Game and Media Technology

Master Thesis

Enhancing 360-Degree Video Collection Exploration: A Novel Interface for Immersive Browsing in Virtual Reality

---

Author:

Bas van Eck

Supervisor:

Wolfgang Hürst

Second examiner:

Remco Veltkamp

July 04, 2024

Utrecht University

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to my supervisor, Dr. Wolfgang Hürst, for his continuous support, guidance, and encouragement throughout my thesis. His insights and expertise were instrumental in shaping this thesis. I am also thankful to my second examiner, Dr. Remco Veltkamp, for his valuable time and effort in evaluating my thesis.

A special thanks to Michel Maasdijk, for providing me with the hardware required to perform my user study and to my friends and family for their moral support during this journey.

Lastly, I would like to thank all those who participated in my user study for their valuable time and insights. Your contributions have been essential to the success of this thesis. Thank you all for making this accomplishment possible.

Contents

Enhancing 360-Degree Video Collection Exploration: A Novel Interface for Immersive Browsing in Virtual Reality

Bas van Eck
Utrecht University
Netherlands
b.vaneck1@students.uu.nl

ABSTRACT

Browsing 360° video libraries in virtual reality is commonly approached with a traditional two-dimensional grid presenting equirectangular or custom thumbnails, similar to how traditional videos are presented. However, this approach fails to take advantage of unique opportunities offered by virtual reality environments, such as immersive displays capable of showing content in three-dimensional space. This study aims to improve the browsing experience by proposing two interfaces that utilize the additional screen space offered by VR to present a detailed video visualization. The filmstrip interface represents the conventional approach to visualizing videos, and the hierarchical interface proposes a new approach where videos can be hierarchically explored. A comparative study was conducted with 32 participants, aged between 18 and 30 years. Participants interacted with both the filmstrip and hierarchical interfaces, performing a timed search task to find a specific video containing a previously viewed frame. They used the interfaces in a counterbalanced order. Performance was assessed based on the time measure and the correctness of the video identification. The user experience was evaluated using the UEQ-S and qualitative feedback from a post-task questionnaire.

The results indicated that participants completed the search task statistically significantly faster using the hierarchical interface compared to the filmstrip interface. Accuracy between the two interfaces showed no significant difference, with almost all participants successfully identifying the correct video in both interfaces. The results considering user experience also indicated a preference for the hierarchical interface. While no significant difference was found in pragmatic quality between the interfaces, the hierarchical interface demonstrated significantly better hedonic and overall quality. These findings indicate that the hierarchical interface results in a better performance and user experience, suggesting that a hierarchical approach provides a more enjoyable and efficient browsing experience for 360° video libraries in VR environments.

1 INTRODUCTION

As virtual reality (VR) technology continues to advance and head-mounted displays (HMDs) become more affordable for the average consumer, an increasing number of people are gaining access to the immersive experience of 360° videos. These videos are typically accessed through applications such as YouTube VR or Oculus Video. The interfaces used in these applications present videos as thumbnails in a two-dimensional (2D) grid pattern, similar to traditional videos are presented. In VR, these interfaces are projected in three-dimensional (3D) space. However, a significant portion of screen space remains underutilized in these VR interfaces. Additionally, 2D thumbnails do not effectively convey the complete context of 360° videos without distorting the images, presenting an opportunity for enhancing the VR browsing experience of 360° videos.

A study by Knoop [21] approached this problem with a focus on the user experience by designing interfaces modeled after real-world physical locations. While this approach had promising results for immersion and user experience, it sacrifices browsing efficiency as users have to navigate the virtual space physically. In this study, we address these challenges by designing an interface that integrates the traditional thumbnail grids with enhanced video visualization, where videos can be hierarchically explored. This method aims to improve both the browsing efficiency as well as the user experience while navigating 360° video libraries in VR. This research focuses on the visualization of video previews with browsing efficiency and user experience as key metrics.

2 RELATED WORK

This review explores enhanced video browsing for traditional videos, 360° videos, and video browsing in VR. Research on 360° video collection browsing is sparse. Rarely are the three key components: 360° videos, Video browsing, and Virtual Reality (VR) combined in a single study. To begin, let us examine traditional video browsing. Numerous platforms, such as YouTube, Google Videos, and Facebook Video, employ comparable methods for video browsing, typically displaying thumbnails in a grid or list format to represent videos. Additionally, most applications offer a preview that starts playing when users hover over the video thumbnails with their cursor. While these applications use literal previews where the video starts playing, the literature has explored other methods to visualize videos.

2.1 Enhanced video browsing

Enhanced video browsers improve upon regular browsing methods by incorporating additional features. Many of these features are focused on navigation within a video, such as controllable playback speed, time jumping, and timeline visualization. For instance, Schoeffmann and Boeszoermenyi [30] introduced navigation summaries that visualize extracted meta-data in a temporal manner in the form of a timeline. Users can select summaries ranging from low-level details(e.g. motion intensity, dominant colors) to high-level details(e.g. positions of commercial breaks, emotions of actors/actresses) based on their preferences. Other studies investigated interaction with the timeline of a video. Hurst et al. [15] and Hurst and Jarvers [18] address the problem of granularity in traditional timeline interfaces. Traditional timeline interfaces can be hard to navigate with a zoomed-out scale. However, with a zoomed-in scale, the user cannot scroll across the whole video. Their work resulted in an interface called ZoomSlider, which was later improved by Hürst [14] and called AV-ZoomSlider. In this interface, users can scroll the timeline by moving the cursor horizontally and change the scale of the timeline by moving the cursor vertically. Additionally, when users reach the border of the player window, scrolling switches to speed-based browsing, where the video plays in a fast-forward fashion.

These interfaces provide advanced browsing controls that help users navigate a video. However, when browsing a video collection, users should instead be provided with a quick and clear overview of a video, often achieved by visualizing the content of a video concisely. This visualization conventionally consists of a filmstrip displaying consecutive frames from the video. Various older studies such as [8] [6], employed the filmstrip approach for video visualization. Drucker et al. [9] integrated the filmstrip method into video browsing, where a filmstrip appears while scrolling through a video, a technique now used by platforms like YouTube. More recent studies such as [2] continued exploring improvements to this approach. Hürst and Darzentas [19] introduced a novel hierarchical storyboard approach. Their interface consists of a grid of temporally ordered thumbnails to represent a video. Users can then select a thumbnail as an anchor and change the granularity of the grid, resulting in a more detailed storyboard of a segment of the video.

2.2 Browsing in Virtual Reality

A valuable source for novel browsing interfaces for video collections in VR is *The Video Browser Showdown* (VBS)[29]. The VBS is an event where teams attempt to solve content-based video retrieval tasks in a competitive setting. Although the event primarily focuses on the retrieval process, it also showcases various innovative methods for visualizing video libraries. While most teams focus on interfaces for 2D screens, some teams build interfaces that utilize VR. Two of those interfaces are *Vitrivr VR* [36] and *EOLAS* [37].

Vitrivr VR is an extension of *Vitrivr*, a content retrieval system that allows users to query video libraries. The results can be explored in VR through a cylindrical, rotatable display, where the position of the media item on the display is based on a similarity score. Media items can then be selected to pop out the media segment inspector, which allows the user to inspect the video in detail. The media segment inspector can be placed anywhere in the virtual space and displays segments of the video as frames in a box that users can riffle through. This method of visualizing the video is very similar to the filmstrip method, by displaying frames in a depth-wise manner instead of displaying them from left to right.

EOLAS represents keyframes of a video by embedding them into latent vector space. They created an interface to explore these keyframes based on spoken textual queries. This interface shows clustered groups of keyframes based on their encoded features in a 3D space. Users can navigate this space and get an initial position based on similarity in feature space.

These interfaces introduce innovative ways to explore video libraries in VR but do not consider 360° videos by relying on conventional 2D thumbnails to visualize the videos.

2.3 Browsing 360° videos

360° videos, also referred to as omnidirectional videos allow users to dynamically change their viewing direction during playback. These videos can be viewed on a computer/laptop, a phone/tablet, or through VR. Although some browsing solutions moved to 3D space, they still use regular thumbnails to represent videos. Conventionally, 360° frames are projected on a 2D plane using equirectangular projection. Vermast and Hürst [38] investigated the use of different shapes to represent a 360° thumbnail. In this study, they found that conventional equirectangular thumbnails are the most effective for recognizing general high-level concepts quickly. However, for identifying low-level details within a video, projecting the frame onto a sphere results in better accuracy without requiring additional time to complete the search task. The spheres additionally resulted in a better user experience, suggesting that using a sphere to represent frames of a video is better when considering low-level details.

Building on this work, Knoop [20] did a comparative study on three designs for a VR video browser: An abstract design based on movies, a design based on a video store, and a design based on a record store. The designs were measured on performance and user experience. Results indicated that the record store design performed worst across all measures. The video store design scores better on pragmatic quality measures, while the abstract design scores better on hedonic quality measures. However, the perceived performance did not show a significant difference. While this study showed promising results in terms of user experience, its browsing speed cannot compare to traditional browsers as users have to move physically through the virtual space. Building on this work, this research proposes a new approach for 360° video browsing in VR with a focus on both user experience and efficiency by combining the traditional grid of thumbnails with enhanced video previews.

3 APPROACH

While exploring a collection of 360° videos, a typical scenario is filtering the content using a search query, followed by selecting content from the query results. The query results are conventionally presented by displaying videos in a grid pattern using equirectangular or custom thumbnails to represent the videos, similar to how traditional videos are presented. Moreover, 360° videos offer an arguably more immersive experience through virtual reality environments compared to traditional 2D screens. Nevertheless, the unique advantages offered by virtual reality environments, such as immersive displays capable of showing content in a 3D space, remain unexplored in the context of 360° video browsing.

This research aims to improve the browsing experience for 360° videos in virtual reality to aid users in efficiently searching for specific videos, by developing and evaluating two interfaces designed to enhance efficiency and user experience. Since this field is relatively underexplored in the context of 360° videos and VR, we consider previous work in the context of enhanced video browsers for traditional videos. Numerous studies experimented with showing additional details, such as neighboring frames, color-coded information about frames, or presenting an overview of all the frames upon selecting a video.

Several studies, performed in the context of traditional videos, addressed this problem by providing a visualization of a video upon selecting it. This study follows that approach and develops two VR interfaces for browsing 360° videos that utilize the additional screen space offered by VR to visualize the content of a video. The first interface is meant to represent the state-of-the-art, translated to the context of 360° videos and VR. It is based on an approach commonly used for browsing traditional videos, where the video is visualized with a filmstrip. A filmstrip temporally presents frames of a video in a sequence. Conventionally the filmstrip is presented as a long row of frames, but studies such as [36] [21] have experimented with presenting the frames in a box that users can riffle through. The second interface takes a novel approach where a video can be hierarchically explored. This approach is motivated by studies such as [19] that visualize videos at different levels of granularity. The details of both interfaces are outlined in section 4. We conduct a comparative analysis to evaluate the effectiveness of the interfaces. The following research questions guide this comparison:

- **RQ1**: Which interface results in better user performance, considering efficiency and accuracy?
- **RQ2**: How do the user experience of the filmstrip interface and the hierarchical interface compare?

To answer these research questions, an experiment is conducted in which participants evaluate the two interfaces after performing a search task. The search task entails finding a video that contains a scene that was explored before. The data collected during the experiment will be analyzed to compare the performance and the user experience between the two interfaces. The metrics to measure the performance are inspection time per video and the correctness of the selected video. The user experience is measured with a standardized questionnaire and qualitative questions. The details of the experiment are clarified in section 5.

4 INTERFACE DESIGN

Users may have different goals while browsing a collection of 360° videos. We identify the following three use cases while designing the interface:

1. **High-level Browsing**; When a user is looking for a video on a general topic (i.e. a video of someone skiing)
2. **Detailed Browsing**; When a user is searching for a video with a specific detail (i.e. a video of someone skiing in Austria using red skis)
3. **Exploring video collection**; When a user is exploring a collection of videos but is not looking for something specific.

The traditional interface works well for use case 1. Users can grasp the high-level concept of a video from a traditional thumbnail [38]. However, users may require more details for use cases 2 and 3, so it can be beneficial to have a detailed overview of the video. In the context of traditional videos, detailed visualization is conventionally approached by displaying multiple frames horizontally or in a grid. Moreover, in the context of 360° videos, Vermast and Hürst [38] demonstrated that low-level details are more accurately discerned when frames are represented as spheres. We propose a novel interface that adds details to the traditional grid of thumbnails to improve the browsing experience across all use cases and perform a comparative analysis against an interface that represents the state-of-the-art.

4.1 Interfaces

Two interfaces have been developed. Firstly the filmstrip interface, which adapts traditional video browsing methods to the context of 360° videos and VR. Secondly, the hierarchical interface, which introduces a novel approach where users hierarchically explore the video content. While both interfaces initially present videos in the same manner, they diverge in their approach to visualizing the video details. Leveraging the additional screen space offered by virtual reality, our design integrates both high-level and low-level browsing. Instead of replacing the original grid of thumbnails, We relocate it to accommodate detailed visualizations. The detailed visualization represents frames as spheres instead of equirectangular thumbnails, as shown to be more effective for distinguishing details by Vermast and Hürst [38]. Drawing inspiration from holographic tables, our approach envisions the traditional grid of thumbnails presented to the user as if laid out on a table. Upon selecting a video, a detailed visualization of the selected video is projected above the table. This visualization is different for the two interfaces. In addition to the thumbnails and the detailed visualization, two more elements are displayed; a timeline and a "lock-in video" button. Users can interact with the timeline to navigate the video, similar to traditional video players. The "lock-in video" button is specifically to confirm a selection during the search task introduced in section 5.

4.1.1 Filmstrip interface.

Detailed visualization of traditional videos is conventionally done with a filmstrip. A filmstrip sequentially shows consecutive frames, allowing for a continuous temporal representation of the video content. In the filmstrip interface, this sequential presentation is retained. However, the frames are projected onto spheres instead of onto 2D planes. The filmstrip thus consists of a row of spheres in 3D space with each sphere representing a frame from the video. Users can interact with the timeline or grab the spheres to scroll through the video. Additionally, the spheres can be rotated using the joystick on the controller. A screenshot of this interface is presented in figure 1.

4.1.2 Hierarchical interface.

Various studies, such as [19] and [14], have experimented with different ways to give users control over granularity. We approach this problem in the context of 360° videos in virtual reality, by visualizing the video details at different layers of granularity. The detailed representation of the video consists of a hierarchy of spheres. Initially, one layer consisting of 5 spheres is shown. Each of the 5 spheres corresponds to a segment of the video. For instance, if the duration of a video is 5 minutes, each sphere represents one minute of the video. The spheres all play their assigned segment simultaneously, collectively presenting the entire video.

Dividing a video into 5 segments might not give the user enough details, especially with longer or more dynamic videos, since a segment might contain multiple scenes. So to give users control over the granularity, they can select any of the spheres to create a second layer (layer 2), which further divides the one-minute clip into five, equally long video segments. A third layer is created similarly. This creates a tree structure of the video for users to

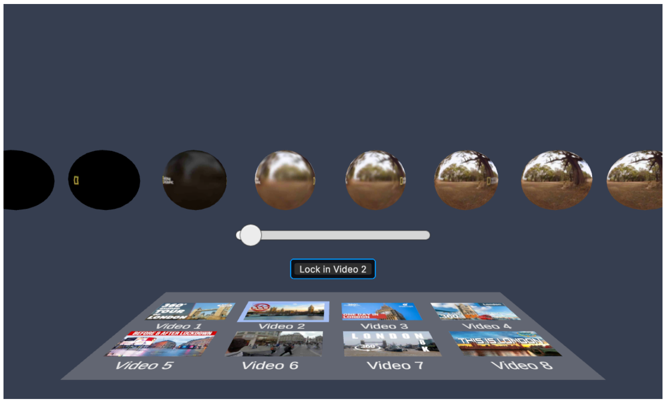


Figure 1: Screenshot of the Filmstrip Interface

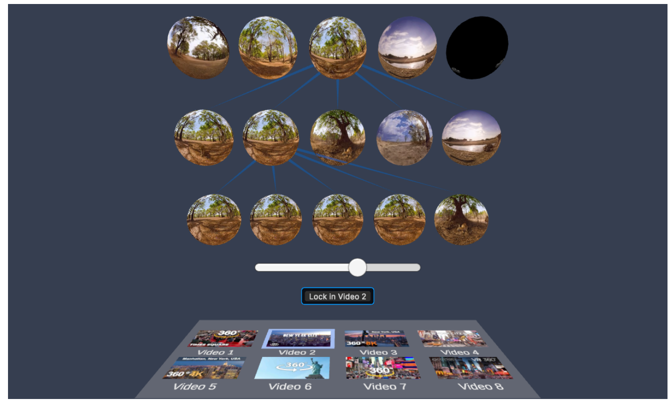


Figure 2: Screenshot of the Hierarchical Interface

interactively explore. Figure 2 shows a screenshot of the interface where all layers are activated.

5 METHODOLOGY

To answer the research questions introduced in section 3, an experiment was conducted where participants evaluated both of the interfaces (see section 4). In this experiment participants were tasked with completing a search task for each interface, followed by answering a questionnaire. During the search task, time and accuracy will be measured to evaluate the participant's performance, while the questionnaire will gather information about the user experience. To ensure that the differences in performance between the two interfaces can be attributed to the interfaces themselves rather than individual differences among participants, a within-subject design was used where the order in which participants used the interfaces was counterbalanced.

5.1 Data collection

5.1.1 User performance.
Application data is collected to answer the first research question. We gather information about the user's performance during the search task. The two primary metrics are the inspection time per video and the correctness of the selected video.

Inspection time per video refers to the time participants look at the detailed visualization of a video. By using this measure instead of total completion time, we account for the fact that participants might inspect a different number of videos between the interfaces. By analyzing the inspection time per video, we can more precisely measure the time required by participants to determine whether they have identified the correct video.

Additionally, the application saves the actions taken by participants to a log file with timestamps to provide additional insight where necessary. The logged information includes: Opening a video, rotation of spheres, changes in the timeline, and selection of spheres[1].

5.1.2 User experience.
User experience (UX) does not have a clear definition. It is a term that describes how a user perceives a system and is thus hard to measure. It is conventionally measured using either the System Usability Scale (SUS) [5], Usability Metric for User Experience (UMUX) [11], or the User Experience Questionnaire (UEQ) [23]. Schrepp et al. [35] state that for ranking two products on their UX quality, the choice of the questionnaire does not matter much, as they measure a quite similar concept. However, while UMUX and SUS focus strongly on pragmatic quality, the UEQ values pragmatic and hedonic quality equally. For this study, we use a short version of the EUQ, since we value both hedonic quality and pragmatic quality. The UEQ contains 26 7-point Likert scale questions that result in six scales. The short version (UEQ-S)[34] includes a selection of 8 of those questions and results in three scales: Overall quality, Hedonic quality, and Pragmatic quality. By employing this shorter version, we balance the need for detailed user feedback with the practical considerations of keeping the questionnaire concise and manageable for participants.

[1]Selection of spheres is only logged in the hierarchical interface as this action is not available in the filmstrip interface

5.2 Questionnaire

The questionnaire used during the experiment consists of four parts. The first part consists of an information sheet, consent form, and demographic questions. It contains questions about age, gender, occupation/degree, experience with VR, and experience with 360° videos. This part of the questionnaire will be filled out before the experiment starts.

5.2.1 During evaluation.
During the evaluation, participants answer questions about each interface individually. First, they will answer the short version of the User Experience Questionnaire (UEQ-S). Then, they will rate the interface and answer open questions about the advantages and disadvantages of the interface.

5.2.2 After evaluation.
After finishing the VR section of the experiment, participants answer final questions aimed at comparing the two interfaces.

5.3 Experiment design

5.3.1 Procedure.

Participants start with the questionnaire containing an information sheet and consent form. Once they agree to the consent form, they proceed to answer demographic questions. Participants are then assigned to either group A or group B. Group A evaluates the Filmstrip interface first, followed by the Hierarchical interface, whereas group B evaluates the Hierarchical interface first and then the Filmstrip interface.

Participants evaluate both interfaces in their assigned order by performing a search task. First, they read the provided instructions. Additional explanation is provided where necessary. Participants start with a tutorial task after adjusting and putting on the VR headset. In the tutorial, participants are introduced to the search task while getting familiar with the interface. The tutorial is the same as the actual search task, with two exceptions. Firstly, the tutorial uses four videos, whereas the actual task uses eight, so participants can find the correct video relatively quickly in the tutorial. Secondly, in the tutorial participants are told to explore the interface and formulate a strategy. In contrast, in the actual test, participants are tasked with finding the correct video as fast as possible.

The search task consists of two steps. In the first step, participants explore a still frame from one of the videos in VR. They are told to remember as much of this frame as possible. In the second step, participants use the assigned interface to figure out which of the eight videos contains the frame that they have just explored. The videos are presented in random order for each participant. However, the order is consistent between interfaces. Thus, while the position of the correct video varies for each participant, it remains identical across both interfaces. Participants click on videos, inspect the detailed visualization, and lock in their choice once they have found the correct video.

Because the UEQ has to be answered right after using the interface, participants return to the computer to fill out the next part of the questionnaire right after the search task. This also gives them a break from using VR, which helps reduce the risk of motion sickness. After performing the last search task, participants answer the final part of the questionnaire with questions focused on comparing the two interfaces.

5.3.2 Environment.

The experiments were conducted in person due to the experiment requiring a VR environment. This allowed for a neutral environment with minimal distractions and assistance when necessary. During the VR portion of the experiment, participants maintained a seated position on a non-rotating chair.

5.4 Participants

A total of 32 participants (56% male and 44% female) were selected for this study through convenience sampling. The participants, all aged between eighteen and thirty, were chosen due to the likelihood of being early adopters of new technologies. Specifically, 22 participants were in the age group of 20-24 years old (68.75%), while 10 participants were in the age group of 25-30 years old (31.25%). 10 participants never used a VR headset before the study (31.25%), while the majority had prior experience. 19 participants (59.38%) indicated to have used a VR headset a couple of times, one occasionally used a VR headset and two participants regularly used a VR headset. Only nine of the participants with prior experience used the VR headset to watch 360° videos (28.13 %). 18 participants watched 360° videos on a phone or tablet (56.25%) and 14 participants watched 360° videos on a computer or laptop (43.75%). Seven participants never watched 360° videos on any platform (21.88%).

5.5 Materials and Technical details

The application, containing both interfaces and a VR viewer, has been implemented in Unity version 2022.3.10f1[2]. Additionally, the VR template provided by Unity, which includes the XR toolkit has been used during development, as it provides a framework for VR experiences. The application is built for the Oculus Quest 3 headset[3].

5.5.1 Videos.

The videos selected for this study must be sufficiently similar to prevent participants from identifying the origin of a specific frame based solely on the thumbnails since this defeats the purpose of a detailed visualization. We have chosen to use videos showcasing specific cities due to their comparable content and general thumbnails. Specifically, tutorials feature videos of Paris and Prague, while videos of New York and London are used for the actual search task. The selection process involves initiating a YouTube search with an arbitrarily chosen city as the search term, filtering the results for 360-degree videos, and then selecting videos based on specified criteria: they must have a duration of 15 minutes or less, have a 4K resolution and must have varying scenes throughout the video. A duration of less than 15 minutes is required to ensure that the lengths of the videos do not vary significantly from each other and to conserve memory space, given the limited working memory of the Oculus Quest 3. The 4K resolution ensures clarity when participants view a video frame in the VR viewer during the search task. Varying scenes are essential as the experiment requires participants to identify a specific video based on a frame; without scene variation, the task would be too easy, defeating the purpose of detailed visualization.

This selection process resulted in 24 videos: 16 for the search tasks and 8 for the tutorials, all downloaded from YouTube. To conserve memory, we scaled down the videos to a resolution of 480x240, which is low enough to save memory while maintaining sufficient detail. Additionally, videos are converted to an equirectangular format if they were not originally formatted as such. Both tasks, as well as extracting the frame shown in the VR viewer are accomplished using FFMPEG[4].

6 RESULTS

6.1 Performance

A timed search task was used to measure participants' performance with each interface. Given that we cannot control the order in which participants select videos, we focus on the time spent per video rather than the total time spent within the interface. This metric

[2]https://unity.com/
[3]https://www.oculus.com/
[4]https://ffmpeg.org/

reflects the duration participants need to determine whether a video contains the frame they are searching for. The results are presented in table 1. During the experiments, four participants experienced a problem with the hierarchical interface that caused the time for the first few inspected videos not to be recorded. However, since our metric is the inspection time per video instead of the total time, their partial data can still be used to give insight into their performance.

Due to the relatively small sample size, a Shapiro-Wilk test was conducted, revealing a significant deviation from normality for both the filmstrip interface (W= .90, p= .008) and the hierarchical interface (W= .88, p= .002). Therefore, a non-parametric test, namely the Wilcoxon Singed-Ranks test, was selected as the appropriate analysis method. The Wilcoxon Signed-Ranks Test indicated that the average time spent per video in the filmstrip interface was statistically significantly higher than in the hierarchical interface (Z= -2.693, p= .007).

Additionally, for each interface, we report the correctness of the video guessed during the search task. For the hierarchical interface, only one participant guessed the wrong video (96.88 % correctly guessed), while two participants guessed an incorrect video for the filmstrip interface (93.75 % correctly guessed). A McNemar's test determined no significant difference in the accuracy of the two interfaces (p= 1.00).

	N	Mean	Std. Deviation	Minimum	Maximum
Filmstrip time per video	32	00:35.87	00:23.77	00:06.30	01:39.55
Hierarchical time per video	32	00:26.17	00:16.29	00:06.48	01:08.66

Table 1: Descriptive statistics for time spent per video in a timed search task, reported in mm:ss.ms

6.2 User Experience

6.2.1 UEQ-S.

The results of the UEQ-S questionnaire are analyzed using an analysis tool provided by the authors [32]. This tool first transforms the questionnaire responses from a 7-point Likert scale to values ranging from -3 (most negative) to +3 (most positive). With these values, the mean Pragmatic, Hedonic, and Overall quality is calculated and reported in table 2.

	Pragmatic Quality	Hedonic Quality	Overall Quality
Filmstrip interface	0.77	0.65	0.71
Hierarchical interface	1.11	1.92	1.52

Table 2: Average pragmatic, hedonic, and overall quality scores obtained from UEQ-S questionnaire analysis

The results are conventionally evaluated against a benchmark dataset, which includes data from 468 studies involving over 21,000 participants [33]. This benchmark, developed using the full UEQ, categorizes results from bad (worst 25% of studies) to excellent (best 10% of studies). The results of the UEQ-S are presented against the benchmark in figure 3. The Hierarchical interface scored just below average on the pragmatic scale. However, it ranked excellent on the hedonic scale, resulting in a good score for the overall quality. The filmstrip interfaces scored below average on all scales. The provided analysis tool performs T-tests to see if the differences between the two interfaces are significant. The t-tests show no significant difference between the Hierarchical and Filmstrip interface for the pragmatic quality (p= .27). However, it does show a significant difference for the Hedonic quality (p= .0001) and the overall quality (p= .002).

Alongside the UEQ-S, participants evaluated each interface by rating it on a scale from 1 to 10. The filmstrip interface received an average rating of 6.31, while the hierarchical interface received an average rating of 7.63. A Wilcoxon-signed rank test revealed a significant difference between these average ratings (p = .002).

6.2.2 qualitative data.

To gain further insights into the interfaces, participants were asked to specify their preferred interface and to discuss the advantages and disadvantages of each. 24 participants (75%) favored the hierarchical interface, while 8 participants (25%) preferred the filmstrip interface. The main reasons mentioned by participants for preferring the filmstrip interface are familiarity, speed, and control over the scrolling speed. In contrast, the main reasons mentioned by participants for preferring the hierarchical interface are speed, simplicity, and a better overview of a video. Furthermore, participants noted various advantages and disadvantages of both interfaces. Table 3 highlights the most notable ones.

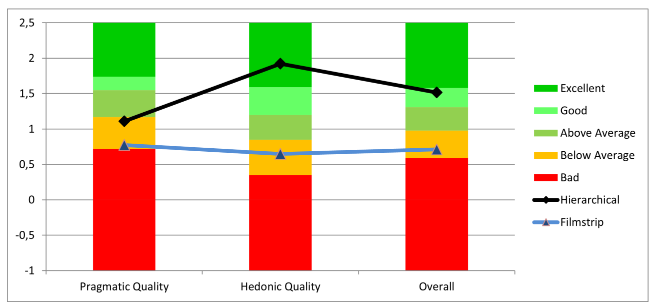


Figure 3: UEQ-S results for the hierarchical interface and the filmstrip interface compared to a benchmark dataset

6.3 Analysis of search patterns

Analyzing the search patterns used during the search task provides us insights into the strengths and weaknesses of the interface, potentially revealing areas for improvement. In the filmstrip interface, the search pattern is relatively straightforward with limited room for variation. After opening the visualization, participants can either scroll or skip through a video and make a decision based on their observations. In contrast, the hierarchical interface has more varied search patterns. We can distinguish these different search patterns by analyzing the time participants spent on each layer.

To analyze participants' search strategies, we categorized them based on the proportion of time spent in each layer. Participants were assigned to a strategy category by comparing the relative time spent across the layers. For instance, if a participant spent the majority of their time in Layer 1, a moderate amount in Layer 2, and

Advantages and disadvantages mentioned by participants	Filmstrip Mentions	Hierarchical Mentions
This interface provides a **clear** overview of video content	3	9
This interface provides an **quick** overview of video content	7	8
This interface is **simple/easy** to navigate	7	5
Scrolling the timeline was very sensitive, causing spheres to move very fast	8	0
The interface was complex	0	5

Table 3: Notable advantages and disadvantages for both interfaces

the least time in Layer 3, their strategy was denoted as L1>L2>L3. To evaluate the effectiveness of each strategy, we compare each participant's strategy to their time spent per video (see 6.1). Table 4 presents the number of participants for each search strategy alongside the average time spent per video.

We found that the majority of participants spent most time in Layers 1 and 2. Additionally, the average time spent per video appears to be noticeably higher for participants that spent most or second most time in layer 3.

Strategy	# Participants	Average time spent per video	Std. Deviation
L1>L2>L3	11	00:24.32	00:14.56
L2>L1>L3	9	00:23.23	00:14.44
L2>L3>L1	5	00:35.59	00:21.90
L3>L2>L1	3	00:43.46	00:12.27
L1>L2=L3	4	00:13.15	00:05.21

Table 4: Distribution of participants across different search strategies based on time spent in each layer of granularity

7 DISCUSSION

This study aims to answer two research questions about improving the browsing experience for 360° video libraries in VR. The first research question is aimed at the performance of the user, while the second research question is aimed at the user experience.

7.1 RQ1: Performance

To answer Research Question 1: *"Which interface results in better user performance, considering efficiency and accuracy?"*, we consider the results gathered from the search task (section 6.1).

Regarding accuracy, no statistically significant difference was found. Nearly all participants identified the correct video, with the exception of one participant using the hierarchical interface and two using the filmstrip interface.

In terms of efficiency, the results indicate that participants were, on average, statistically significantly faster at identifying the correctness of a video with the hierarchical interface. This finding suggests that despite its complexity, the hierarchical interface conveys the content of a video quicker than the filmstrip interface. This is likely due to the hierarchical interface displaying five segments of a video simultaneously.

The filmstrip interface is relatively straightforward. Participants with longer search times often skipped over the correct frame while scrolling through the video, causing them to re-examine all videos. The hierarchical interface is more complex, and section 6.3 demonstrates that the strategy selected by participants influences the time required to identify the correct video. Most participants prioritized layers 1 and 2, which appears to be more effective than prioritizing layer 3. The longest video used in this study is shorter than 15 minutes. A 15-minute video results in 3-minute segments in layer 1, 36-second segments in layer 2, and 7.2-second segments in layer 3. So it makes sense for participants who are prioritizing layer 3 to be less effective because all five spheres in layer 3 will likely show the same scene, especially since most videos used in the study are even shorter than 15 minutes. So for this study, 3 layers most likely provided too much detail. However, the third layer might be effective for analyzing longer videos.

While observing the experiment, three strategies seemed to be the most effective, all prioritizing layer 1 and/or layer 2. For the first strategy, participants exclusively used layer 1 along with the timeline to scroll through the video. In the second strategy, participants explored the video by systematically selecting each sphere in the first layer and thus utilizing mostly the second layer. The final observed strategy used a top-down approach, where participants selected spheres in the first layer only if they had content, or attributes, such as brightness, similar to that of the correct frame. For example, they would skip videos with aerial shots when searching for a frame in the middle of a city.

7.2 RQ2: User Experience

To answer Research Question 2: *"How do the user experience of the filmstrip interface and the hierarchical interface compare?"*, we consider the results from the questionnaire presented in section 6.2.

The results from the UEQ-S indicated that the hierarchical interface has a higher pragmatic quality than the filmstrip interface, although this difference is not statistically significant. The hierarchical interface does have a statistically significantly higher hedonic quality and overall quality compared to the filmstrip interface. Moreover, participants rated the hierarchical interface statistically significantly higher on a scale from 1 to 10, and 75% of the participants preferred the hierarchical interface over the filmstrip interface. Interestingly, participants often explained their preference with the same reasons for both interfaces: speed and simplicity.

A similar trend is observed when considering the advantages and disadvantages mentioned by participants. The advantages of quickly getting an overview of the video and easily navigating the interface are mentioned by roughly the same amount of participants for both interfaces. The advantage of a clear overview provided by the interface is mentioned three times as often for the hierarchical interface. This can be interpreted as both pragmatic and hedonic quality since it influences both the efficiency and the aesthetic of the interface.

Some participants mentioned complexity as a disadvantage of the hierarchical interface. However, the majority of these participants performed better on the search task in this interface, suggesting that the complexity is manageable. A disadvantage of the filmstrip that was mentioned many times is that scrolling the timeline is very sensitive. This caused spheres to move fast, especially for longer videos. This likely contributes to the lower hedonic quality of the filmstrip interface.

7.3 Limitations & Future Work

This study has several limitations that should be considered when interpreting the results. Firstly, there were some minor issues with the implementation due to Unity's inefficient handling of video players. The Oculus Quest was unable to play 15 videos simultaneously, likely because of memory constraints, so only the 5 videos of the most recent layer were played while the videos in the previous layers were paused. Additionally, three participants experienced crashes, but these had minimal impact on the results as they were able to continue after a quick restart. Another issue arose during the search task. The users' view was streamed to a phone for observational purposes, but in some cases the connection was unstable, causing the stream to lag. It is important to note that these observations were not the primary source of data collection and were intended only to provide potential additional insights. Therefore, the impact of this limitation on the overall findings of the study is minimal.

Another limitation of this study is the relatively short time participants had to learn how the interfaces work. This most likely does not have a big effect on the results of the filmstrip interface since it is similar to existing solutions and relatively straightforward. However, more time in the hierarchical interface might have had a bigger effect on the results since the strategy seemed to have an impact on the performance as discussed in section 6.3.

7.3.1 Filmstrip interface.

Several participants indicated difficulty with the sensitivity of the timeline in the filmstrip interface. Improvements to the filmstrip interface could address the issue, for example, by not moving the spheres, but instead only changing the content shown on the spheres. The interface could also take inspiration from the hierarchical interface by introducing adjustable granularity. Another approach could involve increasing the visibility of the spheres, for example by arranging them in a cylindrical shape or an S-shape. This could help participants identify where a new scene begins and scroll to that point.

7.3.2 Hierarchical interface.

The hierarchical approach has proven effective. However, alternative variations of the visualization might further improve the interface. For instance, section 6.3 showed that using only the first layer seems to be a very effective strategy. Therefore, creating a visualization only using just one layer, possibly with more spheres, might be beneficial. Another variation that might be interesting to explore could involve adding a filmstrip layer instead of a hierarchical layer upon selecting a sphere. The biggest problem of the filmstrip interface was the sensitivity of the timeline, but by adding it as a second layer underneath the hierarchical interface, it would be less sensitive since the length of the segment would be shorter.

8 CONCLUSION

This study aimed to enhance the browsing experience of 360° video collections in VR. For this purpose, we designed and implemented two interfaces: the filmstrip interface and the hierarchical interface. The interfaces are evaluated on their efficiency and usability. To evaluate the user experience, the UEQ-S is employed in combination with qualitative questions. The performance was measured with a timed search task, where participants were required to identify a video containing a previously viewed frame.

The results indicate that participants completed the search task statistically significantly faster using the hierarchical interface compared to the filmstrip interface. In terms of accuracy, no significant differences were found, as nearly all participants successfully identified the correct videos with both interfaces. The results from the UEQ-S show no significant difference in pragmatic quality. However, the results demonstrate that the hierarchical interface has a statistically significantly better hedonic and overall quality when compared to the filmstrip interface. Additionally, qualitative feedback demonstrates a preference for the hierarchical interface among participants.

These findings indicate that the hierarchical interface results in a better performance and user experience, suggesting that using a hierarchical approach provides a more enjoyable and efficient browsing experience for 360° video libraries in VR environments. In conclusion, while both interfaces have their strengths, the hierarchical interface stands out for its efficiency and user satisfaction. Continuing development and refinement of this interface will be essential in advancing VR technology for 360° video browsing and offering users a more intuitive experience.

Future work could explore further refinements to the hierarchical interface and address the sensitivity issues identified in the filmstrip interface, potentially enhancing the effectiveness and user experience of both methods.

REFERENCES

The complete list of references is provided at the end of this document.

A APPENDIX

A.1 Experiment Procedure steps

This section describes the procedure during the experiment. These steps served as a guideline for the researcher to promote consistency between participants.

(1) Check the battery of the headset and controllers.
(2) Have the participant fill out the first part of the evaluation.
(3) Have the participant read the instructions and answer any questions.
(4) Run experiment for the correct interface.
   (a) Prepare application (select correct user ID and interface).
   (b) Give controllers to the participant and explain controls (mention to look at them in VR for reminder).
   (c) Give VR headset and help to adjust.
   (d) Start screen casting to your phone to see what the participants see.
   (e) Make sure the participant tries out all the functionalities in the tutorial round.
      (i) Rotating spheres.
      (ii) Resetting orientation.
      (iii) Timeline manipulation.
      (iv) (filmstrip) Grab & moving spheres.
      (v) (hierarchical) Generating layers.
(5) Have the participant fill out the evaluation of the interface.
(6) Redo steps 3 to 5 for the second interface.
(7) Reset experiment.
   (a) Move generated files from headset to laptop.
   (b) Clean VR headset.
   (c) Charge headset while waiting for the next participant.

A.2 Videos used during the experiment

Table 5 presents all the videos used during the experiment. The first 16 videos are used for the two search tasks, and the last 8 videos are used for the tutorials. The videos that are in bold are the videos that participants have to find during the search task. All videos are downloaded from YouTube using Youtube-dl[5], Figure 4 shows the code used to download the video. Additionally, it shows the code used to convert video from YouTube's cubemap format to equirectangular format and the code for extracting a specific frame from the videos, both executed with FFMPEG[6].

[5]https://github.com/ytdl-org/youtube-dl
[6]https://ffmpeg.org/

Video Name	URL	Duration
The Ultimate 360 VR Tour of London (with interesting facts!)	https://www.youtube.com/watch?v=VPdzOmOtvSA	6:14
London, United Kingdom. Virtual travel. 360 video in 8K	https://www.youtube.com/watch?v=KGerjHMa90s	15:32
London City | United Kingdom | 360 VR 8K	https://www.youtube.com/watch?v=p5KpTx9rRRg	4:49
Amazing 360 Guided Tour of London	https://www.youtube.com/watch?v=sEWWHfWBVnk	10:41
London Walk V - VR/360°	https://www.youtube.com/watch?v=J6m58Bpd-mI	5:49
London Before & After Lockdown 360 Video.	https://www.youtube.com/watch?v=v08bmAfBskk	4:10
One Day in London Trailer - VR/360° guided city tour (8K resolution)	https://www.youtube.com/watch?v=v_4_gdbltyM	1:55
This is London (360° VR)	https://www.youtube.com/watch?v=uojSw8niWc8	7:01
Escape Now: New York City in 360° VR | A Guided Tour of the City That Never Sleeps	https://www.youtube.com/watch?v=0qtdheM1cGw	5:14
New York City 8K - VR 360 Drive	https://www.youtube.com/watch?v=2Lq86MKesG4	14:41
360° Video, Manhattan, New York, USA, 4K aerial video	https://www.youtube.com/watch?v=YM6GTu_RcWM	4:28
360° New York City : Evening Walk in Times Square	https://www.youtube.com/watch?v=mNnTt5NVDC4	11:28
NEW YORK CITY - Quick Sightseeing in 360°	https://www.youtube.com/watch?v=hYV8GiZA6i0	4:22
Times Square Virtual Tour | VR 360° Travel Experience | New York City Manhattan	https://www.youtube.com/watch?v=RWNK5oozT1A	4:37
New York, USA. City of Skyscrapers. 360 8K aerial video	https://www.youtube.com/watch?v=swkwX0LMF5I	4:20
[4K 360°] TIMES SQUARE New York City Walking Tour April 2022	https://www.youtube.com/watch?v=s9YvlkY6pnk	7:08
PARIS 360°/ 4K / VR (VIRTUAL REALITY)	https://www.youtube.com/watch?v=LC8ZPLf8Yt4	3:50
Escape Now: Paris in 360° VR | An Enchanting Guided Journey Through the City of Lights	https://www.youtube.com/watch?v=EkshFcLESPU	4:31
360 VR Tour | Paris | Eiffel Tower | Tour Eiffel | All levels | Air panoramic view | No comment tour	https://www.youtube.com/watch?v=nbD8XYTw23Y	6:32
Virtual guided tour of Paris 360 VR Video | Eiffel Tower | Must Visit Bucket List in France	https://www.youtube.com/watch?v=qOMCY5drCqY	8:52
BEST of Prague, 360° Virtual Tour	https://www.youtube.com/watch?v=6-pFvwQG8IM	3:18
Prague 360° Tour Czech Republic (Watch in 5k)	https://www.youtube.com/watch?v=Z8GWVvRtp-s	5:44
Prague Castle 360 VR Tour: Explore the World's Largest Castle	https://www.youtube.com/watch?v=U9tRLfXCwXk	5:17
Prague tourism | Virtual guided tour of Prague 360° VR Video | Charles Bridge | Europe travel 4K	https://www.youtube.com/watch?v=hzohBkJGMUQ	12:18

Table 5: List of videos used during the experiment

```
# Downloading videos from YouTube
youtube-dl -i --write-thumbnail -f bestvideo --no-playlist -a .\VideoLinks.txt

# Converting to equirectangular and rescaling
for %f in (*.*) do ffmpeg -i "%f" -vf "v360=c3x2:e:cubic:in_forder=lfrdbu :in_frot=
000313,scale=480:240,setsar=1:1" -c:v libx264 -crf 21 -c:a copy -movflags +faststart "%~nf_converted.mp4"

# Rescaling (in case video is already equirectangular)
for %f in (*.*) do ffmpeg -i "%f" -vf "scale=480:240,setsar=1:1" -c:v libx264 -crf 21 -c:a copy -movflags
+faststart "%~nf_converted.mp4"

#Extracting frame from video
ffmpeg -ss 1:22 -i [input] -frames 1 -q:v 0 -c:v png FrameToFind.png
```

Figure 4: Code used to download and convert videos

A.3 Experiment Instructions

The next two pages present the instructions that were given to participants during the experiment. They were given right before the search task.

Filmstrip Interface

Task description: You will be shown a frame(still-image) of a 360-degree video**.** Look around carefully in all directions and try to remember what you see, because your **task** is to **find the video** that this frame belongs to **as fast as you can**.

The interface presents 8 videos as thumbnails in front of you on a table (Fig. 1, A). Clicking on any of these thumbnails with the [Trigger button] (Fig. 2) opens the filmstrip visualization of that video, where frames of the video are sequentially visualized as spheres (Fig. 1, B). Points 1 – 4 in Figure 1 explain the details of the filmstrip visualization:

1) Video 2 was selected on the table (A), which means that the filmstrip visualization of the second video is shown (B).
2) This is the filmstrip visualization; each sphere shows a frame of the video. The spheres can be rotated with the [Joystick]. You can also use the [Grab button] (Fig. 2) to grab and move the spheres along the timeline.
3) This is the timeline representing the whole video. Using the [Trigger button] (Fig. 2), you can scroll through the video by dragging the slider or jump to a specific position by clicking anywhere in the timeline.
4) Once you find the correct video (i.e., the one containing the frame that you initially saw), you should click this button.

Before you start the actual test, you will have a **tutorial round** to get familiar with the interface. (The tutorial round will only have 4 videos, not 8). In the tutorial, you can try all the functionalities and then continue with the practice task. During the tutorial, you can also ask questions if any part of the task or interface is unclear!

Figure 1 Screenshot Filmstrip interface (annotated).

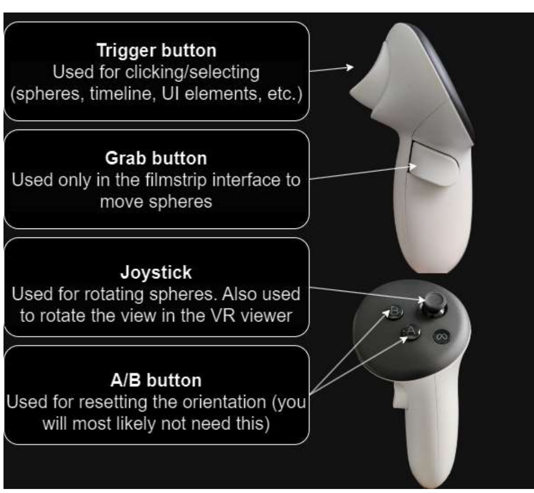


Figure 2 Interaction options.

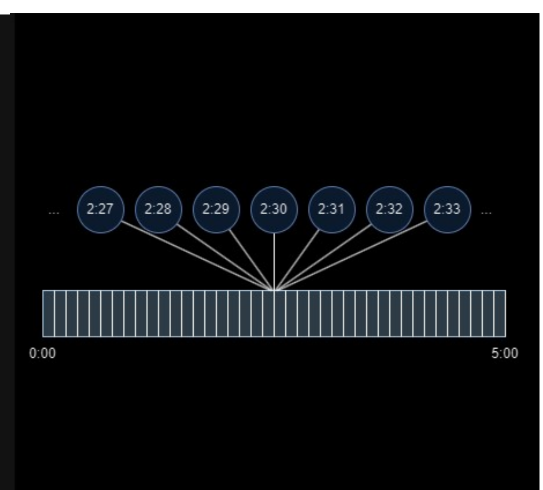


Figure 3 Illustration of the filmstrip interface.

Hierarchical Interface

Task Description: You will be shown a frame(still-image) of a 360-degree video**.** Look around carefully in all directions and try to remember what you see, because your **task** is to **find the video** that this frame belongs to **as fast as you can**.

The interface presents 8 videos as thumbnails in front of you on a table (Fig 1, A). Clicking on any of these thumbnails with the [Trigger button] (Fig. 2) opens the hierarchical visualization of that video, where segments of the video are visualized as spheres, sorted by time, and played back automatically (Fig. 1, B). Points 1 – 4 in Figure 1 explain the details of the hierarchical visualization:

1) Video 2 was selected on the table (A), which means that the hierarchical visualization of the second video is shown (B).
2) This is the hierarchical visualization. It allows you to look at the details of the video in up to three different layers of granularity. The screenshot (Fig. 1) shows an example where all three layers are activated. (See Figure 3 for an abstract illustration of the interface). When you click on a thumbnail, the first layer (Layer 1) will be shown. Each of the five spheres in this layer corresponds to a segment of the video. For instance, if a video is 5 minutes long, each sphere represents one minute of the video. Clicking on a sphere with the [Trigger button] (Fig 2) creates the second layer (Layer 2), which further divides the one-minute clip into five, equally long video segments. Layer three is created similarly if one clicks on a sphere in layer 2.
3) This is the timeline for the lowest layer (layer 3 in this example). Using the [Trigger button] (Fig. 2), you can scroll through the video by dragging the slider or jump to a specific position by clicking anywhere in the timeline.
4) Once you find the correct video (i.e., the one containing the frame that you initially saw), you should click this button.

Before you start the actual test, you will have a **tutorial round** to get familiar with the interface. (The tutorial round will only have 4 videos, not 8). In the tutorial, you can try all the functionalities and then continue with the practice task. During the tutorial, you can also ask questions if any part of the task or interface is unclear!

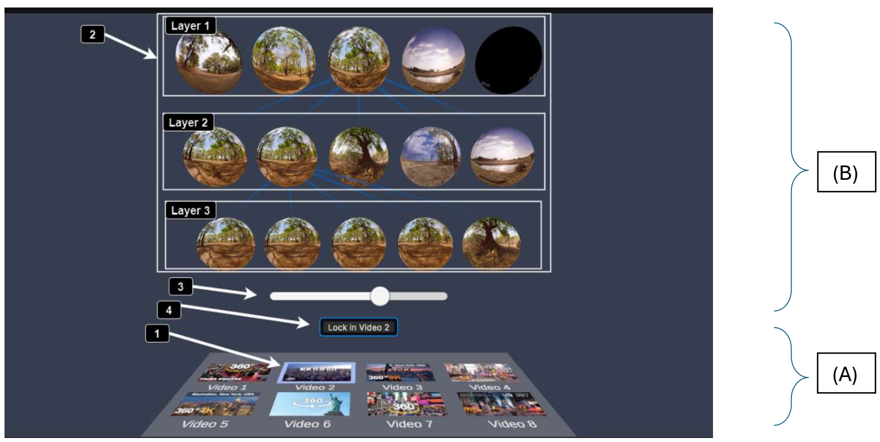


Figure 1 Screenshot Hierarchical interface (annotated)

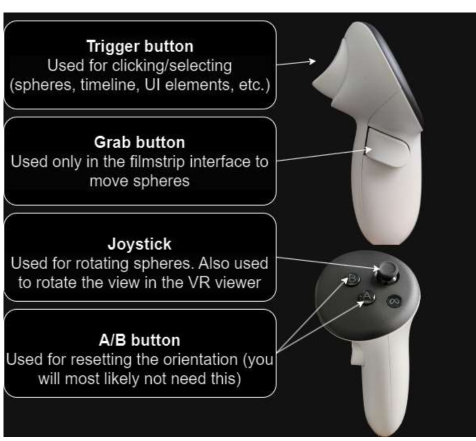


Figure 2 Interaction options.

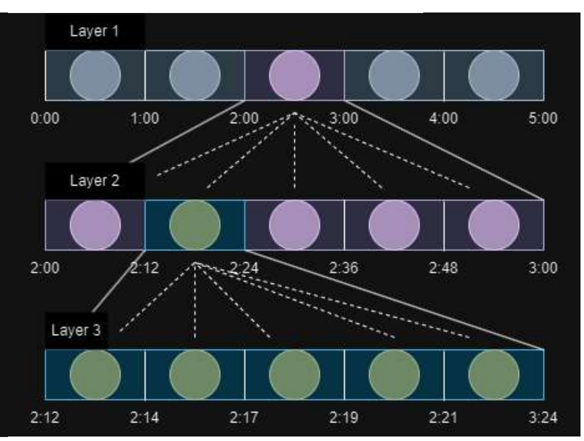


Figure 3 Abstract Illustration of the Hierarchical interface.

A.4 Counterbalance order

Table 6 presents the order in which the interfaces were presented to participants during the search task. The interface order was assigned with the questionnaire, which kept the number for both groups equal.

Participant	Filmstrip interface	Hierarchical interface
1	1	2
2	2	1
3	1	2
4	2	1
5	2	1
6	1	2
7	1	2
8	2	1
9	1	2
10	2	1
11	1	2
12	2	1
13	2	1
14	1	2
15	2	1
16	1	2
17	2	1
18	1	2
19	1	2
20	2	1
21	2	1
22	1	2
23	1	2
24	2	1
25	1	2
26	2	1
27	1	2
28	2	1
29	1	2
30	2	1
31	1	2
32	2	1

Table 6: Order in which the interfaces were presented to the participants

A.5 Questionnaire

The following pages contain the questionnaire used during the experiment, described in section 5.2. The questions about the specific interface (pages 22 - 26) are repeated for both interfaces. The question on page 25 is only displayed if the previous question "Did you experience any problems with the interface" is answered with "Yes".

Tools

Taal

Nederlands

Information Sheet

Title: Enhancing 360-Degree Video Collection Exploration: A Novel Interface for Immersive Virtual Reality Browsing

You have been invited to participate in a scientific study on enhancing the browsing experience for 360-degree videos in virtual reality (VR). You will evaluate two different interfaces by performing a search task using the Oculus Quest 3 VR headset and fill out related questionnaires. The experiment will take about 30 minutes.

This study is carried out by Bas van Eck(b.vaneck1@students.uu.nl) as part of my master thesis under the supervision of Wolfgang Hürst(huerst@uu.nl).

PARTICIPATION

The purpose of this study is to collect and analyze data about how people interact with the two interfaces. You have the freedom to withdraw at any time, without providing a reason. In such cases, your results and information will be deleted and will not be used for the study. Opting out or choosing not to participate will have no negative consequences for you. You will receive no direct benefits from participating in this study.

RISKS

Participating in virtual reality experiences may induce motion sickness symptoms like nausea and dizziness. If you experience any discomfort, please cease activity immediately.

CONFIDENTIALITY

Your data is stored in a password-protected environment. Your data will be stored anonymously, meaning that after processing the data, the data can no longer be traced back to you or your answers.

If you have any questions or comments regarding this evaluation, please contact me at **b.vaneck1@students.uu.nl**

Volgende pagina

Uitgevoerd met Qualtrics

Tools

Taal

Nederlands

*Consent form for participation in the research project "Enhancing 360-Degree Video Collection Exploration: A Novel Interface for Immersive Virtual Reality Browsing"**

Please read the statements below and tick the final box to confirm you have read and understood the statements and upon doing so agree to participate in the experiment

I confirm that I am 18 years of age or over.

I confirm that the research project "Enhancing 360-Degree Video Collection Exploration: A Novel Interface for Immersive Virtual Reality Browsing" has been explained to me. I have had the opportunity to ask questions about the project and have had these answered satisfactorily. I had enough time to consider whether to participate.

I understand that my participation in this study is voluntary and that I may withdraw from the study at any time without providing a reason, and that if I withdraw any personal data already collected from me will be erased.

I consent to allow the fully anonymized data to be used in future publications and other scholarly means of disseminating the findings from the research project.

I understand that the data acquired will be securely stored and anonymized by researcher, and will not be identifiable in any published materials.

☐ I confirm that I have read and understood the above statements, and agree to participate in the study (Check the box)

< Go Back Volgende pagina >

Uitgevoerd met Qualtrics

Tools

Utrecht University

Taal

Nederlands

How old are you? (leave empty if you prefer not to say)

*Which of the following terms best describes your gender identity?

- Male
- Female
- Non-binary
- Not listed:
- Prefer not to answer

Go Back Volgende pagina

Uitgevoerd met Qualtrics

Tools

Utrecht University

Taal

Nederlands

*What is your current field of occupation or degree of study?

Go Back Volgende pagina

Uitgevoerd met Qualtrics

Utrecht University

Taal

Nederlands

*Please indicate how often you have used each of the following technologies

	Never	A couple times	occasionally	Regularly	Very often
Used a Virtual Reality headset	○	○	○	○	○
Used a computer or laptop to watch 360-degree vidoes	○	○	○	○	○
Used a phone to watch 360-degree videos	○	○	○	○	○
Used Virtual Reality to watch 360-degree videos	○	○	○	○	○

Go Back Volgende pagina

Taal

Nederlands

*It is time for the evaluation of the first interface.

The procedure is as follows:

1. Read the instructions that will be given to you by the researcher.
2. Do a tutorial/practice task.
3. Do the actual experiment.
4. Return to this questionnaire.

Info for the researcher:

Group A

ID: 0

☐ I have finished the search task

Volgende pagina >

Tools

Utrecht University

Language
English

*Evaluate the interface

The questions consist of pairs of contrasting attributes that may apply to the interface. You can express your agreement with the attributes by ticking the circle that most closely reflects your impression.

Please decide spontaneously. Don't think too long about your decision to make sure that you convey your original impression.

Sometimes you may not be completely sure about your agreement with a particular attribute or you may find that the attribute does not apply completely to the particular product. Nevertheless, please tick a circle in every line.

It is your personal opinion that counts. Please remember: there is no wrong or right answer!

	1	2	3	4	5	6	7	
obstructive	○	○	○	○	○	○	○	supportive
complicated	○	○	○	○	○	○	○	easy
inefficient	○	○	○	○	○	○	○	efficient
confusing	○	○	○	○	○	○	○	clear
boring	○	○	○	○	○	○	○	exciting
not interesting	○	○	○	○	○	○	○	interesting
conventional	○	○	○	○	○	○	○	inventive
usual	○	○	○	○	○	○	○	leading edge

Go Back Next page

Tools

Utrecht University

Taal

Nederlands

*How would you rate this interface on a scale from 0 to 10?

Filmstrip interface

0 10

*Did you experience any problems with the interface (lagging, stuttering, videos not loading, etc.)

- No
- Yes

Go Back Volgende pagina

Uitgevoerd met Qualtrics

Tools

Taal

Nederlands

*What problem did you experience?

Go Back Volgende pagina

Uitgevoerd met Qualtrics

Tools

Taal

Nederlands

*What would you consider advantages of the filmstrip interface?

*What would you consider disadvantages of the filmstrip interface?

Go Back Volgende pagina

Tools

Language

English

Thanks! Please continue to some final questions.

Go Back

Next page

Tools

Utrecht University

Language

English

*Which interface would you prefer while browsing 360-degree videos?

- Filmstrip Interface
- Hierarichal Interface

*Why do you prefer this interface?

Do you have any other insights/ideas about the interfaces?

Thanks for your participation!

Go Back Next page

A.6 Implementation

This section contains screenshots of the application used during the experiment. The application starts with two menus meant for the researcher. One to enter the participant number (Figure 5) and the next one to select the correct interface (Figure 6).

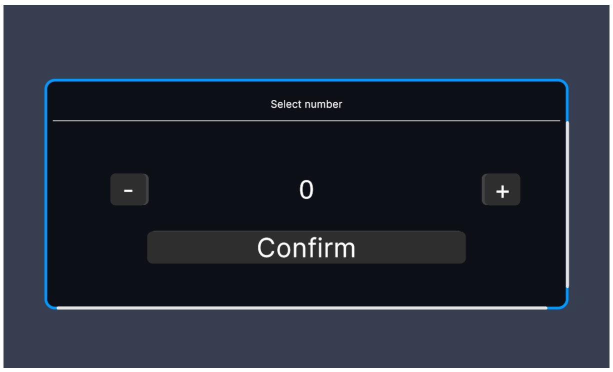


Figure 5: Screenshot participant number selection screen

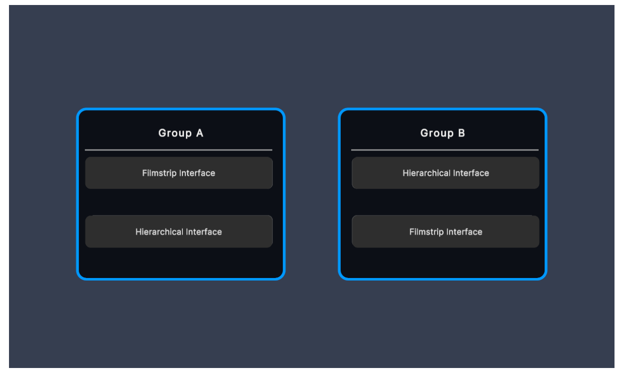


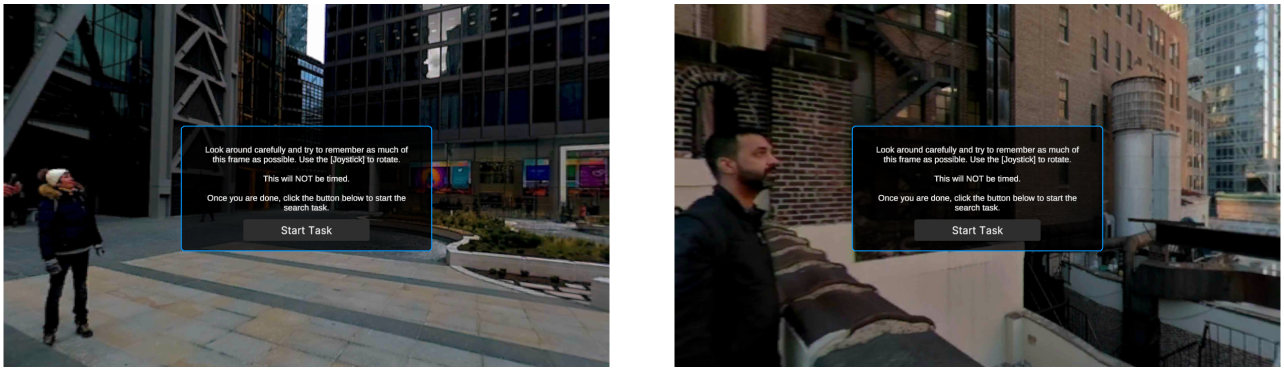
Figure 6: screenshot interface selection screen

The VR headset will then be given to the participant, who will start in a VR viewer showing the frame that they will have to find during the tutorial search task. Figure 7 presents the frames participants will have to find during the first (a) and second (b) tutorial. Keep in mind that participants can rotate to look around in VR, so they have a bigger scene to explore.

(a) **(b)**

Figure 7: (a) Screenshot of VR viewer with frames of pairs (a) and Prague (b) participants will have to find for the tutorial

Next, participants will proceed to the tutorial. The tutorial interfaces are identical to those used during the timed task, with the only difference being the number of videos involved. Therefore, we will not describe them separately here. Descriptions of the interfaces can be found in section A.6.1 and section A.6.2. For the actual search task, participants will be given a new frame to locate. Figure 8 shows the frames from London (a) and New York (b) that participants will have to find during the timed search tasks. Depending on their assigned order, participants move on to either of the interfaces.



(a) **(b)**

Figure 8: (a) Screenshot of VR viewer with frames of Londen (a) and New York (b) participants will have to find during the timed search task

A.6.1 Filmstrip interface.

Figure 9 showcases the filmstrip interface. In this example, video 5 is selected, and the timeline is scrolled to about halfway through the video. The spheres above the timeline provide a visual representation of the content of the video. Scrolling the timeline moves the spheres, displaying different parts of the video. Details of the filmstrip interface are outlined in section 4.1.1.

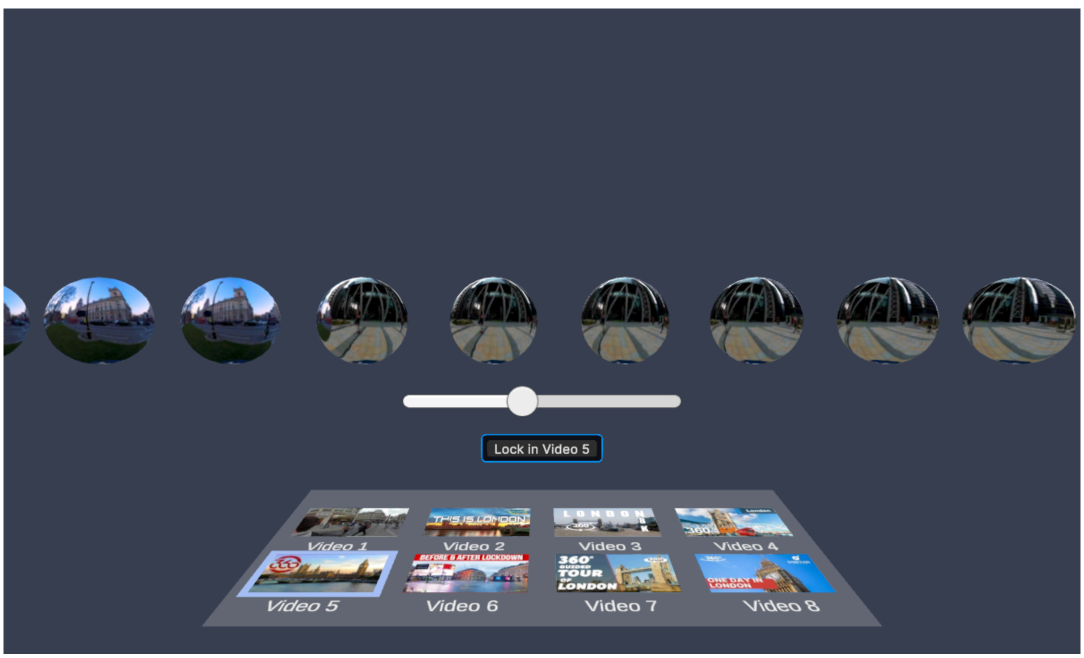


Figure 9: Screenshot Filmstrip interface

A.6.2 Hierarchical Interface.

The details of the hierarchical interface are outlined in section 4.1.2. Figure 10 showcases the hierarchical interface right after video 5 has been selected. The video has been split up into five segments and each sphere plays its assigned segment. Users can click any of the spheres to create a new layer that further divides that specific segment. Figure 11 showcases the same interface where the middle sphere of the first layer, and the second sphere of the second layer are selected.

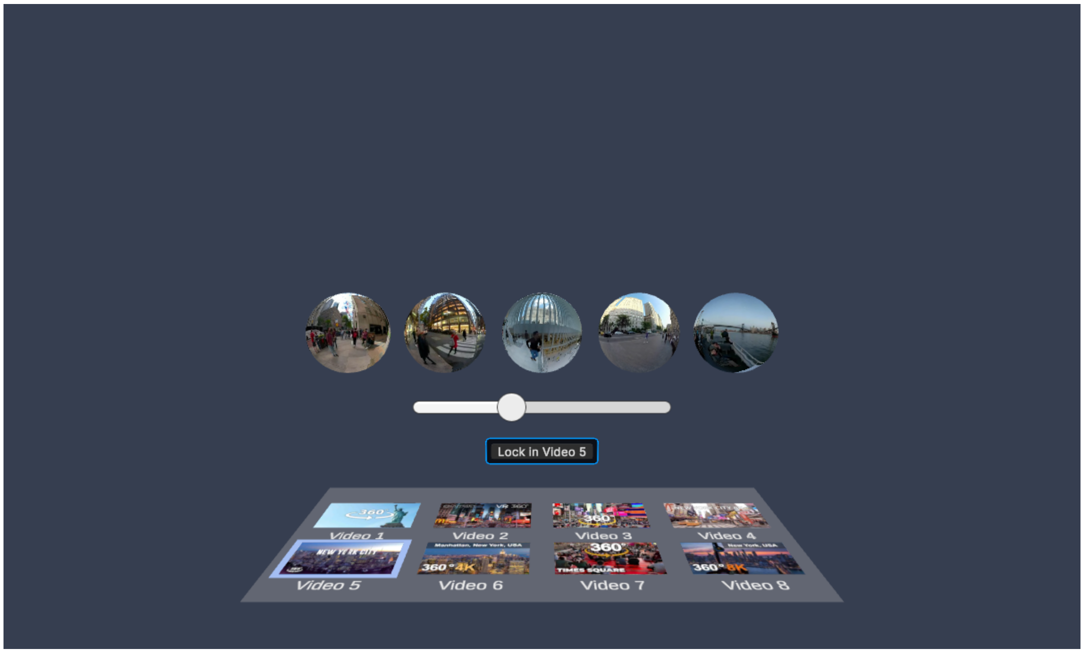


Figure 10: Screenshot Hierarchical Interface with one layer activated

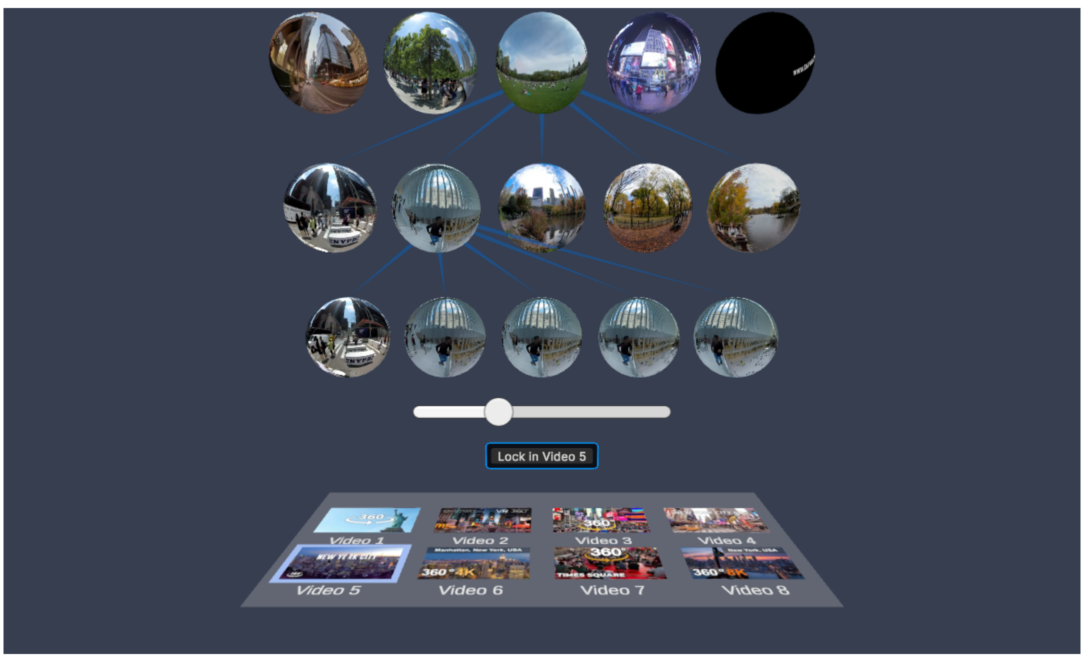


Figure 11: Screenshot Hierarchical Interface with three layers activated

B ADDITIONAL RESULTS

B.1 Demographic information

This section presents more insights into the demographics of the participants who participated in the study. Figure 12 presents the almost equal distribution of males and females from a total of 32 participants. The distribution of ages among the participants is presented in Figure 13. All participants are between the ages of 18 and 30, with 24 participants in the range of 21-25 years old and 8 in the range of 25-30 years old.

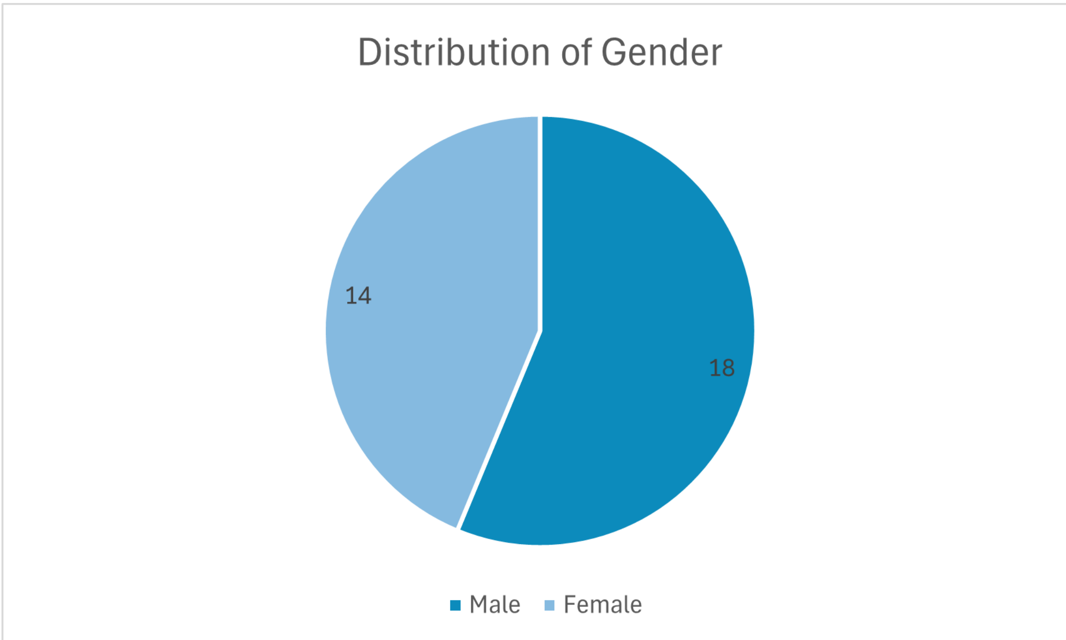


Figure 12: Distribution of gender among participants

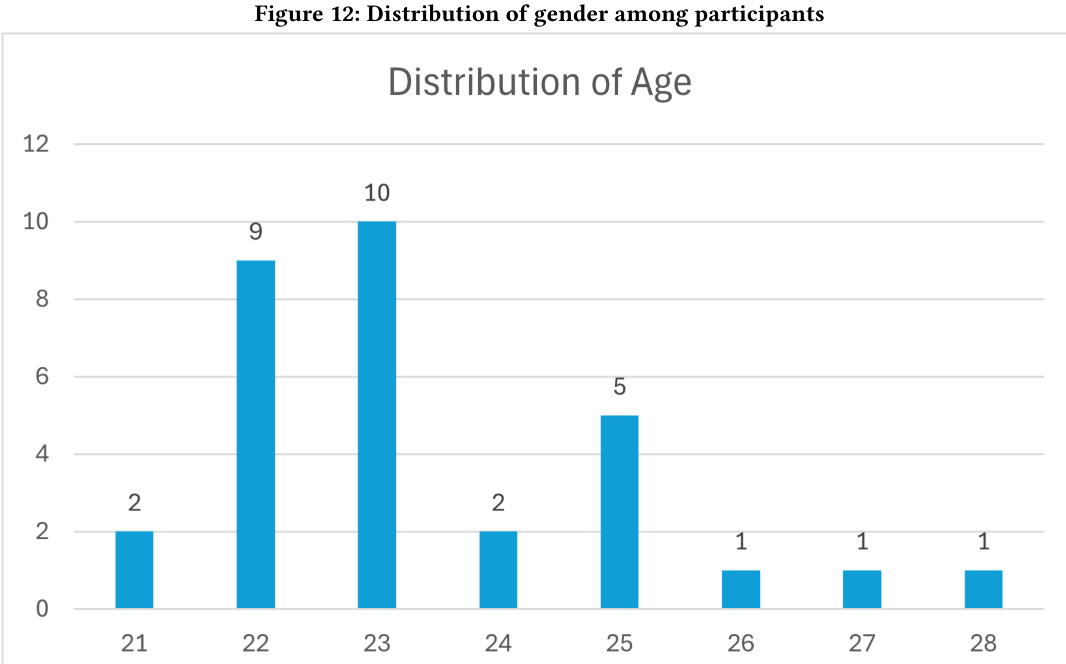


Figure 13: Distribution of age among participants

B.1.1 Familiarity with relevant technologies among participants.

The following graphs illustrate participants' familiarity with technologies relevant to this study. Figure 14 shows that most participants have either used a VR headset a couple of times or never at all, with only two participants reporting regular use. Of the 22 participants who had prior experience with a VR headset, only 9 used it to watch a 360° video (Figure 15). Participants were more familiar with watching 360° videos on a phone or computer/laptop. Over half of the participants have watched a 360° video on a phone (Figure 16), although all of them indicated doing so only a couple of times. Nearly half of the participants also watched 360° videos on a computer or laptop, with two participants indicating to watch them occasionally (Figure 17).

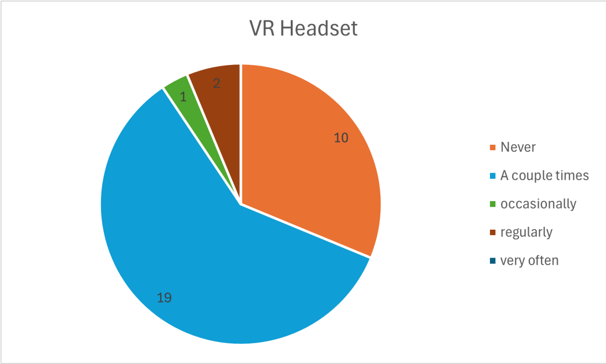


Figure 14: The distribution of familiarity with using a VR headset

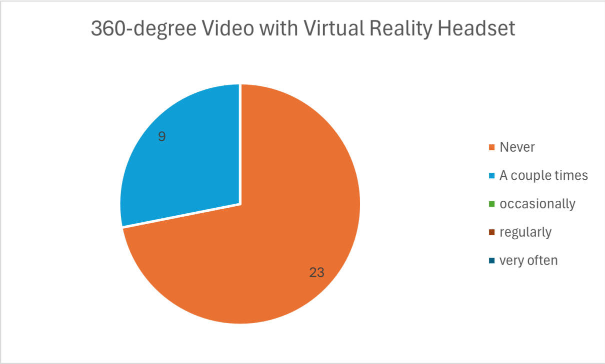


Figure 15: The distribution of familiarity with watching 360° videos with a VR headset

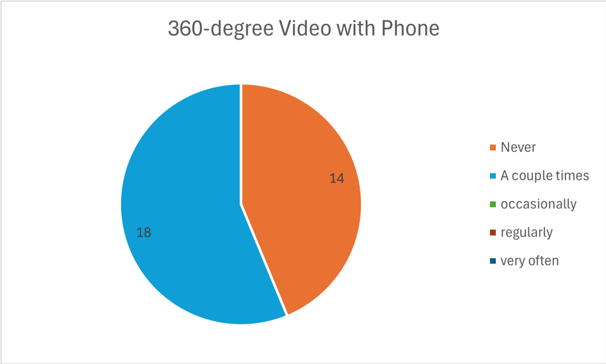


Figure 16: The distribution of familiarity with watching 360° videos with a Phone

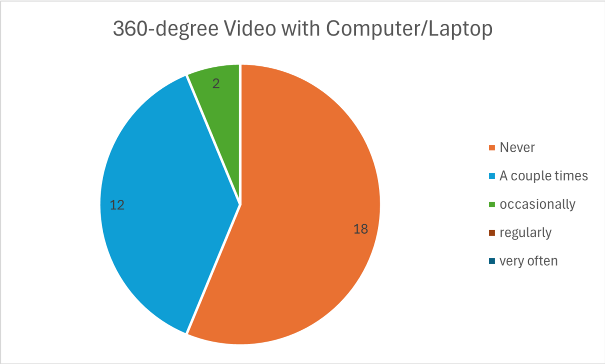


Figure 17: The distribution of familiarity with watching 360° videos with a computer or laptop

B.2 Total time spent

In section 6.1 we discuss the participants' performance in the search task. We utilize the inspection time per video as a measure since this best represents the time a participant needs to determine whether a video is correct for the search task. However, we also measured participants' total time to complete the search task. This measure is less precise, as variance might be introduced, for example when participants open a different number of videos between the interfaces. Nevertheless, exploring this measure is valuable since the interface with the fastest time for a few participants differs depending on which measure is used. However, the results come to the same conclusion when looking at averages.

The results are presented in table 7. The results of four participants are excluded because the total time for either of the interfaces was influenced. Three of the four participants experienced a crash, so the total time could not be reconstructed. The other participant thought he found a pattern in the presentation order of the thumbnails which turned out to be right in his specific case, causing him to immediately click the correct video for one of the interfaces.

	N	Mean	Std. Deviation	Minimum	Maximum
Filmstrip total time	28	03:30.00	02:57.46	00:14.70	13:20.17
Hierarchical total time	28	02:09.11	01:37.92	00:16.11	06:44.76

Table 7: The average time spent during the search task in mm:ss.ms

In line with the findings from section 6.1, the results indicate that, on average, participants are faster at finding the correct video with the hierarchical interface. A Shapiro-Wilk test was conducted, revealing a significant deviation from normality for both the filmstrip interface ($W= .22$, $p< .001$) and the hierarchical interface ($W= .19$, $p= .003$). The Wilcoxon Signed-Ranks test then demonstrated that the total search time in the filmstrip interface was statistically significantly longer than in the hierarchical interface ($Z= -2.824$, $p= .005$). Notably, the maximum search time in the filmstrip interface nearly doubled that of the hierarchical interface.

B.3 Time spend per layer

During the experiments, it was observed that most participants who did not find an effective search strategy spent a lot of time in the third layer. In section 6.3 we analyzed the data to see if this observation was consistent across all participants. Table 8 presents the data used for the analysis. It contains the time spent in each individual layer as well as the time spent per video for all the participants.

Participant	Time spend in layer 1	Time spend in layer 2	Time spend in layer 3	Time spend per video
Participant 1	0:00:51.90	0:01:19.81	0:01:30.09	0:00:31.969
Participant 2	0:00:55.47	0:01:34.07	0:00:33.02	0:00:24.332
Participant 3	0:01:16.82	0:02:15.83	0:01:27.22	0:00:37.063
Participant 4	0:03:44.34	0:03:04.96	0:01:12.70	0:01:01.236
Participant 5	0:00:01.50	0:00:42.28	0:00:17.95	0:01:08.660
Participant 6	0:01:00.35	0:01:16.67	0:00:08.09	0:00:22.610
Participant 7	0:01:08.53	0:00:00.00	0:00:00.00	0:00:17.910
Participant 8	0:00:15.39	0:00:27.32	0:00:01.96	0:00:12.597
Participant 9	0:00:54.47	0:00:03.13	0:00:00.00	0:00:19.877
Participant 10	0:02:04.06	0:00:00.00	0:00:00.00	0:00:16.594
Participant 11	0:01:08.75	0:00:17.52	0:00:00.00	0:00:18.405
Participant 12	0:00:55.61	0:02:18.44	0:02:18.68	0:00:42.030
Participant 13	0:00:26.92	0:00:06.35	0:00:03.25	0:00:18.920
Participant 14	0:00:28.61	0:00:04.08	0:00:00.00	0:00:40.500
Participant 15	0:01:09.08	0:00:08.76	0:00:00.00	0:00:20.852
Participant 16	0:00:03.62	0:00:00.00	0:00:00.00	0:00:06.480
Participant 17	0:00:07.67	0:00:27.51	0:00:14.16	0:00:14.750
Participant 18	0:01:40.39	0:03:35.33	0:01:09.26	0:00:54.329
Participant 19	0:01:09.60	0:00:08.03	0:00:03.46	0:00:11.063
Participant 20	0:01:49.10	0:00:29.54	0:00:04.07	0:00:18.680
Participant 21	0:00:16.14	0:00:00.00	0:00:00.00	0:00:11.630
Participant 22	0:00:05.64	0:00:18.33	0:00:00.00	0:00:09.473
Participant 23	0:01:17.09	0:02:20.36	0:03:33.08	0:00:56.387
Participant 24	0:00:11.99	0:00:17.11	0:00:11.78	0:00:34.195
Participant 25	0:00:07.93	0:01:22.20	0:00:00.00	0:00:15.792
Participant 26	0:00:26.10	0:01:05.79	0:00:37.03	0:00:16.592
Participant 27	0:02:17.75	0:00:53.23	0:00:20.07	0:00:27.379
Participant 28	0:00:12.17	0:00:17.15	0:00:00.00	0:00:08.785
Participant 29	0:01:24.88	0:00:02.89	0:00:00.00	0:00:18.328
Participant 30	0:00:33.39	0:02:18.67	0:00:31.46	0:00:26.916
Participant 31	0:01:14.85	0:00:07.88	0:00:00.00	0:00:12.270
Participant 32	0:00:35.70	0:03:21.70	0:01:16.70	0:00:40.862

Table 8: Time spent per layer for each participant

B.4 Additional results from the User Experience Questionnaire

This section provides detailed results from the UEQ-s covered in section 6.2. Table 9 present the means and standard deviations of each individual question. The first four questions relate to pragmatic quality, while the last four relate to hedonic quality. Values between -0.8 and 0.8 are considered a neutral evaluation, indicated by a yellow right-pointing arrow. Values above 0.8 are considered a positive evaluation, indicated with a green upward-pointing arrow. On average, neither interface received a negative evaluation for any of the questions. The hierarchical interface received positive evaluations for seven of the eight questions, whereas the filmstrip interface received a positive evaluation for only three. Comparing the two interfaces, we can see that the filmstrip interfaces only scored better on questions 2 and 4. Given the simplicity of the filmstrip interface compared to the hierarchical interface, it is not surprising that the filmstrip interface is rated as easier and slightly clearer.

UEQ-S question	Filmstrip		Hierarchical	
	Mean	**Std. deviation**	**Mean**	**Std. deviation**
1: Obstructive - Supportive	→ 0.4	1.6	↑ 1.5	1.4
2: Complicated - Easy	↑ 1.2	1.8	→ 0.5	1.6
3: Inefficient - Efficient	→ 0.3	1.7	↑ 1.3	1.6
4: Confusing - Clear	↑ 1.2	1.4	↑ 1.1	1.4
5: Boring - Exciting	→ 0.3	1.5	↑ 1.8	1.3
6: Not interesting - Interesting	↑ 0.9	1.5	↑ 2.0	1.2
7: Conventional - Inventive	→ 0.6	1.8	↑ 1.8	1.5
8: Usual - Leading Edge	→ 0.7	1.8	↑ 2.1	1.1

Table 9: Mean and Standard Deviation of each UEQ-S question for Filmstrip interface and Hierarchical interface

B.5 Advantages and disadvantages of the interfaces, as mentioned by participants

B.5.1 Filmstrip interface.

The following tables (Table 10 and Table 11) present statements regarding the filmstrip interface, along with the frequency at which each statement is mentioned in the questionnaire answers.

Advantages	Amount of mentions
This interface provides a **clear** overview of video content	3
This interface provides an **quick** overview of video content	7
This interface is **simple/easy** to navigate	7
This interface is **fun** to use	2
This interface shows the video in **chronological** order	1
This interface is **intuitive**	1
This interface feels more familiar than the hierarchical interface	1
This interface is simpler than the hierarchical interface	2
This interface felt faster than the hierarchical interface	2
With this interface I can't miss any frames	1

Table 10: Advantages of the filmstrip interface

Disadvantages	Amount of mentions
Scrolling the timeline was very sensitive, causing spheres to move very fast	8
Fast-moving spheres made me a little dizzy	3
It took some time to understand this interface	2
This interface was harder for the search task due to possibly missing frames	2

Table 11: Disadvantages of the filmstrip interface

B.5.2 Hierarchical interface.

The following tables (Table 12 and Table 13) present statements regarding the hierarchical interface, along with the frequency at which each statement is mentioned in the questionnaire answers.

Advantages	Amount of mentions
This interface provides a **clear** overview of video content	9
This interface provides a **quick** overview of video content	8
This interface was **simple/easy**	5
This interface shows playing previews instead of frames	3
This interface lets me change granularity	1
This interface felt faster than the filmstrip interface	5
This interface was easier to use than the filmstrip interface	3
This interface gave a better overview of the video than the filmstrip interface	5
This interface was more fun than the filmstrip interface	2

Table 12: Advantages of the hierarchical interface

Disadvantages	Amount of mentions
The interface was complex (showing much information at once)	5
It took some time to understand this interface	3
The interface could use a button to go back to the first layer	1

Table 13: Disadvantages of the hierarchical interface

C PRECEDING LITERATURE STUDY

C.1 360° videos

360° videos, also referred to as omnidirectional videos, enable users to dynamically change their viewing direction during playback, offering an immersive experience where users can explore their surroundings. 360° videos can be consumed from various platforms, including mobile devices, computers, and head-mounted displays (HMD). Producing and distributing these videos presents several challenges. These challenges are not expected to significantly impact the study, as the primary focus is browsing 360° video libraries. A concise overview of the production and distribution challenges will be provided below.

C.1.1 Production.

Creating a 360° video requires multiple cameras aimed at different but slightly overlapping directions. Captured footage from these cameras is integrated into a 360° video with a process commonly referred to as stitching. A simple stitching method involves minimizing the pixel-to-pixel dissimilarities, but several more advanced stitching methods have been explored for better results [39].

After capturing and stitching, the video gets transformed into a planar representation using projection. Various projection techniques have been proposed (e.g. pyramid maps, frustum maps, equal-area projection)[12]. Notably, the predominant choice for 360° videos is *equirectangular projection* [41]. Another popular projection technique, widely used in gaming applications due to its space-efficient nature, is *cubemap projection*

C.1.2 Streaming.

Adaptive streaming of both 360° and traditional videos is an active field of research. Due to the higher bandwidth requirements, streaming 360° videos is more challenging then streaming traditional videos. Yaqoob et al. [41] classifies streaming solutions for 360° videos into three categories: viewport-independent, viewport-dependent, and tile-based streaming. Viewport-independent streaming is similar to streaming traditional videos as the whole frame is streamed in equal quality. Viewport-dependent streaming reduces the transmitted bitrate by transmitting areas of the video frame that are outside the viewport in reduced quality. Adaptions of the original video are prepared server-side. The client fetches the right adaption depending on the estimated viewing orientation. For tile-based streaming, video frames are segmented into smaller chunks of which the quality can be adjusted [13]. Recent studies such as [10], [28], [27], use viewport prediction to determine the quality of the tiles.

C.2 Interaction

Currently, interaction with 360° videos often mirrors methods employed for interacting with conventional videos. Users currently interact with videos using one of three modalities, namely: Touch, Mouse and keyboard, and spatial interaction. The first two involve interaction on a 2D screen, while spatial interaction involves a Head Mounted Display (HMD). Browsing applications commonly project the 2D interface to the virtual environment of a HMD, without utilizing the unique advantages provided by a virtual environment.

C.2.1 Navigation and Browsing.

The primary type of interaction is navigation, denoting the interaction within a video during playback including pausing and resuming, altering playback speed, navigating along the timeline, and, in the context of 360° videos, exploring the spatial dimensions of the video. When video navigation is combined with other interaction features like video retrieval or video summarization we call it Browsing.

A lot of research has been done for enhanced video browsers, by adding simple features like changing playback speed, time jumping, and timeline of visual frames [24][26][7], or by adding more advanced features.

Barbieri et al. [3] used a colored timeline slider to display additional information about the current frame. Each frame is represented by a color based on 1) the dominant color in that frame, or 2) the volume of the audio track. A smoothing filter was then used to help the user differentiate between segments.

Moraveji [25] followed a similar approach. In this experiment, The timeline is annotated with unique colors, assigned to image features. Segments with colors too narrow for clickable interaction are removed.

Extending this work, Schoeffmann and Boeszoermenyi [30] introduced navigation summaries. Navigation summaries are visual representations of either low-level details or high level information about the video. This visualization is an extension or replacement of the traditional timeline. It behaves similarly, meaning that when a user clicks on a position the video playback continues from that point in time. Users have the ability to choose from multiple navigation summaries to display. This work progressed into the development of a tool featuring multiple navigation summaries and *region of interest* querying, elaborated in [31].

Other work introduced ZoomSlider, an interface for interaction with the timeline [15] [18]. Traditional timeline interfaces can be hard to navigate with a zoomed-out scale. However, with a zoomed-in scale, the user can not scroll across the whole video. ZoomSlider was introduced to solve this scaling problem. Instead of using a timeline widget, the user can scroll horizontally anywhere in the video to move in time. In addition to that, the user can scroll vertically to modify the scale of the timeline.

This idea was later improved by integrating speed-based navigation into the ZoomSlider to create an audio-visual video browser [14]. In this advanced version of ZoomSlider (AV-ZoomSlider), speed-based navigation was added to the borders of the video player. scrolling vertically across the border changed to playback speed of the video to a slow-motion or fast-forward speed. Later this method was also implemented for mobile devices [17] [16].

Various studies employed video frames to visualize a video as an overview of the content. [22] [8] [6]. These visualizations are commonly referred to as filmstrips or storyboards. Drucker et al. [9] implemented "SmartSkip", a filmstrip approach for skipping through videos using a remote controller. Comparative analysis with conventional time skipping and fast-forwarding was conducted. Interestingly, although the smart skip method performed worse on quantifiable metrics such as the number of clicks required, the distance from the desired frame, and the time to get to the desired frame, users mostly rated smart skip as the best interface.

Hürst and Darzentas [19] introduced a browsing interface with a hierarchical storyboard design. This interface consists of a grid of temporally ordered thumbnails to represent a video. The thumbnails are generated using a time-based brute-force approach. Users can then select a thumbnail as an anchor and change the granularity of the grid, resulting in a more detailed storyboard of a segment of the video.

C.3 Browsing 360° video libraries

Interfaces for browsing 360° video libraries are predominantly designed for 2D screens. When using a Head-Mounted Display (HMD), a projected version of the 2D interface is commonly shown. These interfaces typically present videos using a grid pattern with either equirectangular or custom thumbnails. Research efforts aimed at enhancing the browsing experience have mostly been directed at improving this 2D interface.

C.3.1 Enhanced video browsers.

The Video Browser Showdown (VBS)[29] is an event where teams try to solve content-based video retrieval tasks in a competitive setup. This event results in specialized video browsing systems with different methods to query and browse videos [4]. This is a great source for state-of-the-art video browsing interfaces, although a lot of effort goes towards query formulation.

Xu et al. [40] worked on browsing videos in 3D. They proposed three keyframe extraction methods and compared them to existing methods. They then created an innovative prototype for 3D visualization of those keyframes. The visualization consists of three windows; *KFView*, *CloseView*, and *VideoView*. *KFView* is the overview of a video. It presents the frames in the shape of a helix along the z-axis. To navigate, the user can adjust the zoom level, as well as scroll up and down. The user can then click a frame for closer inspection in *CloseView* and *VideoView*. *VideoView* is a conventional video player with some additional features like 3D manipulation, and *CloseView* shows a number of neighboring frames on a multiple-sided polygon prism.

C.3.2 Virtual Reality.

In the domain of Virtual reality, research on enhanced video browsers is relatively scarce. In 2021 the VBS was introduced to *Vitrivr VR*, a VR multimedia retrieval system [36]. Its user interface can be divided into three categories: query formulation, result exploration, and media item inspection. Queries accepted by the system can be concept, text, Boolean, and geo-spatial based. The results can be explored through a cylindrical, rotatable display, where the position of the media item on the display is based on a similarity score. Media items can then be selected to pop out the *media segment inspector*, which allows the user to inspect the video in detail. The media segment inspector can be placed anywhere in the virtual space and displays segments of the video as frames in a box that users can riffle through.

Another interface that was introduced at the VBS in 2021 is *EOLAS* [37]. *EOLAS* represents keyframes of a video by embedding it into latent vector space. They created an interface to explore these keyframes based on a spoken textual query. This interface shows clustered groups of keyframes based on their encoded features in a 3D space. Users can navigate this space and get an initial position based on similarity in feature space.

C.3.3 360 ° videos.

These enhanced video browsers use thumbnails to represent frames of videos. 360° video frames cannot be shown in a 2D frame without using some transformation. Conventionally, equirectangular projection is used which results in a stretched version of the content that fits in the 2D frame. Vermast and Hürst [38] investigated the use of different shapes to represent the thumbnail. In this study, it is found that for recognizing general high-level concepts, it is best to use conventional equirectangular thumbnails. However, for recognizing low-level details of the video, projecting the frame on a sphere resulted in better accuracy without needing more time to complete the task. The sphere additionally resulted in a better user experience.

Building on this work, Knoop [21] did a comparative study on three designs for a VR video browser: An abstract design based on movies, a design based on a video store, and a design based on a record store. The designs were measured on perceived and measured performance. Results show a worse performance on all measures for the record store design. The video store design scores better on pragmatic quality measures, while the abstract design scores better on hedonic quality measures; however, the perceived performance did not show a significant difference.

C.4 Metrics

Measuring the success of a system can be difficult since the definition of success depends on the goals of the system. When assessing the quality of a system, we can measure the user's performance when using the system. However, this does not cover the entire scope of what makes a system successful. Other aspects, such as the user's frustration, excitement, and the system's intuitiveness of the system are just as important, but less trivial to measure because of their subjective nature. To provide a subjective comparison of systems, validated measures have been developed.

C.4.1 Usability.

Kirakowski [20] provided a definition for usability as "the effectiveness, efficiency, and satisfaction with which specified users can achieve specified goals in a particular environment". To measure the usability of a system, the industry standard is to use the System Usability Scale (SUS) developed by Brooke [5]. It consists of ten carefully selected questions presented on a 5-point Likert scale. The scale yields a score within the range of 0 to 100 which can be used to compare (versions of) systems. Although the score should not be used to directly assess the value, it does give a good indication of the usability of a system. According to Aaron Bangor and Miller [1], systems with a score above 70 are at least passable, and systems with a score below 50 will almost certainly have usability issues.

In 2010, Finstad [11] developed The Usability Metric for User Experience (UMUX). This is a usability metric with only four Likert scale questions. Results show that this metric is reliable, valid, and a sensitive alternative for the SUS.

C.4.2 User experience.

User experience (UX) does not have a clear definition. It is a term used to describe how a user perceives a system. Measuring UX is commonly done using the User Experience Questionaire (UEQ) developed by Laugwitz et al. [23]. The UEQ contains 26 7-point Likert scale questions resulting in six scales: Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty. Schrepp et al. [34] later developed UEQ-S, a short version of the UEQ. It contains only 8 questions and results in three scales: Overall quality, Hedonic quality, and Pragmatic quality. An Excel sheet for simple conversions of the results, as well as a benchmark are provided on the website[7].

[7]https://www.ueq-online.org/

REFERENCES

[1] Philip T. Kortum Aaron Bangor and James T. Miller. 2008. An Empirical Evaluation of the System Usability Scale. *International Journal of Human–Computer Interaction* 24, 6 (2008), 574–594. https://doi.org/10.1080/10447310802205776 arXiv:https://doi.org/10.1080/10447310802205776

[2] Abir Al Hajri, Matthew Fong, Gregor Miller, and Sidney Fels. 2014. Fast Forward with your VCR: Visualizing Single-Video Viewing Statistics for Navigation and Sharing. *Proceedings - Graphics Interface.*

[3] M. Barbieri, G. Mekenkamp, M. Ceccarelli, and J. Nesvadba. 2001. The color browser: a content driven linear video browsing tool. In *IEEE International Conference on Multimedia and Expo, 2001. ICME 2001.* 627–630. https://doi.org/10.1109/ICME.2001.1237798

[4] Minh-Triet Tran Duc-Tien Dang-Nguyen Anita Min-Chun Hu Binh Huynh Thi Thanh Benoit Huet Björn Þór Jónsson, Cathal Gurrin. 2022. *MultiMedia Modeling.* Springer Cham. 487–585 pages. https://doi.org/10.1007/978-3-030-98355-0

[5] John Brooke. 1996. *SUS – a quick and dirty usability scale.* 189–194.

[6] Linjun Chang, Yichen Yang, and Xian-Sheng Hua. 2008. Smart video player. In *2008 IEEE International Conference on Multimedia and Expo.* 1605–1606. https://doi.org/10.1109/ICME.2008.4607760

[7] Kai-Yin Cheng, Sheng-Jie Luo, Bing-Yu Chen, and Hao-Hua Chu. 2009. SmartPlayer: User-Centric Video Fast-Forwarding. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09).* Association for Computing Machinery, New York, NY, USA, 789–798. https://doi.org/10.1145/1518701.1518823

[8] M.G. Christel, A.G. Hauptmann, A.S. Warmack, and S.A. Crosby. 1999. Adjustable filmstrips and skims as abstractions for a digital video library. In *Proceedings IEEE Forum on Research and Technology Advances in Digital Libraries.* 98–104. https://doi.org/10.1109/ADL.1999.777702

[9] Steven M. Drucker, Asta Glatzer, Steven De Mar, and Curtis Wong. 2002. SmartSkip: Consumer Level Browsing and Skipping of Digital Video Content. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Minneapolis, Minnesota, USA) *(CHI '02).* Association for Computing Machinery, New York, NY, USA, 219–226. https://doi.org/10.1145/503376.503416

[10] Ching-Ling Fan, Jean Lee, Wen-Chih Lo, Chun-Ying Huang, Kuan-Ta Chen, and Cheng-Hsin Hsu. 2017. Fixation Prediction for 360° Video Streaming in Head-Mounted Virtual Reality. In *Proceedings of the 27th Workshop on Network and Operating Systems Support for Digital Audio and Video* (Taipei, Taiwan) *(NOSSDAV'17).* Association for Computing Machinery, New York, NY, USA, 67–72. https://doi.org/10.1145/3083165.3083180

[11] Kraig Finstad. 2010. The Usability Metric for User Experience. *Interacting with Computers* 22, 5 (05 2010), 323–327. https://doi.org/10.1016/j.intcom.2010.04.004 arXiv:https://academic.oup.com/iwc/article-pdf/22/5/323/1992916/iwc22-0323.pdf

[12] Mario Graf, Christian Timmerer, and Christopher Mueller. 2017. Towards Bandwidth Efficient Adaptive Streaming of Omnidirectional Video over HTTP: Design, Implementation, and Evaluation. In *Proceedings of the 8th ACM on Multimedia Systems Conference* (Taipei, Taiwan) *(MMSys'17).* Association for Computing Machinery, New York, NY, USA, 261–271. https://doi.org/10.1145/3083187.3084016

[13] Jeroen Van der Hooft, Maria Torres Vega, Stefano Petrangeli, Tim Wauters, and Filip De Turck. 2019. Tile-based adaptive streaming for virtual reality video. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 15, 4 (2019), 1–24.

[14] Wolfgang Hürst. 2006. Interactive Audio-Visual Video Browsing. In *Proceedings of the 14th ACM International Conference on Multimedia* (Santa Barbara, CA, USA) *(MM '06).* Association for Computing Machinery, New York, NY, USA, 675–678. https://doi.org/10.1145/1180639.1180781

[15] W. Hurst, G. Gotz, and T. Lauer. 2004. New methods for visual information seeking through video browsing. In *Proceedings. Eighth International Conference on Information Visualisation, 2004. IV 2004.* 450–455. https://doi.org/10.1109/IV.2004.1320183

[16] Wolfgang Hürst, Georg Götz, and Martina Welte. 2007. Interactive Video Browsing on Mobile Devices. In *Proceedings of the 15th ACM International Conference on Multimedia* (Augsburg, Germany) *(MM '07).* Association for Computing Machinery, New York, NY, USA, 247–256. https://doi.org/10.1145/1291233.1291284

[17] Wolfgang Hürst, Georg Götz, and Martina Welte. 2007. A New Interface for Video Browsing on PDAs. In *Proceedings of the 9th International Conference on Human Computer Interaction with Mobile Devices and Services* (Singapore) *(MobileHCI '07).* Association for Computing Machinery, New York, NY, USA, 367–369. https://doi.org/10.1145/1377999.1378037

[18] W. Hurst and P. Jarvers. 2005. Interactive, dynamic video browsing with the zoomslider interface. In *2005 IEEE International Conference on Multimedia and Expo.* 4 pp.–. https://doi.org/10.1109/ICME.2005.1521484

[19] Wolfgang Hürst and Dimitrios Darzentas. 2012. HiStory - A hierarchical storyboard interface design for video browsing on mobile devices. *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia, MUM 2012*, 17:1–17:4. https://doi.org/10.1145/2406367.2406389

[20] J. Kirakowski. 1995. *Evaluating Usability of the Human-Computer Interface.* Springer Berlin Heidelberg, Berlin, Heidelberg, 21–32. https://doi.org/10.1007/978-3-642-85220-6_2

[21] Mike Knoop. 2023. Exploring Interactive Experiences to Browse Extensive 360° Video Libraries in Virtual Reality.

[22] Anita Komlodi and Laura Slaughter. 1998. Visual Video Browsing Interfaces Using Key Frames. In *CHI 98 Conference Summary on Human Factors in Computing Systems* (Los Angeles, California, USA) *(CHI '98).* Association for Computing Machinery, New York, NY, USA, 337–338. https://doi.org/10.1145/286498.286805

[23] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and Evaluation of a User Experience Questionnaire. *USAB 2008* 5298, 63–76. https://doi.org/10.1007/978-3-540-89350-9_6

[24] Francis C. Li, Anoop Gupta, Elizabeth Sanocki, Li-wei He, and Yong Rui. 2000. Browsing Digital Video. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (The Hague, The Netherlands) *(CHI '00).* Association for Computing Machinery, New York, NY, USA, 169–176. https://doi.org/10.1145/332040.332425

[25] Neema Moraveji. 2004. Improving Video Browsing with an Eye-Tracking Evaluation of Feature-Based Color Bars. In *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries* (Tuscon, AZ, USA) *(JCDL '04).* Association for Computing Machinery, New York, NY, USA, 49–50. https://doi.org/10.1145/996350.996363

[26] Kadir Peker and Ajay Divakaran. 2004. Adaptive fast playback-based video skimming using a compressed-domain visual complexity measure. 2055–2058. https://doi.org/10.1109/ICME.2004.1394669

[27] Feng Qian, Bo Han, Qingyang Xiao, and Vijay Gopalakrishnan. 2018. Flare: Practical Viewport-Adaptive 360-Degree Video Streaming for Mobile Devices. 99–114. https://doi.org/10.1145/3241539.3241565

[28] Yago Sanchez, Gurdeep Bhullar, Robert Skupin, Cornelius Hellge, and Thomas Schierl. 2019. Delay Impact on MPEG OMAF's tile-based viewport-dependent 360° video streaming. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* PP (02 2019), 1–1. https://doi.org/10.1109/JETCAS.2019.2899516

[29] Klaus Schoeffmann. 2019. Video Browser Showdown 2012-2019: A Review. In *2019 International Conference on Content-Based Multimedia Indexing (CBMI).* 1–4. https://doi.org/10.1109/CBMI.2019.8877397

[30] Klaus Schoeffmann and Laszlo Boeszoermenyi. 2009. Video Browsing Using Interactive Navigation Summaries. In *2009 Seventh International Workshop on Content-Based Multimedia Indexing.* 243–248. https://doi.org/10.1109/CBMI.2009.40

[31] Klaus Schoeffmann, Mario Taschwer, and Laszlo Boeszoermenyi. 2010. The Video Explorer: A Tool for Navigation and Searching within a Single Video Based on Fast Content Analysis. In *Proceedings of the First Annual ACM SIGMM Conference on Multimedia Systems* (Phoenix, Arizona, USA) *(MMSys '10).* Association for Computing Machinery, New York, NY, USA, 247–258. https://doi.org/10.1145/1730836.1730867

[32] Martin Schrepp. [n. d.]. Data analysis tool (ueq-s). https://www.ueq-online.org/Material/Short_UEQ_Data_Analysis_Tool.xlsx Accessed on 12-06-2024.

[33] Martin Schrepp. 2023. *User Experience Questionnaire Handbook.* https://www.ueq-online.org/Material/Handbook.pdf Version 11, Accessed on 12-06-2024.

[34] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2017. Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence* 4 (01 2017), 103. https://doi.org/10.9781/ijimai.2017.09.001

[35] Martin Schrepp, Jessica Kollmorgen, and Jörg Thomaschewski. 2023. A Comparison of SUS, UMUX-LITE, and UEQ-S. *J. User Exper.* 18, 2 (jun 2023), 86–104.

[36] Florian Spiess, Ralph Gasser, Silvan Heller, Mahnaz Parian-Scherb, Luca Rossetto, Loris Sauter, and Heiko Schuldt. 2022. Multi-Modal Video Retrieval Innbsp;Virtual Reality Withnbsp;vitrivr-VR. In *MultiMedia Modeling: 28th International Conference, MMM 2022, Phu Quoc, Vietnam, June 6–10, 2022, Proceedings, Part II* (Phu Quoc, Vietnam). Springer-Verlag, Berlin, Heidelberg, 499–504. https://doi.org/10.1007/978-3-030-98355-0_45

[37] Ly-Duyen Tran, Manh-Duy Nguyen, Thao-Nhu Nguyen, Graham Healy, Annalina Caputo, Binh T. Nguyen, and Cathal Gurrin. 2021. A VR Interface for Browsing Visual Spaces at VBS2021. In *MultiMedia Modeling: 27th International Conference, MMM 2021, Prague, Czech Republic, June 22–24, 2021, Proceedings, Part II* (Prague, Czech Republic). Springer-Verlag, Berlin, Heidelberg, 490–495. https://doi.org/10.1007/978-3-030-67835-7_50

[38] Alissa Vermast and Wolfgang Hürst. 2023. Introducing 3D thumbnails to access 360-degree videos in virtual reality. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (2023). https://doi.org/10.1109/tvcg.2023.3247462
[39] LYU Wei, Zhou Zhong, Chen Lang, and ZHOU Yi. 2019. A survey on image and video stitching. *Virtual Reality & Intelligent Hardware* 1, 1 (2019), 55–83.
[40] Qing Xu, Yu Liu, Xiu Li, Zhen Yang, Jie Wang, Mateu Sbert, and Riccardo Scopigno. 2014. Browsing and exploration of video sequences: A new scheme for key frame extraction and 3D visualization using entropy based Jensen divergence. *Information Sciences* 278 (09 2014), 736–756. https://doi.org/10.1016/j.ins.2014.03.088
[41] Abid Yaqoob, Ting Bi, and Gabriel-Miro Muntean. 2020. A survey on adaptive 360 video streaming: Solutions, challenges and opportunities. *IEEE Communications Surveys & Tutorials* 22, 4 (2020), 2801–2838.

List of Tables

List of Figures