 10	Yes		Yes	
Participant 11	Yes	Yes	Yes	No
Participant 12	Yes	Yes	Yes	No
Participant 13	No	No		Yes
Participant 14	Yes	No	No	Yes
Participant 15				
Participant 16	No	No		No
Participant 17	Yes	No	Yes	
Participant 18			Yes	Yes
Participant 19			Yes	No
Participant 20	No	No	No	Yes
Participant 21	No	Yes		
Participant 22	No	Yes	No	Yes
Participant 23	No		Yes	No
Participant 24			No	Yes
Participant 25	Yes		Yes	No
Participant 26	No	No	Yes	Yes
Participant 27	Yes	No	No	Yes
Participant 28		Yes	Yes	Yes
Participant 29	No		Yes	Yes
Participant 30	Yes		Yes	Yes
Participant 31	No	Yes	Yes	Yes
Participant 32	No		No	
REMEMBERED	38,46%	57,14%	65,38%	65,38%

5.3.4 Subjective Data - Average Scores

These values are the averages of the scores assigned by our test participants to the questions under each post-task survey in the questionnaire (the numbers for REMEMBERED SCENE represent the percentage of remembered scenes vs total scenes).

		TC	FC	TG	FG
AVERAGE	OVERVIEW SCORE	3,59	3,97	3,47	3,97
	OVERVIEW EASE	4,16	4,50	3,56	4,16
	REMEMBERED SCENE	38,46%	57,14%	65,38%	65,38%
	SCENE-FINDING EASE	4,28	4,28	4,31	3,88
	DISCOMFORT SCORE	1,03	1,09	1,09	1,06

The next few sections contain the individual scores for each question.

5.3.5 Overview Score

OVERVIEW SCORE	GAMEPAD TIMELINE	GAMEPAD FAST-FORWARD	GAZE TIMELINE	GAZE FAST-FORWARD
Participant 1	4	5	2	2
Participant 2	5	4	5	4
Participant 3	3	5	4	5
Participant 4	3	4	5	5
Participant 5	4	4	3	4
Participant 6	4	4	4	4
Participant 7	4	4	3	5
Participant 8	4	3	4	3
Participant 9	2	3	3	3
Participant 10	4	5	4	2
Participant 11	4	4	4	5
Participant 12	4	3	3	4
Participant 13	5	5	3	5
Participant 14	4	4	2	3
Participant 15	2	3	3	5
Participant 16	4	3	3	3
Participant 17	3	4	5	3
Participant 18	2	3	3	4
Participant 19	3	3	3	5
Participant 20	4	4	4	4
Participant 21	4	5	5	5
Participant 22	3	3	2	4
Participant 23	5	4	4	5
Participant 24	4	5	3	4
Participant 25	5	5	4	5
Participant 26	4	4	5	1
Participant 27	5	3	3	4
Participant 28	1	5	4	5
Participant 29	2	3	2	4
Participant 30	2	3	3	4
Participant 31	4	5	4	4
Participant 32	4	5	2	4

5.3.6 Overview Ease

OVERVIEW EASE	GAMEPAD TIMELINE	GAMEPAD FAST-FORWARD	GAZE TIMELINE	GAZE FAST-FORWARD
Participant 1	3	4	1	3
Participant 2	5	5	4	4
Participant 3	3	5	5	5
Participant 4	4	5	4	5
Participant 5	3	4	2	2
Participant 6	5	5	5	5
Participant 7	5	4	3	4
Participant 8	4	3	3	3
Participant 9	5	4	3	4
Participant 10	5	5	3	4
Participant 11	5	4	5	5
Participant 12	5	4	2	4
Participant 13	5	5	3	5
Participant 14	4	5	2	3
Participant 15	2	3	4	5
Participant 16	3	5	4	2
Participant 17	3	5	5	4
Participant 18	4	4	4	5
Participant 19	4	4	2	5
Participant 20	3	5	4	5
Participant 21	4	5	5	4
Participant 22	3	4	3	3
Participant 23	4	4	5	5
Participant 24	5	5	3	5
Participant 25	5	5	4	4
Participant 26	5	5	5	4
Participant 27	5	5	3	4
Participant 28	5	5	5	5
Participant 29	4	4	2	4
Participant 30	3	4	2	4
Participant 31	5	5	5	5
Participant 32	5	5	4	4

5.3.7 Scene-Finding Ease

SCENE-FINDING EASE	GAMEPAD TIMELINE	GAMEPAD FAST-FORWARD	GAZE TIMELINE	GAZE FAST-FORWARD
Participant 1	4	5	5	4
Participant 2	5	4	5	3
Participant 3	3	4	5	4
Participant 4	4	5	5	4
Participant 5	4	2	4	1
Participant 6	5	5	5	4
Participant 7	4	2	4	4
Participant 8	5	4	4	3
Participant 9	5	4	3	4
Participant 10	5	5	5	4
Participant 11	5	4	3	5
Participant 12	5	5	5	4
Participant 13	5	5	4	5
Participant 14	4	5	2	5
Participant 15	1	4	4	5
Participant 16	3	4	5	5
Participant 17	4	4	4	2
Participant 18	3	5	5	5
Participant 19	4	5	4	4
Participant 20	4	5	5	5
Participant 21	4	4	5	2
Participant 22	4	5	4	4
Participant 23	4	4	5	5
Participant 24	4	4	4	5
Participant 25	5	3	5	3
Participant 26	5	5	5	3
Participant 27	5	5	2	5
Participant 28	5	4	5	5
Participant 29	5	5	4	2
Participant 30	4	5	3	4
Participant 31	5	5	5	5
Participant 32	5	2	5	1

5.3.8 Discomfort Score

DISCOMFORT	GAMEPAD TIMELINE	GAMEPAD FAST-FORWARD	GAZE TIMELINE	GAZE FAST-FORWARD
Participant 1	1	1	1	1
Participant 2	1	1	1	1
Participant 3	1	1	1	1
Participant 4	1	1	1	1
Participant 5	1	1	1	1
Participant 6	1	1	1	1
Participant 7	1	1	1	1
Participant 8	1	1	1	1
Participant 9	1	1	1	1
Participant 10	1	1	1	1
Participant 11	1	1	1	1
Participant 12	1	1	1	1
Participant 13	1	1	1	1
Participant 14	1	1	2	2
Participant 15	1	1	1	1
Participant 16	1	1	1	1
Participant 17	1	1	1	1
Participant 18	1	2	1	1
Participant 19	1	1	1	1
Participant 20	1	1	1	1
Participant 21	1	1	1	1
Participant 22	1	1	1	1
Participant 23	2	2	1	1
Participant 24	1	1	2	2
Participant 25	1	2	2	1
Participant 26	1	1	1	1
Participant 27	1	1	1	1
Participant 28	1	1	1	1
Participant 29	1	1	1	1
Participant 30	1	1	1	1
Participant 31	1	1	1	1
Participant 32	1	1	1	1

6 Bibliography

References

[1] A. Arashov, "Drawing attention problem in 360-degree video," Master's thesis, 2018.

[2] S. P. Sargunam, K. R. Moghadam, M. Suhail, and E. D. Ragan, "Guided head rotation and amplified head rotation: Evaluating semi-natural travel and viewing techniques in virtual reality," *IEEE Virtual Reality (VR)*, 2017.

[3] E. D. Ragan, S. Scerbo, F. Bacim, and D. A. Bowman, "Amplified head rotation in virtual reality and the effects on 3d search, training transfer, and spatial orientation," *IEEE Transactions on Visualization and Computer Graphics*, 2017.

[4] J. Karlsson Sellén, "Performance of amplified head rotation in virtual reality under different object distances," 2017.

[5] S. Hong and G. J. Kim, "Accelerated viewpoint panning with rotational gain in 360 degree videos," *VRST '16*, 2016.

[6] R. Kopper, C. Stinson, and D. A. Bowman, "Towards an understanding of the effects of amplified head rotations," 2012.

[7] B. Bolte, G. Bruder, F. Steinicke, K. Hinrichs, and M. Lappe, "Augmentation techniques for efficient exploration in head-mounted display environments," 2010.

[8] L. T. Nielsen, M. B. Møller, S. D. Hartmeyer, T. C. M. Ljung, N. C. Nilsson, R. Nordahl, and S. Serafin, "Missing the point: An exploration of how to guide users' attention during cinematic virtual reality," *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, 2016.

[9] S. Grogorick, M. Stengel, and M. Magnor, "Subtle gaze guidance for immersive environments," *Proceedings of SAP*, 2017.

[10] Y.-T. Lin, Y.-C. Liao, S.-Y. Teng, Y.-J. Chung, L. Chan, and B.-Y. Chen, "Outside-in: Visualizing out-of-sight regions-of-interest in a 360 video using spatial picture-in-picture previews," *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, 2017.

[11] L. Sassatelli, A.-M. Pinna-Déry, M. Winckler, S. Dambra, G. Samela, R. Pighetti, and R. Aparicio-Pardo, "Snap-changes: a dynamic editing strategy for directing viewer's attention in streaming virtual reality videos," *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*, 2018.

[12] N. Waldin, M. Waldner, and I. Viola, "Flicker observer effect: Guiding attention through high frequency flicker in images," *Computer Graphics Forum*, 2017.

[13] S. Rothe, H. Humann, and A. Mathias, "Diegetic cues for guiding the viewer in cinematic virtual reality," *Proceedings of the Second International Symposium of Chinese CHI*, 2014.

[14] M. Gödde, F. Gabler, D. Siegmund, and A. Braun, "Cinematic narration in vr – rethinking film conventions for 360 degrees," 2018.

[15] E. Ahmed, N. Edenhofer, and J. Wimmer, "Immersive storytelling in 360-degree videos: An analysis of interplay between narrative and technical immersion," *Journal For Virtual Worlds Research*, vol. 11, no. 1, 2018.

[16] J. R. Smith, D. Joshi, B. Huet, W. Hsu, and J. Cota, "Harnessing a.i. for augmenting creativity: Application to movie trailer creation," in *Proceedings of the 25th ACM International Conference on Multimedia*, ser. MM '17. New York, NY, USA: ACM, 2017, pp. 1799–1808.

[17] E. Amer and A. Nabil, "A framework to automate the generation of movies' trailers using only subtitles," in *Proceedings of the 7th International Conference on Software and Information Engineering*, ser. ICSIE '18. New York, NY, USA: ACM, 2018, pp. 126–130.

[18] G. Irie, T. Satou, A. Kojima, T. Yamasaki, and K. Aizawa, "Automatic trailer generation," in *Proceedings of the 18th ACM International Con-*

ference on Multimedia, ser. MM '10. New York, NY, USA: ACM, 2010, pp. 839–842.

[19] T. Kjær, C. B. Lillelund, M. Moth-Poulsen, N. C. Nilsson, R. Nordahl, and S. Serafin, "Can you cut it?: An exploration of the effects of editing in cinematic virtual reality," in *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, ser. VRST '17. New York, NY, USA: ACM, 2017, pp. 4:1–4:4.

[20] G. W. Corder and D. I. Foreman, "Nonparametric statistics for non-statisticians," 2011.

[21] W. J. Conover and R. L. Iman, "On multiple-comparisons procedures," *NM, USA: Los Alamos Scientific Lab*, pp. 1–17, 1979.

[22] S. V. Cobb, S. Nichols, A. Ramsey, and J. R. Wilson, "Virtual reality-induced symptoms and effects (vrise)," *Presence: Teleoperators & Virtual Environments*, vol. 8, no. 2, pp. 169–186, 1999.

[23] E. Patrick, D. Cosgrove, A. Slavkovic, J. A. Rode, T. Verratti, and G. Chiselko, "Using a large projection screen as an alternative to head-mounted displays for virtual environments," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 2000, pp. 478–485.

[24] "A london city guided tour (360 vr video)," Youtube. [Online]. Available: https://www.youtube.com/watch?v=m9EClKA1VeQ

[25] "Rome: A guided city tour (360 vr video)," Youtube. [Online]. Available: https://www.youtube.com/watch?v=1ziMH$_lAUW0$

[26] "Free solo 360 — national geographic," Youtube. [Online]. Available: https://www.youtube.com/watch?v=FRGF77fBAeM

[27] "360 underwater national park — national geographic," Youtube. [Online]. Available: https://www.youtube.com/watch?v=v64KOxKVLVg