Introducing 3D Thumbnails to Access 360-Degree Videos in Virtual Reality

Alissa Vermast and Wolfgang Hürst

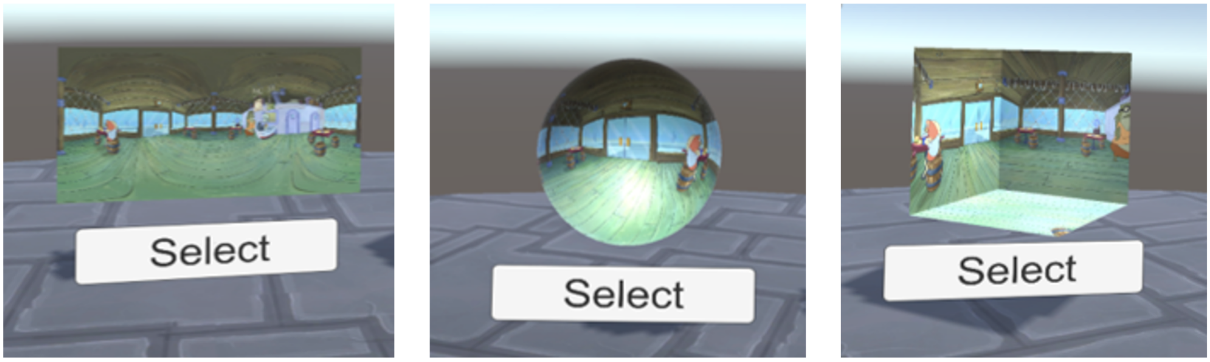

Fig. 1: Different thumbnails representing 360° videos in a 3D VR environment: a state-of-the-art 2D equirectangular projection (left), a 3D representation projecting the 360° video onto a sphere (center), and a cube-shaped projection (right). The screenshots are from an experiment comparing them with respect to performance and user experience when accessing and exploring a collection of 360° videos in VR.

Abstract—360° videos provide an immersive experience, especially when watched in virtual reality (VR). Yet, even though the video data is inherently three-dimensional, interfaces to access datasets of such videos in VR almost always use two-dimensional thumbnails shown in a grid on a flat or curved plane. We claim that using spherical and cube-shaped 3D thumbnails may provide a better user experience and be more effective at conveying the high-level subject matter of a video or when searching for a specific item in it. A comparative study against the most used existing representation, that is, 2D equirectangular projections, showed that the spherical 3D thumbnails did indeed provide the best user experience, whereas traditional 2D equirectangular projections still performed better for high-level classification tasks. Yet, they were outperformed by spherical thumbnails when participants had to search for details within the videos. Our results thus confirm a potential benefit of 3D thumbnail representations for 360-degree videos in VR, especially with respect to user experience and detailed content search and suggest a mixed interface design providing both options to the users. Supplemental materials about the user study and used data are available at https://osf.io/5vk49/.

Index Terms—360-degree video, video search, 360-degree video interaction, interfaces for video collections

1 Introduction

Due to the continuing development of virtual reality (VR) technology and the affordability of technologies like head-mounted displays (HMDs) for consumers, more and more people have the opportunity to enjoy 360° videos in VR [9]. This is often done with popular applications like YouTube VR or Facebook 360. Yet, the interaction design of these tools is mostly adapted from their two-dimensional counterparts (i.e., YouTube and Facebook Video). For example, like in 2D, interfaces to access and explore collections of 360° videos in VR generally consist of 2D grid-like structures with thumbnails and some meta-data such as the titles of the videos. Because these interfaces are all two-dimensional, a projection method is used to represent the 360° video in a flat way. An equirectangular projection is the most commonly used method in this context (see Fig. 1, left). However, this transfer from a 3D image to a 2D space comes at the price of high distortions in parts of the 2D representation. In addition, various guidelines for developing interfaces for VR advise favoring three-dimensional interfaces over two-dimensional ones, with the main argument being increased immersion[1]. We therefore suggest that a 3D representation of a 360° video may be beneficial when accessing individual videos in VR as well. Typical tasks people perform when browsing large video archives or searching for particular video content are, for example, to quickly spot videos of a certain type (e.g., underwater videos) or containing concrete objects (e.g., a sea turtle). While textual meta data, such as video titles, are often used for this, there are always situations where users will want or need to directly explore parts of the content, which is why basically all video search engines, both on 2D screens as well as immersive VR displays, always show a thumbnail-style representation of the videos.

We propose the use of three-dimensional thumbnails for previewing 360° videos where the 360° video is projected onto a 3D shape such as a sphere or a cube. We claim that such a 3D representation better resembles the original video and is therefore more intuitive to explore and easier to comprehend than the omnipresent equirectangular thumbnail visualization. Contents and spatial relations are easier to identify due to less distortions in the projection. Potential disadvantages include the fact that only parts of the projection is visible at a time, and users need to actively change their point of view or rotate the 3D representation to see what is on the back of it. Yet, this obstacle might even turn out to be beneficial, because such an interaction can lead to a better

• *Both authors are with Utrecht University. E-mail: huerst@uu.nl*

Manuscript received 14 October 2022; revised 13 January 2023; accepted 30 January 2023. Date of publication 24 February 2023; date of current version 29 March 2023. Digital Object Identifier no. 10.1109/TVCG.2023.3247462

[1] Oculus Best Practices, https://static.oculus.com/documentation/pdfs/intro-vr/latest/bp.pdf

Fig. 2: 360° image of a bedroom, in the equirectangular format. (Image source: https://pixexid.com/image/cm1btua-360-image-of-a-room, Creative Commons License)

Fig. 3: 360° image of a bedroom, in a fbrlud (front, back, right, left, up, down) cube-format

spatial understanding of locations of objects shown in the 360° video and be more enjoyable and thus preferred by users. In this paper, we present a comparative study verifying these claims and exploring if 3D thumbnails can provide a better alternative to the 2D representations currently used by almost all 360° video libraries in their VR interfaces.

2 Background and Related Work

Even though 360° media has existed for quite some time, the technology is still considered novel with plenty of unanswered research questions. Many of these address performance-related aspects about better hardware and/or optimizations, for example, in relation to streaming [9] and higher-quality imaging [19].

Research about usage of and interaction with 360° videos is relatively sparse and only recently gained larger attention. We can classify these works into two categories: in-video and outside-video navigation. The first concerns interaction methods within a single video. Examples for related research include work on presence [2], immersion [34], different interaction modes [27], new usage scenarios [25] and interfaces for editing of 360° videos [26]. An increasing amount of work is focusing on how to steer people's attention or how to enable them to interactively explore different viewing directions [2,12,15,21,23,35]. Some of these offer solutions that could be applied to our scenario, that is, gaining a quick overview of a video's type and content. For example, [35] uses an equirectangular projection thumbnail of the whole scene to enable people of seeing what is "behind them" when exploring a 360° video in VR. [33] uses equirectangular projections to aid navigation in and editing of 360° videos of social conversations. Yet, neither of these examples explore the potential of 3D visualizations as video representations. Even Nguyen et al's system still uses equirectangular projections to aid navigation within the videos (see Fig. 3 in [26]) despite using the high distortion of such projections as motivation for their "in-headset" video editing approach. The same holds for interaction and navigation approaches for within 360° video navigation on desktops, such as [20]. Likewise, most of these works focus on detailed in-video navigation, whereas our solutions address high-level video classification tasks.

Outside-video navigation concerns itself with how to browse a collection of videos. For traditional videos, most standard interfaces use a grid-like representation of individual videos, which in turn are represented by a small image or frame from the video and commonly referred to as video thumbnail. For 360° videos, projection methods are mostly used to visualize all viewing directions via such a flat, 2D image. Out of the wide variety of projection methods (see, e.g., [36]), equirectangular and cube map projections have the most widespread usage in literature but also commercial applications [29]. An example of an equirectangular projection can be seen in Figure 2. This visualization is used for thumbnails by both YouTube and Facebook, as well as applications like YouTube VR and Facebook 360. Furthermore, it is commonly used by video editing software, such as Adobe Premiere Pro, for editing immersive videos. In addition to the widespread usage, another advantage of this projection is the logical mathematical relationship between the position of a pixel and its corresponding location in the spherical projection. However, this does not necessarily mean that someone less knowledgeable about this considers this relationship as intuitive. A major disadvantage of the equirectangular projection is that pixels at the bottom and the top of the spherical video have to be stretched, and therefore distort part of the image [28,29]. This inherent need for distortion can be explained by the Theorema Egregium, proven by Carl Gauss [10]. This theorem states that any curved surface cannot be displayed on a single plane without distortion or stretching, caused by the difference in the size of the circumference of a circle at the top and bottom compared to in the middle of the sphere. Therefore, the projection could pose problems regarding users not understanding distance and angle related information [18]. The high amount of distortion can also pose problems for existing compression software [4,28], amplifying the problem of performance and optimization.

An occasionally used alternative to the equirectangular projection is the cubemap projection introduced by Greene [11] and variations thereof [4,13]. Here, the views of six perpendicular viewing directions are visualized as squares. An example of a cubemap projection with all sides in one row can be seen in Figure 3. An obvious disadvantage is that it is difficult to associate connections across the individual squares, no matter in which order they are presented. For this reason, they are sometimes shown with left-front-right-back views in one row and the square representing the top and bottom are placed above and below, respectively. Yet, this representation is not very space efficient. In addition, it remains difficult to keep track of moving objects and the exact spatial relation between objects [18]. Compared to equirectangular projections, the amount of distortion is significantly reduced but still present, mainly at the corners and edges of the cube.

Using miniature 3D visualization for interaction with larger 3D data is a well-know approach in VR, most prominently represented by the World-in-Miniature (WiM) approach [30]. While our approach follows the same principle, it is different with respect to the actual visualization (3D shapes such as cubes and spheres) and intended usage. Traditional WiM approaches in VR are commonly used to manipulate and interact with the virtual world. Elvins et al. use them to assist people in navigation [5–7]. Englmeier et al. present a newer example using a sphere visualization [8], but again with a use case (discrete locomotion) that is very different than identification and classification of content, which is the focus of our work.

People have also experimented with spherical representations of 360° content independent of VR and 360° video. For example, Li et al. [22] and Miyafuji et al. [24] studied the usefulness of physical spherical displays. Their results support our claim that a spherical 3D representation of a 360° video could be useful for our use case scenario as well. There are several examples for the usage of spheres in other contexts, such as navigation [3], aids for scene transitioning [16], and scene representation for remote collaboration [31]. The closest usage of spheres for the visualization of 360° content to our work, both in terms of the actual visualization as well as its usage, is the integration of Street View into Google Earth VR [17]. However, their context differs from ours insofar as they focus on still 360° images that are represented in context of the environment, which should simplify content identification. Although we are not aware of any related scientific study of their approach, their implementation supports our claim that such a visualization may be beneficial in the context of 360° video search and classification as well.

3 Idea, Research Questions, and Approach

In this section, we introduce our solution (Section 3.1), define related research questions and our approach to answer them (Section 3.2), and describe the implementation for the experiment (Sections 3.3 to 3.5).

3.1 3D thumbnails

Immersive 360° videos are commonly rendered as if the viewer is located in the center of a sphere. The video is then projected on the inside of that sphere. We therefore claim that representing a video as a miniaturized 3D sphere that a user can explore from the outside, such as illustrated in Figure 1, center, may be more intuitive than using the flat, distorted 2D projection shown in Figure 1, left. Alternatively, a cube-shaped representation, as shown in Figure 1, right, could be used, which might provide a better spatial orientation due to its more discrete character making it easier to associate with different viewing directions, despite featuring more distortions of the actual content than a spherical representation. We purposely restricted this initial study to these basic shapes in order to establish the validity of the concept of 3D thumbnails for the representation of 360° videos. If proven successful, further, more complex shapes could be explored in future follow-up work.

3.2 Research questions and approach

To evaluate the potential benefits of these two visualizations of 3D thumbnails (sphere and cube) we present an experiment, in which they will be compared against the most commonly used 2D thumbnail representation of 360° videos, that is, an equirectangular projection. We assume a scenario in which multiple 360° videos are represented in VR and the user wants to explore them with different intentions. For example:

1. Users might be looking for videos that contain a specific element or object of interest (e.g., sea turtles in an underwater video). In such a use case, it can be important that a user is able to explore a video in all viewing directions, because the target might appear anywhere in the 360° video.
2. Users might want to browse 360° videos freely in search for a certain type of video (e.g., underwater videos). When looking at a collection of video thumbnails, the user wants to quickly identify which files are relevant for them, and which ones are not. In such a use case, it might be sufficient to only see parts of the video and only some viewing directions to find the relevant ones.
3. Users might not have a concrete search goal but just want to find random videos of interest or explore the whole database to get a better idea of its content. In this case, detailed information about the content might be less relevant, but the browsing experience, that is, how much a user enjoys using the system and feels comfortable exploring it, plays an important role.

Interfaces for the access of video archives usually do not just represent the actual video content, for example, with a thumbnail, but also other information, such as the video's title and maybe a short text description. Yet, all use cases illustrated above often rely heavily on visual exploration. For example, it might be easier and faster to spot an underwater video by looking at a thumbnail rather than reading a video's title or textual description (use case 1). Likewise, visual content is generally preferred for random exploration (use case 3). Because we are interested in finding the best visual representation, we therefore excluded additional information about a video's content from the evaluation but solely presented the related thumbnails. For each of the use cases, we introduce a related research question:

RQ1: How fast can users identify whether a video contains a certain item of interest depending on the thumbnail version?

RQ2: How fast can users identify the high-level subject of a video based on its thumbnail representation?

RQ3: How do users rate the user experience with different thumbnails?

In the experiment, we use two tasks to address the first two research questions. These tasks simulate the related search behavior by presenting a user with a collection of video thumbnails and a selection prompt. This prompt instructs users what search criteria they should employ when selecting videos from the video collection. The third research question is addressed by verifying the experience of the users when performing these two tasks. The concrete tasks and exact measures that are used in this context are detailed in the methodology in Section 4.

We expect that the user experience results will identify 3D thumbnails to be an exciting and engaging concept because they better represent the 3D nature of 360° video and foster a more interactive experience, with better usability.

Furthermore, we suspect that participants will be able to identify the high-level subject of a video just as well or faster when using 3D thumbnails because we expected that participants do not need the full 360° video to correctly identify its subject. Thus the fact that parts of a 3D thumbnail are always hidden should not have a major impact. If anything, we expect that the lesser degree of distortion in the 3D thumbnails may decrease the time needed to identify the video's subject.

Finally, we hypothesize that participants will be slower at finding certain objects in the video with 3D thumbnails compared to 2D ones, because rotating the view of the thumbnail requires active interaction. Depending on the interaction method, the time difference might be minimized though.

3.3 Implementation: thumbnail shape

360° videos watched with head-mounted displays are supposed to create the experience as if the viewer is in the center of the scene depicted by the video. For this reason, most 360° video players are implemented by projecting the video onto the inside of a sphere and placing the viewer in the center of it. Likewise, the video can be projected onto another "enclosing" 3D shape, of which a cube is the most basic one. The idea with the 3D thumbnails is to place the viewer outside of this 3D shape to give them a better, but proportionally more accurate overview of the scene – in contrast to the geometrically inaccurate and highly distorted view of a equirectangular 2D projection. However, simply displaying the video on the outside of the 3D shape with the observer next to it would inherently change the image by creating a "mirrored" version of it. This unwanted effect was avoided by having the normals of the shape's faces be pointing inwards so that the video plays on the inside of the shape.

3.4 Implementation: thumbnail previews

Video is a time-dependent medium, which is why a still thumbnail only represents a very small fraction of its content. In video archives such as YouTube, it is therefore common to play a small snippet of its content once a viewer hovers over a video's thumbnail with the mouse pointer. In VR, a similar implementation is often used for 360° videos, that is, the still image of the equirectangular projection is replaced by a short video clip play a small snippet of its content once a viewer hovers over the thumbnail with a controller. We take a similar approach for 3D thumbnails. That is, the 3D thumbnails will start off to be static (a single frame of the video). However, when the users hovers the ray-caster of the VR controller over the thumbnail or grabs the thumbnail to manipulate the orientation of it, a 30 second preview of the video will play until the controller is moved away or the thumbnail is released again. If the user holds or hovers over a thumbnail for more than 30 seconds, the video will freeze on the last frame of the preview. If the users stops holding or hovering and then grabs or hovers over the thumbnail again, the preview will start playing from the beginning again. Some state-of-the-art applications include previews that are longer. We keep the previews relatively short as to minimize experiment duration as well as application memory needed for the videos. The previews are 30 seconds of a single uninterrupted scene, meaning there are no jump-cuts in the video preview. This was done because the focus of our research is on simple classification tasks common to the initial selection of videos from a larger set and not on the more sophisticated, detailed exploration of a single video's content. All the videos in the application are muted, because our aim is to evaluate the best visual representation, independent of audio feedback.

3.5 Implementation: Interaction design

To interact with the application, we use three buttons on the headset's controllers: "Index Trigger", "Hand Trigger", and the "A" button (or

Fig. 4: The right controller of the Oculus Quest used in the experiment. The left one features an 'X'-Button instead of an 'A'-Button but is otherwise identical.

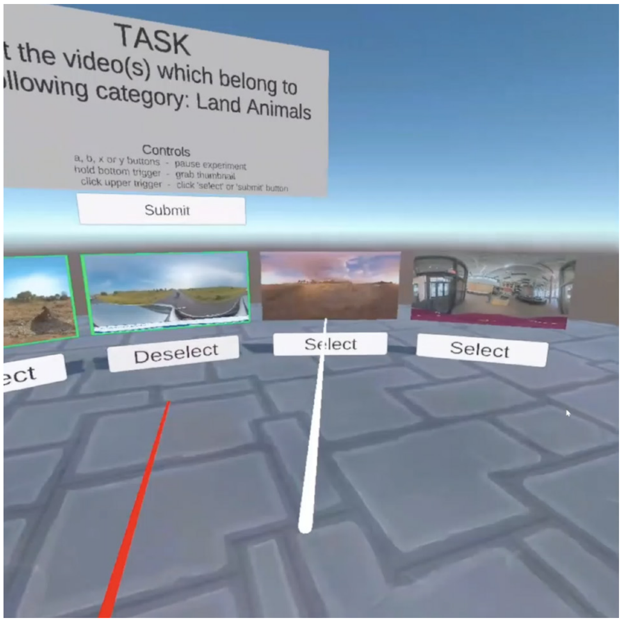

Fig. 5: Flat, equirectengular thumbnails (screenshot from experiment).

"X" button on the controller for the left hand – see Figure 4).

To explore the 3D thumbnails, the implemented environment needs to enable users to interact with them, for example, by picking them up or hovering over them (see the previous subsection). In addition, certain standard interactions need to take place, for example, confirming that target videos are found by pressing a related button. For this, we use ray casting, which is one of the most common approaches for interaction in VR. That is, raycasts are used to point and interact with elements in the application. When aiming the raycast of a controller at a thumbnail or a UI element such as a button, the ray will change color (from red to white) indicating the possibility for interaction. When aiming at a button, the user can use the Index Trigger to press the button. This aligns with VR interaction conventions. The Hand Trigger is used to pick up a thumbnail. Picking up a thumbnail works as follows. The user aims the raycast at the desired thumbnail. The ray changes to white. The user presses and holds down the Hand Trigger. The thumbnail will then teleport and snap to the position of the controller. The thumbnail will also be scaled by 0.5 to compensate for the decreased distance to the thumbnail.

When the user moves their hand or wrist, they can rotate the thumbnail as desired in order to explore different viewing directions of the video. When letting go of the Hand Trigger, the thumbnail will teleport back to its original position, while being scaled back up by 2. In the case of a 3D thumbnail, the orientation of the thumbnail will be the orientation the thumbnail had when the user let go of the thumbnail. Not only is this arguably more intuitive, it also allows the user to pick up the same thumbnail again and rotate it some more, in case the user does not have enough wrist flexibility to do the rotation permutation in one go. Since the participant has two controllers it is possible to hold two different thumbnails at the same time, one using each hand.

The two dimensional thumbnail, that is, the flat equirectangular projection, behaves similarly with the exception of the rotation, which will swap back to a position perpendicular to the user's view. We consider this difference in interaction between the different thumbnails inherent to the type of thumbnail, as it would not make sense to place the two-dimensional thumbnail rotated in any other way. This is also not a feature that is present in state-of-the-art applications. While standard players using 2D thumbnails normally do not allow users to pick them up for closer inspection, this feature is supported by our application because it is consistent with the implementations of the 3D thumbnails. In addition, it might provide a benefit by enabling users to explore the 2D thumbnails' content at a closer distance.

4 Methodology

In the following (Section 4.1), we specify the measures used in the experiment to answer the research questions introduced in Section 3. We describe the experiment design (Section 4.2) and provide details about the participants (Section 4.3) along with the used data and implementation of the text environment (Section 4.4).

4.1 Experiment goal and measurements

The experiment was designed to compare the different kinds of thumbnails (sphere, cube, equirectangular), which was therefore the independent variable of this study. Dependent variables are the data that is gathered by questionnaires (RQ3), the time it takes the participants to complete a task (RQ1, RQ2), and whether or not they completed the task successfully (RQ1, RQ2).

The participants had to fulfill certain tasks in VR, simulating the search goals discussed in Section 3. These tasks are divided into two different groups: the *categorization tasks* and the *item search tasks*. Both types of task have a similar setup. The participant is presented with five thumbnails of the same kind (either equirectangular, sphere or cube) and a task they have to fulfill. In the case of a categorization task, the participant is asked to select the video(s) that belong to a certain high-level category. The possible categories are city, indoors, roller coaster, land animals, underwater and winter sports. The measurements that are taken during the categorization tasks answer RQ2. In case of an item search task, the participant is asked to select the video(s) that contain a certain item. The measurements that are taken during the item search tasks answer RQ1. Two different aspects were measured: the time taken to complete the task and the correctness/error rate. With this data, we can infer how fast and how correct participants are able to recognize the imagery of the thumbnails.

To answer RQ3, the participants filled in the UEQ-S questionnaire each time after experiencing the categorization tasks and the item search tasks for one thumbnail type. The use of the UEQ-S is chosen over the use of the commonly used SUS because the SUS is more useful for a fully functional system, while the thumbnails are part of what could be a fully functional video browsing system. The experiment also included questions about cybersickness, about what participants considered to be advantages and disadvantages of each thumbnail type, and a question asking the participant to rate the thumbnail types from best to worst based on their experience with them.

Some additional interaction data was logged as well, which could possibly provide more insight into why certain tasks might have been performed better or worse and why they took more or less time to complete. This includes whether the participant picked up thumbnails to inspect them more closely and possibly rotate the view. Furthermore, since the participant has the possibility to pause the experiment, we measured the amount of times the participants took breaks as well.

Screenshots from the experiment environment are shown in Figure 5 for the equirectangular thumbnails and Figure 6 for the sphere thumb-

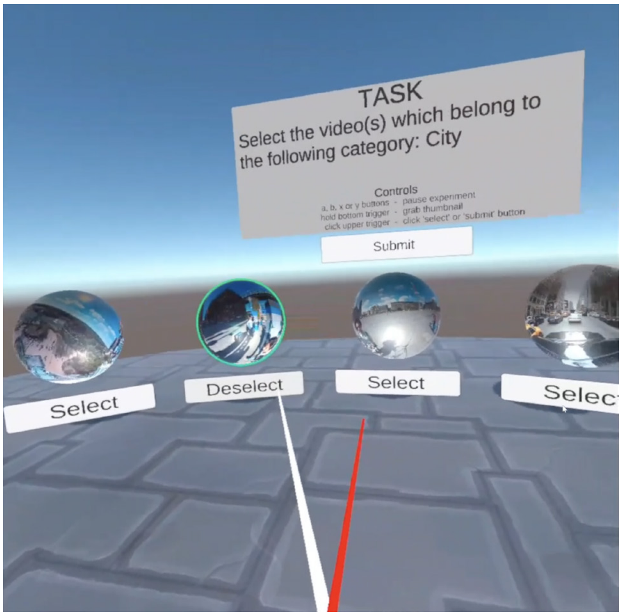

Fig. 6: Sphere thumbnails (screenshot from experiment).

Fig. 7: Closer inspection of sphere thumbnail (screenshot from experiment).

nails. For each task, there is a row of five thumbnails of which the participant needs to chose one (item search task) or several (categorization task). The screenshot in Figure 7 illustrates a situation where a participant picked up a thumbnail for further inspection.

4.2 Experiment design

Procedure. Before starting the experiment, the participant received an explanation about the procedure and was asked to sign an informed consent form. The experiment took about 30 minutes per participant. They were seated on a non-rotating chair to minimize cybersickness and allowed to take breaks if needed. The VR headset (Oculus Quest) was cleaned thoroughly after each participant used it. The experiment consisted of the following three steps.

1. *Questions before VR*. This section includes questions about general participant characteristics like age and gender, as well as some questions about the participants' experience with VR and 360° videos. These questions were answered on a laptop computer by filling out an online form.
2. *The VR portion*. This is the part of the experiment in which the participant wore the VR headset. The participant was aided in putting on the VR headset. Then, the participant was presented with a welcoming explanation scene, followed by a tutorial. This tutorial was in place to make sure the participants got familiar with 360° video and the interaction with those videos, as well as explaining what will be asked of them in the experiment. The step-by-step introduction of the tutorial guarantees that test subjects who are less familiar with VR in general or unfamiliar with 360° video understand and are able to operate the environment correctly. The tutorial is followed by an example scene, with an example task. This was included so that the participant is fully prepared for the tasks that belong to the experiment. Then, the following steps are repeated for each of the three thumbnail kinds (sphere, cube and equirectangular). First, the participant gets two categorization tasks, then two item search tasks and then one scene with the UEQ-S questionnaire.
3. *Concluding questions after VR*. These questions focused on comparing the three different thumbnails and their corresponding advantages and disadvantages. It also asked the participant to rate the thumbnails from best to worst. The participant enters their answer on the laptop computer again.

The order in which the participants experienced the different types of thumbnails was balanced across participants.The type of tasks and videos the participant will encounter was determined as follows. First, the tasks for the item search tasks are chosen. We specified eighteen different tasks (three for each video category) The tasks are made such that the item to be searched is as well-known as possible (not requiring any specific knowledge). The aim was to minimize possible different interpretations for the tasks. The amount of correct videos to be selected (at least one, maximum all five) has been distributed over the tasks as evenly as possible. Target items were randomly distributed over the horizontal and vertical axis of the video to avoid any search bias, with the majority of targets being slightly above the horizon level (e.g., people, cars, street lights), few above (e.g., moon, ceiling fan), and very few below (e.g., sharks in underwater videos). We therefore do not treat the placement of the item to be searched as a variable of influence, as the positions of the item search target are objectively placed over all conditions. Since the participant sees six item search tasks in total (two for each thumbnail type), we select one of the three possible tasks for each category at random (thus making sure the participant gets an item search tasks for each of the six categories). The order of these tasks is randomized. The participant also sees six tasks for the categorization task. The order of these is randomized as well. These categorization tasks are made with the videos that were not used by the item search tasks (ensuring a participant never sees the same video thumbnail multiple times during the experiment). The correct amount of videos to select (again at least one, maximum all five) is randomized.

A within subject design is chosen for better comparability between the use of different thumbnails and because it requires less test subjects, which were hard to accommodate due to the ongoing COVID situation.

Setup and physical environment. The experiments were conducted in person under consideration of all COVID-regulations imposed during that time to allow for a controlled environment and the possibility to assist the participants when necessary. The experiments were always conducted in a neutral environment (minimizing distraction, background sounds, differences in temperatures etc.). Furthermore, this in person setup allowed for a better opportunity to gather qualitative data by observing the participant and conducting a semi-structured discussion with the participant after the experiment.

4.3 Participants

A total of 31 participants (11 female, 20 male) participated in this study. All participants were students between 18 and 30 years old, with most of them being in their 20s (6.5% 18-20 years, 54.8% 21-25 years, 35.5% 26-30 years, 3.2% 31-35 years). This age range is often considered as early adopters of new technologies and is therefore also the target group of this research. Almost all participants had used VR at least once, but the majority (67.7%) had only used a VR headset a couple of times, only 3.2% never at all. The participants were quite unfamiliar with 360° videos. 51.6% of them had never watched 360° videos on a laptop or desktop computer, with the remaining participants only having done so a couple of times. The familiarity with watching 360° videos on a mobile phone was higher. 32.3% had never watched 360° videos on a mobile phone, 64.5% had done so a couple of times, and 3.2% (one participant) had done so quite a bit. Participants were most unfamiliar with watching 360° videos in VR. 61.3% had never done this before, with the remaining participants saying they had only done this a couple of times. From this data we can conjecture that while all participants are somewhat familiar with VR, the majority of participants is quite unfamiliar with technologies regarding 360° video.

4.4 Material and implementation

Test environment. A testing environment with the 2D and 3D thumbnails was implemented in Unity version 2021.16f1[2]. Several packages were needed to ensure interaction with a VR HDM worked properly. The packages used for this purpose are Oculus XR Plugin, XR Interaction Toolkit and XR Plugin Management. An Oculus Quest[3] VR headset (first generation) was used in the experiment.

Tools for analysis. To analyze the results, Python 3.8[4] was used to transform the data and carry out calculations. In addition, the Short UEQ Data Analysis Excel tool [5] was used to aid the analysis of the UEQ-S results.

Data (videos). The 90 videos (fifteen for each of the six categories) that were used for the thumbnails in the experiment were gathered from YouTube. YouTube supports a multitude of different 360° video formats. However, some are more easily convertible into other formats than others. For this research, only videos that were in equirectangular format or in YouTube's own cubemap format were used. The videos were all downloaded in 1080p quality. A resolution of 4K is generally considered to be the bare minimum for watching 360° videos in VR [14], but this is not necessary here, since the video will much smaller relative to the user in thumbnail form, compared to playing the video all around the user when watching a single 360° video normally. FFmpeg[6] [32] was used for preprocessing the video data. If the video that was downloaded from YouTube was in YouTube's own cub emap format, they were first converted into the equirectangular video format. All the videos were cut down to a preview of 30 seconds of continuous video (meaning without any cuts) of one scene only. This was done so that the thumbnail portraits the scene in the video accurately. The videos are muted, as the different tasks the participant has to fulfill focus on visual search tasks only. After this, the videos were converted into a custom cubemap format, that is needed to make the cube thumbnails, see Figure 3. This format is then UV-mapped onto the inverted cube to make the thumbnails. Thus, for each video, both an equirectangular as well as a cubemap version is saved.

5 Results

In this section, we summarize our results, starting with the quantitative measures (Sections 5.1 to 5.5), followed by qualitative results (Sections 5.6 and 5.7).

5.1 Time needed for the tasks

To inspect whether the thumbnail kind has an influence on the amount of time that it takes to complete the selection task, we measured the time it took each participant to complete each task. Every participant

	Time	
	Categorization	Item search
Flat	49.03 s (22.60)	56.90 s (19.37)
Sphere	65.81 s (22.98)	58.06 s (21.88)
Cube	61.32 s (21.97)	51.35 s (24.20)
	Correctness	
	Categorization	Item search
Flat	9.65 (1.15)	7.77 (1.72)
Sphere	9.94 (0.25)	9.87 (0.34)
Cube	9.81 (0.47)	7.39 (1.38)
	Pickups	
	Categorization	Item search
Flat	6,00 (3.61)	10.77 (2.34)
Sphere	8.42 (4.05)	9.81 (2.61)
Cube	7.55 (4.43)	10.58 (3.65)

Table 1: Overview of the results from measuring the time needed for the tasks (in seconds), the correctness (the amount of correctly selected or not selected thumbnails out of 10) and pickups (the amount of times a participant picked up a thumbnail). The table includes the means and the standard deviations of these values, per thumbnail type.

has done two categorization tasks and two item search tasks for each thumbnail. We conducted our statistical analysis on the time that a participant needed for two of the same kinds of tasks. To examine the possible connection between the thumbnail kind and the time needed for the task, we conducted a one-way repeated measures ANOVA on all 31 participants for the categorization task, as well as for the search item task. The results can be seen in Table 1 (top).

For the categorization task, the results show that the thumbnail type leads to statistically significant differences in time needed for the task ($F(2,60) = 3.7849, p = 0.03$). To investigate this significant difference further, we conducted post hoc analysis using Tukey's test. This shows significant difference between the flat and sphere thumbnails ($p = 0.01$), but no significant difference between the flat and the cube thumbnails ($p = 0.09$) and between the cube and the sphere thumbnails ($p = 0.72$).

For the item search task, the results of the one-way repeated measures ANOVA show no statistically significant differences in the amount of time in seconds needed to complete the task due to the different thumbnail type ($F(2,60) = 0.8841, p = 0.42$

5.2 Correctness of the tasks

To examine whether there is a correlation between the thumbnail type and the correctness in the selections the participants made, we conducted a one-way repeated measures ANOVA on all 31 participants for the categorization task, as well as for the search item task. This was done by comparing the selections the participants made with the correct selections for each task. The results can be seen in Table 1 (center).

For the categorization task, we found no statistically significant differences in the correctness of the the tasks ($F(2,60) = 1.1131, p = 0.36$). For the item search task, the results of the one-way repeated measures ANOVA show a statistically significant difference in the correctness of the the tasks between the thumbnail types ($F(2,36) = 22.5527, p < 0.0001$). ($F(2,60) = 33.4640, p < 0.01$). To evaluate this further we conduct a post hoc analysis between the different combinations of thumbnail types using Tukey's test. This test results in a statistically significant difference in correctness between the flat and the sphere thumbnails ($p < 0.01$), statistically significant difference between the sphere and the cube thumbnails ($p < 0.01$), but no statistically significant difference between the flat and cube thumbnails ($p = 0.48$).

5.3 Thumbnail pickups

To examine whether the thumbnail kind has an influence on the amount of times participants pick up the thumbnails, we conducted a one-way repeated measures ANOVA on all 31 participants. This is done for the

[2]https://unity.com/
[3]https://www.oculus.com/
[4]https://www.python.org/downloads/release/python-380/
[5]https://www.ueq-online.org/
[6]https://ffmpeg.org/

	Flat		Sphere		Cube	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Pragmatic Quality	0,895	1,261	1,484	0,959	0,581	1,277
Hedonic Quality	-0,798	1,040	1,661	0,768	1,202	0,843
Overall	0,048	0,872	1,573	0,670	0,891	0,854

Table 2: The two quality scales averaged over the UEQ-S answers of all participants per thumbnail type.

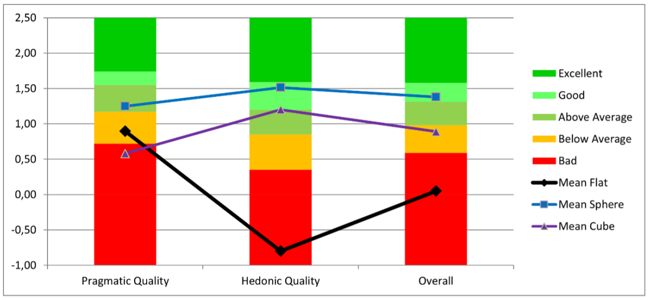

Fig. 8: Overview of the results per thumbnail type, compared to a benchmark dataset.

pickups during the categorization task and for the pickups during the item search task. Results can be seen in Table 1 (bottom).

For the categorization task, it shows that the thumbnail type leads to statistically significant differences in amount of thumbnail pickups ($F(2,36) = 5.2791, p < 0.01$). ($F(2,60) = 6.2964, p = 0.03$). To investigate this significant difference further, we conduct a post hoc Tukey's test. However, this test show that there are no pair-wise comparisons that are statistically significant. The results between the pickups for flat and sphere thumbnails fall short of being statistically significant ($p = 0.06$), there is no statistically significant difference between the sphere and the cube thumbnails ($p = 0.68$) and no statistically significant difference between the flat and the cube thumbnails ($p = 0.3$).

For the item search task, the one-way repeated measures ANOVA results show no statistically significant differences in the amount of thumbnail pickups due to thumbnail type ($F(2,60) = 0.9773, p = 0.38$).

5.4 User Experience (UEQ-S)

The results of the UEQ-S are analyzed with the help of the UEQ Data Analysis Tool by M. Schrepp [1]. The values filled into the questionnaire are first transformed from values between 1 and 7 to values between -3 (the most negative response) and +3 (the most positive response). The first four questions of the questionnaire focus on pragmatic quality, while the last four questions focus on hedonic quality. The means of both these qualities per thumbnail type can be seen in Table 2. The pragmatic quality is the highest for the sphere thumbnail and the lowest for the cube thumbnail. The hedonic quality is the highest for the sphere as well, and the lowest for the equirectangular thumbnail type.

To evaluate and compare the usability of the three thumbnail kinds in a meaningful way, it is common to compare the data to an existing benchmark dataset. The benchmark dataset contains data from over 21.000 participants from 468 different studies concerning a wide variety of different products (such as business software, web pages, web shops and social networks). This benchmark is based on the full UEQ version and not the UEQ-S (as the UEQ-S is still relatively new and there is no such benchmark yet), so it is best to interpreted the results as a rough approximation. It is still useful to illustrate the relative quality of the different thumbnail types. In Figure 8 we see the values from Table 2 presented against the benchmark. This presents the sphere thumbnail to be in the 'above average' to 'good' range. The cube thumbnail falls in the category 'bad' for the pragmatic quality, but in the 'above average' category for the hedonic quality, ending up with an overall score of 'below average'. The flat thumbnails have a pragmatic quality that is in the 'below average' category, and a terribly low hedonic quality, resulting in a bad overall score.

Thumbnail	1st place	2nd place	3rd place	μ	σ	σ^2
Flat	32.26% (10)	19.35% (6)	48.39% (15)	2.16	0.88	0.78
Sphere	51.61% (16)	45.16% (14)	3.26% (1)	1.52	0.56	0.31
Cube	16.13% (5)	35.48% (11)	48.39% (15)	2.32	0.74	0.54

Table 3: Results of the ranking of the thumbnail types. Values in brackets indicate number of participants who ranked this thumbnail at that position. The last columns show the mean ranking place, standard deviation, and variance.

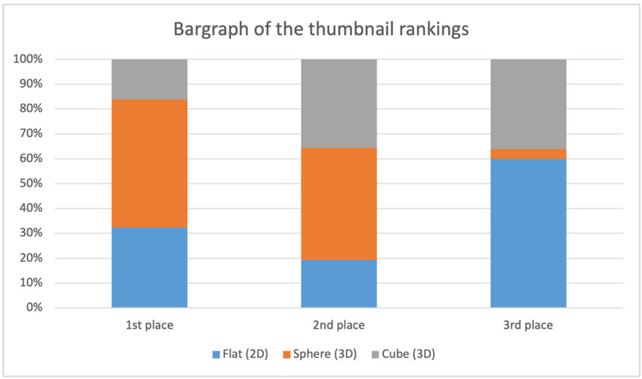

Fig. 9: Bar graph of the rankings per thumbnail type (in %).

5.5 Rankings of the thumbnails

In the last question of the questionnaire that the participant filled in during the experiment, asks for a ranking of the three thumbnail kinds from best to worst, based on the experience with them. Participants were not given an explicit measure to compare them with, but were encouraged to decide what 'best' and 'worst' meant to them. The results of these rating can be seen in Table 3. An average, the participants ranked the sphere thumbnails the highest, resulting in the lowest mean value of 1.52 and the lowers standard deviation and variance. The flat and the cube thumbnails have quite comparable mean placements of 2.16 and 2.36, respectively. The percentages from Table 3 are visualized in Figure 9, illustrating that the sphere thumbnail is placed on the first place most often and ranked last only once. The flat thumbnail is ranked last the most often.

5.6 Advantages and disadvantages

The final questionnaire of the experiment also asked to describe what participants consider to be advantages and disadvantages of each thumbnail type. The results of those answers are summarized in this section.

Flat thumbnails. There is not much variety in the answers about the advantages of the flat thumbnails. 22 out of the 31 participants mentioned that it is nice that you can see the entire view of the 360° video at once. Another advantage is the fact that flat thumbnails are conventional and previously known, which is mentioned by six participants. The disadvantage that is noted most often is the distortion, as mentioned by 17 participants. Other disadvantages are that the flat thumbnails do not make use of the 3D possibilities that VR can provide (mentioned by six participants) and that the flat thumbnails are less immersive (mentioned by five participants). Six characterized them as "boring", "plain", or "less fun". Four participants mentioned that orientation and perspective was harder to see with the flat thumbnails. Additionally, four participants answered that grabbing the thumbnail does not add much value. Details are hard to see on the flat thumbnails (mentioned by three participants). Lastly, two participants said that it was annoying that the left and right side of the projection are not connected.

Sphere thumbnails. The most commonly mentioned advantage of the sphere thumbnails is that they feel like a more logical and intuitive way of representing the video (mentioned by eleven participants). Seven participants mentioned that the orientation and perspective of the video is easy to understand. The appearance of the thumbnails was characterized as positive (e.g.,looking "good", "cool", or "interesting") seven times. The fact that the sphere thumbnails were fun to pick up was mentioned by seven participants as well. Another advantage is the little to no distortion (mentioned by five participants). Lastly, four participants said that the sphere thumbnails provided greater immersion, and three stated that they are easy to understand or easy to see from afar. As for the disadvantages of the spheres, the one that was mentioned most often (13 times) is that you cannot see the entire view of the video without rotating. Seven participants mentioned that they experienced some strain on the wrist or annoyance with a lack of wrist flexibility. Three participants mentioned that it is hard to see and interpret what is in the video when looking at the thumbnail from a distance.

Cube thumbnails. Five participants answered that they could not think of any advantages for the cube thumbnails. Six answered that they look cool, fun and/or interesting. Four participants mentioned positively that that there is less distortion and stretching of the video compared to the flat thumbnail. Four characterized the cube thumbnails as clear and easy to understand. The appearance of the cube thumbnails is brought up as well. Two participants simply answered that the fact that the thumbnail is 3D is an advantage. It is clearer to interpret from a distance, compared to a sphere (answered by two participants) and the orientation and perspective is easy to understand (mentioned by three participants). Three stated that they show a great amount of the scene at once. There are two main disadvantages that were mentioned most often: The fact that one cannot see the entire full view at once (answered by eleven participants) and that the edges and corners of the cube have distortion or do not look nice (13 participants). Seven participants said that the cube feels non-intuitive or unnatural. Lastly, two participants mentioned that the ceiling and the floor of the cube were harder to interpret than the sides, and two said that it was difficult on the wrist to fully turn them.

5.7 Cybersickness

To investigate a possible effect on wellness, we measured cybersickness by means of a self-report. Immediately after the participants completed the VR portion of the experiment, they answered questions regarding cybersickness. This includes questions about whether they experienced certain well-known symptoms of cybersickness, as well as rating their own cybersickness on a scale from 0 to 10. 10% of participants experienced slight headaches. Around 10% experienced some dizziness. Only about 3% reported some nausea. About 16% of participants reported they felt uncomfortable at some point during the VR portion of the experiment. The mean value of cybersickness on the scale from 0 to 10 is 0.81 (std. deviation 1.22 and variance 1.49). Only one participant rated cybersickness with four, and all others gave lower rating with a large majority (61%) expressing no problems (rated zero). The participants were made aware that they could pause the experiment at any time, to take the headset off and take a break, or to stop the experiment completely. None of the participant needed or wanted a break during the VR session.

6 Discussion

In the following, we discuss the above results with respect to the research questions introduced in Section 3.2: RQ1: How fast can users identify whether a video contains a certain item of interest? RQ2: How fast can users identify the high-level subject of a video? RQ3: How do users rate the user experience?

6.1 RQ1: Finding items in thumbnails

To answer RQ1, we look at the results from the measuring the time participant took for the item search tasks (Section 5.1), the correctness of these tasks (Section 5.2) and the amount of pickups of thumbnails (Section 5.3).

The results from analyzing the time measurements taken during item search tasks show that there is no statistically significant difference between the time needed to complete the task for any of the thumbnail types. Furthermore, we found no statistically significant difference in the amount of thumbnail pickups. Although the actual numbers are slightly higher for the sphere, this is a surprisingly positive outcome because we expected a larger increase in time here due to their more interactive nature and a need to pick them up if one wants to explore all viewing directions. We did observe statistically significant differences in the correctness. The correctness was significantly higher for the sphere thumbnails, compared to the other two thumbnail kinds. From these results, we conclude that the sphere thumbnails lead to the best performance in the item search task, without the need to spend significantly more time on it.

6.2 RQ2: Identifying high-level subject

To answer RQ2, we look at the results from measuring the time participant took for the categorization tasks (Section 5.1), the correctness of these tasks (Section 5.2) and the amount of pickups of thumbnails (Section 5.3).

The results from analyzing the time measurements show that the categorization tasks with the flat thumbnail are completed considerably faster than with the sphere or the cube thumbnails. To some degree, this was expected, because people can easily see all viewing directions at once and for a high-level categorization of the content (e.g., underwater videos), the distortion of parts of the image will have a much lesser effect than when searching for details in it. Yet, one could argue equally that the fact of not being able to see all viewing directions is not that relevant for a classification task, which does not seem to be the case. Another explanation could be that people were encouraged to pick them up, even when not really necessary. For the correctness, we found no statistically significant difference between any of the thumbnail types. The results on whether there are more thumbnail pickups is inconclusive, as the one-way repeated measures ANOVA does show statistically significant differences, but the post hoc Tukey's test does not. All in all, we can conclude that recognizing high-level subject of a video from the 2D thumbnail is the fastest, without compromising the selection accuracy.

6.3 RQ3: User Experience

To answer RQ3, we look at the results from the UEQ-S (Section 5.4), the participant's rankings (Section 5.5), the advantages and disadvantages mentioned by the participants (Section 5.6) and the cybersickness (Section 5.7).

The results from the UEQ-S showed that the sphere thumbnails had the highest pragmatic quality, the highest hedonic quality and therefore also the highest overall quality. The cube had a bad pragmatic quality, but also a good hedonic quality. An important finding is the very bad hedonic quality of the flat thumbnails. This could be explained by the disadvantages mentioned by the participants, such as the distortion and the fact that the flat thumbnails were experienced as less immersive. This result point towards the fact that 3D thumbnails are considered way more enjoyable to use than the flat thumbnail. This trend is similar when looking at the rankings of the thumbnails. The sphere thumbnail is ranked in first place by the majority of the participants and is never placed last. The cubes seem to be favoured less, as it is rarely placed on first place. The opinion about the flat thumbnails seem to be more divided, as participants rarely place the flat thumbnail on second place, but often on first or last place in their ranking. The participants reported only little cybersickness, which is why we consider its possible effect on the user experience to be negligible. From all these results, we can conclude that 3D thumbnails (in particular the sphere thumbnail) lead to a positive user experience.

6.4 Limitations

During the study, we experienced a few performance-related issues. Ideally, the view from the VR headset would be wirelessly streamed to a separate screen, so that the researcher could see what the participant of the experiment was seeing at all times. This would be useful when participants had questions or needed clarification. However, this was not done because, combined with the high rendering requirements, it resulted in an occasional overheating of the Oculus Quest, which would have impacted the experiments negatively. Another performance-related issue was that sometimes the thumbnails took a noticeable amount of time (anywhere from zero to two seconds) to load the previews onto the thumbnail shapes, despite being saved locally on the Oculus Quests instead of streamed from an external source.

We also noticed that there were a few tasks that might have been interpreted differently across participants. Effort was taken to minimize this as much as possible when creating the search tasks, but some instances of confusing occurred nonetheless.

Lastly, because 3D thumbnails are a novel concept introduced in this paper, there is no knowledge yet about how to best interact with them. The interaction method that was designed was mostly inspired by literature and implementations common in other domains. Although we did not observe any interaction problems, exploring other options might be worthwhile in future work. Novelty could have impacted the results about user experience positively for the sphere and cube, because often a non-standard approach is rated higher with respect to experience due to its uniqueness, and performance negatively, because people might perform better with familiar approaches. Yet, given that most participants were rather unfamiliar with 360° videos in VR, we expect a rather limited influence on our results.

7 Conclusion and Future Work

The goal of this research was to verify the usefulness of 3D thumbnails as a better alternative to standard 2D representations when exploring and accessing large video collections in VR. To do this, we investigated which 360° video presentation leads to the best user experience, which is the most effective at conveying the high-level subject matter of a video, and which is the most effective when looking for a certain specific subject in the video. We found that the flat thumbnails lead to the fastest performance in the categorization tasks, without compromising accuracy. Despite this, we conclude that the sphere thumbnails have the most potential because we found that the sphere thumbnail leads to the best user experience. Furthermore, we found that the item search tasks were performed most accurately with the sphere thumbnails, without needing more time for these tasks, compared to the other thumbnail kinds. Since 3D thumbnails for 360° videos in VR is an entirely novel concept, we hope this research can inspire researchers to investigate this concept in more depth. When more knowledge is gathered about three-dimensional thumbnails, they can be used in video browsing applications, and possibly influence what these applications will look like, enhancing the 360° video watching experience. Our current results therefore suggest a combined approach. A flat thumbnail representation might be best as start, because it allows users to get a quick overview of the content and is sufficient for high-level classification. But for further inspection, when searching for individual items in a video, or just to provide a better, more immersive user experience, our new concept of 3D thumbnails clearly showed a benefit that might even increase in future work. Possible avenues to further explore are different 3D shapes. The cube fell surprisingly short in the results, despite its obvious advantage to provide a better spatial orientation due to implicitly highlighting fixed dedicated locations via its edges. Thus, a mixture between a cube and a sphere (e.g., cube versions with corners that are flattened more and more – with either sharp or soft edges) seems promising. One participant suggested projecting the entire image onto half of a sphere, creating a mixture between a sphere and an equirectangular projection, which might therefore combine the advantages of both.

Another possibility for future research is exploring other interaction methods with 3D thumbnails. One possibility could be to investigate other ways of rotating the 3D thumbnails, as some participants mentioned the lack of wrist flexibility or strain on their wrist when trying to view the full view of the thumbnails. The option to rotate a sphere like a real-world globe would prevent this.

To take it further, we also want to investigate how to represent these thumbnails within the 3D environment. For this evaluation, we used the common flat arrangement of thumbnails along a grid. Yet, we can take advantage of the full 3D space. Promising options that come to mind include non-flat arrangement, but also groups of thumbnails for videos of the same category where distances between videos and groups reflect the level of relationship.

References

[1] Short UEQ data analysis tool.xlsx. https://www.ueq-online.org/. 7

[2] T. Aitamurto, S. Zhou, S. Sakshuwong, J. Saldivar, Y. Sadeghi, and A. Tran. Sense of presence, attitude change, perspective-taking and usability in first-person split-sphere 360° video. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–12. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3174119 2

[3] J. Ardouin, A. Lécuyer, M. Marchal, and E. Marchand. Navigating in virtual environments with 360° omnidirectional rendering. In *2013 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 95–98, 2013. doi: 10.1109/3DUI.2013.6550203 2

[4] C. Brown. Bringing pixels front and center in VR video. Google AR & VR blog posting, https://blog.google/products/google-ar-vr/bringing-pixels-front-and-center-vr-video/, March 2017. 2

[5] T. T. Elvins, D. R. Nadeau, and D. Kirsh. Worldlets—3d thumbnails for wayfinding in virtual environments. In *Proceedings of the 10th Annual ACM Symposium on User Interface Software and Technology*, UIST '97, p. 21–30. Association for Computing Machinery, New York, NY, USA, 1997. doi: 10.1145/263407.263504 2

[6] T. T. Elvins, D. R. Nadeau, R. Schul, and D. Kirsh. Worldlets: 3d thumbnails for 3d browsing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '98, p. 163–170. ACM Press/Addison-Wesley Publishing Co., USA, 1998. doi: 10.1145/274644.274669 2

[7] T. T. Elvins, D. R. Nadeau, R. Schul, and D. Kirsh. Worldlets: 3-D Thumbnails for Wayfinding in Large Virtual Worlds. *Presence: Teleoperators and Virtual Environments*, 10(6):565–582, 12 2001. doi: 10.1162/105474601753272835 2

[8] D. Englmeier, W. Sajko, and A. Butz. Spherical world in miniature: Exploring the tiny planets metaphor for discrete locomotion in virtual reality. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 345–352, 2021. doi: 10.1109/VR50410.2021.00057 2

[9] C.-L. Fan, W.-C. Lo, Y.-T. Pai, and C.-H. Hsu. A survey on 360° video streaming: Acquisition, transmission, and display. *ACM Comput. Surv.*, 52(4), aug 2019. doi: 10.1145/3329119 1, 2

[10] K. F. Gauss and P. Pesic. *General Investigations of Curved Surfaces*. Courier Corporation, Oct. 2005. Google-Books-ID: VsMsAwAAQBAJ. 2

[11] N. Greene. Environment mapping and other applications of world projections. *IEEE Computer Graphics and Applications*, 6(11):21–29, Nov. 1986. Conference Name: IEEE Computer Graphics and Applications. doi: 10.1109/MCG.1986.276658 2

[12] S. Grogorick, M. Stengel, E. Eisemann, and M. Magnor. Subtle gaze guidance for immersive environments. In *Proceedings of the ACM Symposium on Applied Perception*, SAP '17. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3119881.3119890 2

[13] Y. He, X. Xiu, P. Hanhart, Y. Ye, F. Duanmu, and Y. Wang. Content-adaptive 360-degree video coding using hybrid cubemap projection. In *2018 Picture Coding Symposium (PCS)*, pp. 313–317, 2018. doi: 10.1109/PCS.2018.8456280 2

[14] M. Hosseini. View-aware tile-based adaptations in 360 virtual reality video streaming. In *2017 IEEE Virtual Reality (VR)*, pp. 423–424, Jan. 2017. doi: 10.1109/VR.2017.7892357 6

[15] H.-N. Hu, Y.-C. Lin, M.-Y. Liu, H.-T. Cheng, Y.-J. Chang, and M. Sun. Deep 360 pilot: Learning a deep agent for piloting through 360° sports videos. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1396–1405, 2017. doi: 10.1109/CVPR.2017.153 2

[16] M. Husung and E. Langbehn. Of portals and orbs: An evaluation of scene transition techniques for virtual reality. In *Proceedings of Mensch Und Computer 2019*, MuC'19, p. 245–254. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3340764.3340779 2

[17] J. Kim. Get a closer look with Street View in Google Earth VR. Google AR & VR blog posting, https://blog.google/products/google-ar-vr/get-closer-look-street-view-google-earth-vr/, September 2017. 2

[18] H. Kuperus. Can you recognize it? How size and movement affect recognizability of equirectangular projections from immersive videos. MSc thesis, https://studenttheses.uu.nl/handle/20.500.12932/41311. 2

[19] Y. Lee and C.-W. Tang. Early skip mode decision of versatile video coding on 8k 360-degree videos. In *2021 IEEE International Conference on Consumer Electronics (ICCE)*, pp. 1–3, 2021. doi: 10.1109/ICCE50685.2021.9427611 2

[20] J. Li, J. Lyu, M. Sousa, R. Balakrishnan, A. Tang, and T. Grossman. Route tapestries: Navigating 360° virtual tour videos using slit-scan visualizations. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, UIST '21, p. 223–238. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3472749.3474746 2

[21] K. Li, Z. Wu, K.-C. Peng, J. Ernst, and Y. Fu. Tell me where to look: Guided attention inference network. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9215–9223, 2018. doi: 10.1109/CVPR.2018.00960 2

[22] Z. Li, S. Miyafuji, T. Sato, and H. Koike. Omnieyeball: Spherical display embedded with omnidirectional camera using dynamic spherical mapping. In *Adjunct Proceedings of the 29th Annual ACM Symposium on User Interface Software and Technology*, UIST '16 Adjunct, p. 193–194. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2984751.2984765 2

[23] Y.-T. Lin, Y.-C. Liao, S.-Y. Teng, Y.-J. Chung, L. Chan, and B.-Y. Chen. Outside-in: Visualizing out-of-sight regions-of-interest in a 360° video using spatial picture-in-picture previews. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, p. 255–265. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3126594.3126656 2

[24] S. Miyafuji, T. Sato, Z. Li, and H. Koike. Qoom: An interactive omnidirectional ball display. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, p. 599–609. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3126594.3126607 2

[25] A. Nassani, L. Zhang, H. Bai, and M. Billinghurst. Showmearound: Giving virtual tours using live 360 video. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI EA '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411763.3451555 2

[26] C. Nguyen, S. DiVerdi, A. Hertzmann, and F. Liu. Vremiere: In-headset virtual reality video editing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, p. 5428–5438. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3025453.3025675 2

[27] T. Pakkanen, J. Hakulinen, T. Jokela, I. Rakkolainen, J. Kangas, P. Piippo, R. Raisamo, and M. Salmimaa. Interaction with WebVR 360° video player: Comparing three interaction paradigms. In *2017 IEEE Virtual Reality (VR)*, pp. 279–280, 2017. doi: 10.1109/VR.2017.7892285 2

[28] D. Pio and E. Kuzyakov. Under the hood: Building 360 video. Facebook Engineering, https://engineering.fb.com/video-engineering/under-the-hood-building-360-video/, July 2018. 2

[29] R. Skupin, Y. Sanchez, Y.-K. Wang, M. M. Hannuksela, J. Boyce, and M. Wien. Standardization status of 360 degree video coding and delivery. In *2017 IEEE Visual Communications and Image Processing (VCIP)*, pp. 1–4, 2017. doi: 10.1109/VCIP.2017.8305083 2

[30] R. Stoakley, M. J. Conway, and R. Pausch. Virtual reality on a WIM: Interactive worlds in miniature. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, p. 265–272. ACM Press/Addison-Wesley Publishing Co., USA, 1995. doi: 10.1145/223904.223938 2

[31] T. Teo, A. F. Hayati, G. A. Lee, M. Billinghurst, and M. Adcock. A technique for mixed reality remote collaboration using 360 panoramas in 3d reconstructed scenes. In *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology*, VRST '19. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3359996.3364238 2

[32] S. Tomar. Converting video formats with ffmpeg. *Linux Journal*, 2006(146):10, 2006. 6

[33] A. Truong and M. Agrawala. A tool for navigating and editing 360 video of social conversations into shareable highlights. In *Proceedings of Graphics Interface 2019*, GI 2019. Canadian Information Processing Society, 2019. doi: 10.20380/GI2019.14 2

[34] A. Tse, C. Jennett, J. Moore, Z. Watson, J. Rigby, and A. L. Cox. Was I there? Impact of platform and headphones on 360 video immersion. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, p. 2967–2974. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3027063.3053225 2

[35] S. Yamaguchi, N. Ogawa, and T. Narumi. Now I'm not afraid: Reducing fear of missing out in 360° videos on a head-mounted display using a panoramic thumbnail. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 176–183, 2021. doi: 10.1109/ISMAR52148.2021.00032 2

[36] Y. Ye, E. Alshina, and J. Boyce. Jvet-h1004: Algorithm descriptions of projection format conversion and video quality metrics in 360lib. *Joint Video Exploration Team (JVET) of ITU-T SG*, 07 2017. 2