Universiteit Utrecht

Viewpoint-controlled Navigation in Virtual Reality for Exploratory Image Browsing

Author:
Joris BOLWERK

Supervisor:
Dr. Wolfgang HÜRST

Examiner:
Prof. dr. Marc van KREVELD

A thesis submitted in fulfillment of the requirements for the degree of Master of Science

at the

Department of Information and Computing Sciences

January 3, 2017

UTRECHT UNIVERSITY

Abstract

Faculty of Science
Department of Information and Computing Sciences

Master of Science

Viewpoint-controlled Navigation in Virtual Reality for Exploratory Image Browsing

by Joris BOLWERK

Virtual reality allows users to explore virtual worlds in a close-to-natural way. As the current trend of gaming and tele-presence applications suggests, this has much potential for realistic experiences. But there is more. Not only is the experience realistic, it is also intuitive and thus we aspire to explore the potential of virtual reality for abstract environments. In this thesis we set out to explore the suitability of virtual reality for exploratory image browsing. In particular viewpoint control and the effect of visualizations are examined. Our participants expressed more enthusiasm for exploratory image browsing in a confined domain of their expertise than for broad applications to discover new images. During the navigation, some of the participants looked at center of the view, others towards the edges. Some used the structure to navigate, others the passing images. The fisheye views were appreciated if the view highlighted regions of interest. Distortion of the sphere fisheye view approach was acceptable for layouts with other information such as structure and similarity, whereas the stacking fisheye view approach suffered from incorrectly rectifying depth levels in the stack. Using the viewpoint to navigate was very easy and intuitive to learn and quite satisfactory, although the resolution of head-mounted displays cannot compete with that of high resolution screens yet.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Over three billion photos a day were shared on social media in 2015. Snapchat and WhatsApp both took account for just over 1 billion per day. Slightly less than half a billion photos were shared on Facebook, around a quarter billion on Facebook Messenger, 100 million photos on Instagram, and 1 million on Flickr. (Meeker, 2016) (Smith, 2016)

In order to keep their collection of images accessible, subsets of images are retrieved by means of keyword searches or albums associated with a user. Flickr also offers the functionality to explore the whole collection of publicly available images. This could be interesting for other image sharing platforms as well, because it opens up the scope of the user and thus adds a new layer of social connectivity. For instance, someone could browse through images of arbitrary fashion bloggers rather than settle for only those images uploaded by acquainted bloggers.

1.1 User needs for image browsing

Fortunately, the user needs for image browsing are not limited to fashion blogs. A news article could call for a photo that serves the context, one could look up pictures of last night's event to share with friends, or one could sit on the couch and just want to look at beautiful images such as art, landscapes and sunsets.

These scenarios usually call for different ways of interaction. Finding pictures of last night's event is easy if the collection is arranged by time. Searching a photo for a news article requires some form of interaction to specify what the article is about, whereas checking out beautiful images calls for a system that alleviates the interaction such that the user is able concentrate on the beauty of the images.

In case of the latter, we believe that accessing virtual reality through a head-mounted display offers a great opportunity for intuitive interaction and thus more enjoyable experiences.

1.2 Image browsing in virtual reality

In order to validate this belief, we delve into the characteristics of virtual reality. According to Bowman and McMahan (2007), immersion in a virtual reality can lead to engaging experiences, enhanced peripheral awareness, and reduction of information clutter. Especially peripheral awareness can be exploited to interact with the environment rather than isolated search windows. We presume that this helps the user to process information about the search more subconsciously and, as a result, focus on the presented images.

1.2.1 Peripheral awareness

Peripheral awareness is closely related to peripheral vision which is the vision outside the fovea. The vision on the fovea covers a field of view of $60°$. The range from $60°$ to almost $180°$ in horizontal direction belongs to the peripheral vision (Strasburger, Rentschler, and Jüttner, 2011). The quality of vision degrades towards the edges. Therefore, the peripheral vision is sometimes characterized by its blurriness (Strasburger, Rentschler, and Jüttner, 2011). Others argue that it should not be considered as losing focus, but rather as losing the quality of form (Lettvin, 1976). This distinction is important for our concept of peripheral awareness. This means that users are able to detect movements in the peripheral vision, even though it may be unclear what kind of object is moving. This thesis considers peripheral awareness as the ability to combine peripheral vision with knowledge of the environment in order to be aware of this environment even while it changes. Additionally, Bateson (1994) argued that peripheral vision enables us to see new things beyond our focus that originate from our expectations and cultural preferences.

Thus, the virtual environment should provide information all around the user to improve peripheral awareness. It is our assumption that this allows the user to focus on a particular image while still being presented with the information of neighboring images. If any of these neighboring images draws attention, the user navigates towards it. For this purpose, we consider spatial navigation based on the location of the viewpoint.

1.2.2 Spatial navigation

Spatial navigation is the ability to navigate through a structure according to the spatial location in that structure. In case the viewpoint is used to control the spatial location, users rotate their heads into the same direction as the target while maintaining the ability to scan the environment with their eye gaze. Hence, we hypothesize that viewpoint-control allows users to easily switch between looking at images and navigating through the layout.

1.2.3 Link-based navigation

To the best of our knowledge, there is just one pilot study concerned with image browsing in virtual reality. Khanwalkar, Balakrishna, and Jain (2016) implemented a system with two modes. In the first mode, concepts extracted from the featured image can be selected to render a new set of images. The second mode works quite similar, but with a hierarchy of categories instead of concepts related to the featured image. They evaluated their system with these two modes and observed that participants had a positive opinion of exploring images in infinite virtual space and all but one would use such an application in the future.

In spite of these promising results, it is our expectation that their approach is less suitable in case the search intention cannot be expressed in categories or concepts. Instead of validating a system, we strive to identify how people interact with a layout of images surrounding them. This way, we aim for a better understanding of what makes virtual reality suitable for image browsing rather than just showing that it is appreciated.

1.3 Research purpose

The purpose of this thesis is to explore spatial navigation methods for exploratory image browsing in virtual reality. In particular, visualizations of the viewpoint are explored. For this purpose, we examine the question:

Is viewpoint-controlled navigation in virtual reality suitable for exploratory image browsing?

A combination of questionnaires and interviews is used to gain insight into the user experience. Besides general user experience, we are also interested in the effect of highlighting the spatial location of the viewpoint and whether this is related to the layout.

1.4 Overview

Chapter 2 discusses related work and section 2.4 specifies the sub-questions that arose during the literature study. Next, Chapter 3 describes the techniques used to implement our system, which decisions were made and on what grounds. This is followed by Chapter 4 that describes the evaluation method and procedure. Chapter 5 shows the results of the interviews and analyses of the questionnaires and discusses the realization of the subgroups and factors to analyze. Chapter 6 subsequently discusses the findings in relation to the sub-questions and research purpose. The research question and sub-questions are recalled. Finally, Chapter 7 discusses interesting research directions and limitations of the current experiments that came up during this thesis or were put on hold to keep focus on the research purpose.

Chapter 2

Related Work

2.1 Image browsing

Large image collections are usually diverse and thus relevant for a variety of user needs. The first part of this section elaborates on the user needs for image browsing through large image collections and how the experience of a virtual reality could contribute to the needs. The second part discusses interfaces for browsing large image collections in virtual reality.

2.1.1 Professional needs for image browsing

Markkula and Sormunen (2000) studied the search strategies of journalists in photo archives of newspapers and observed that browsing is an essential strategy. To identify the needs of archivists, they examined the following field studies of respectively Enser, Ørnager and Keister:

- 70 percent of the requests received by the Hulton Deutsch Collection from newspapers and magazines were of specific entities, events or locations (Enser, 1993).
- Archivists from 13 newspaper archives estimated that half of the requests were clear, concerning persons, and the other half was abstract, concerning themes (Ornager, 1995).
- The archive of the National (United States) Library of Medicine received around one third to one half descriptions of how the image should look like. The other requests were topical definitions without specific visual requirements (Keister, 1994).

Their field study was concerned with journalists sending requests rather than archivists receiving them. Their results showed that the archives were used mostly to satisfy the need for images of named entities or current news events. Results also showed that it was considered difficult to find generic object types or themes. This led to the observation that browsing is essential to develop ideas and apply search criteria that are hard to express in words. Although essential, they identified that browsing sometimes required a lot of time and effort. Hence, there is a demand for new interfaces to retrieve images easily in a short amount of time.

Virtual reality offers new options to design interfaces. Unfortunately, the effect of these phenomena on the duration of the search tasks seems meager as there is no direct relation between them. Given the additional issues of current virtual reality displays, such as the relatively low resolution and high processing power, it does not seem wise to pursue virtual reality interfaces for such professional needs at this time.

2.1.2 Leisure needs for image browsing

Jansen (2008) tried to figure out how users construct search queries for images on the web. Three evaluators examined and classified the queries into 4 categories: unique, unique with refiners, non-unique and non-unique with refiners. A unique query is usually a specific person, object or event. Non-unique queries are less specific. Searching for 'Donald Trump' would be categorized as a unique query, whereas the query 'Politicians with an odd hairstyle' would be non-unique. Refiners are additional search attributes. In the case that all three evaluators disagreed, that is if none of the categories was chosen more than once, the query was discarded. They found that 87.1 percent of the queries had at least one refiner. Almost half, 49.4 percent, of the refiners specified an image collection and from the queries with refiners, 71.9 percent was categorized as non-unique and 15.2 percent as unique.

Thus, users aimed for a collection of images in almost half of the inspected queries and did not search for specific person, object or event in a majority of them. This indicates that there are many cases in which people tend to search for broad subsets of results.

André et al. (2009) studied the number of clicks and query trails per session in the log files of a major web search engine. The query trails consisted of a sequence of queries. The sessions were ended by periods of more than 30 minutes without activity. They found that 30 percent of the image queries and 43 percent of the web queries contained a click. However, the average number of clicks per image query was larger than for web queries. Also, the depth of image queries was more than twice the depth of text-based web queries on average. The depth of a query was measured by the number of consecutive result pages viewed before the query was refined.

The query trails were categorized into three groups: unrelated, tangential and related. Related queries were usually refinements upon the previous query (Hawaiian spiders -> Hawaiian sugarcane spiders). Tangential queries were inspired by the previous query, but not related (horned lizard desert -> desert color). Unrelated queries did not have anything in common (fitness -> deal or no deal). From the image query trails, 70 percent was judged as related, 10 percent as tangential and 20 percent as unrelated. For web query trails, the percentages were respectively 60, 5 and 35 percent.

The results of André et al. contribute to the speculation that there are many cases in which people tend to search for broad subsets of images. The number of result pages viewed for images was more than twice as high as for text-based queries. Furthermore, the image queries were more often inspired by or related to the previous queries than text-based web queries. A browsing interface could take this into account by providing related refinements. A keyword-based interface can allocate space for these recommended refined queries. A context-based interface may arrange the collection by a set of characteristics, such that the user is able to refine the results based on knowledge of the arrangement.

2.1.3 Interfaces for image browsing

Keyword-based and similarity-based are the two dominant types of interfaces for image search and browsing through large collections (Yee et al., 2003). Major search engines apply the same keyword-based interface for images as traditional web search interfaces (André et al., 2009). These traditional web search interfaces consist of an input field for the keywords and a paginated ranked list of results. Many similarity-based interface systems extract features from the images when added to the collection and use these features to arrange the collection (Rodden et al., 2001).

A simple way to arrange the collection is sorting on a single feature per dimension. If there are more features to consider than dimensions available for representation, images with a small distance between their feature vectors can be clustered in a lower dimensionality. Another option is the use of anchor images. These anchors are chosen with a variety of similarity distances from each other. Upon the selection of an anchor, a new set of anchors is chosen with a variety of similarity distances, but close to the selected anchor (Datta, Li, and Wang, 2005).

Anchor-based approaches show a diverse subset of images from over the whole collection and upon selection the subset is replaced by a new subset of more related images. Since our envisioned environment consists only of images, replacing them seems to break the immersion. Therefore, this section focuses on sorting and clustering.

Clustering on similarity features

To visualize the clusters in a three-dimensional environment, the number of dimensions should be reduced to a few dimensions and at most three of them can be displayed at a time. Two dimensionality reduction techniques are multidimensional scaling and self-organizing maps. Multidimensional scaling algorithms aim to preserve the higher-dimensional similarity distances in the reduced space as well as possible. Self-organizing maps use unsupervised competitive learning to map the input vectors to a position in the map space.

Chen et al. (1998) showed that their implementation of a Self-organizing map algorithm was appreciated for broad browsing tasks due to the graphical aspects of the map, whereas the users performing a directed search or browsing using an alphabetical order expressed that the algorithm did not work well. Kleiman et al. (2015) designed an interface that only considers the local neighbors of images using a k-nearest neighbor algorithm and greedy map generation. This idea was based on the assumption that for high-dimensional data, such as images, the neighbors at a short distance are more relevant than those farther away and thus it would suffice to only consider the k closest neighbors per image in the data set. This made their algorithm computationally inexpensive and dynamic, but at the cost of a bit of uncertainty about the outcome due to locality.

Sorting on metadata

Besides clustering images with similar features in a low-dimensional representation, the images can also be sorted on low-dimensional characteristics such as location, time, and other metadata associated with the images. This facilitates the same representations as with the clustering methods, but uses these characteristics instead of the similarity distances between the feature vectors.

Yee et al. (2003) evaluated a multifaceted metadata interface for a collection of art images and compared it to a keyword-based interface. Half of the participants preferred the keyword-based interface over the metadata interface to find images of roses. For refined searches, however, their metadata interface was clearly favored over the baseline. Given the high percentage of refiners in the study of Jansen and the large amount of related queries in the study of André et al. as discussed in section 2.1.2, image searches on the web may benefit from a metadata interface as well. Unfortunately, the interface proposed by Yee et al. was designed specifically for the art collection of 35,000 images and thus it seems unlikely that it scales well to large and diverse image collections.

Representation

What kind of context is suitable depends on the needs of the user. For example, Location-based context may be desired when the user intent is to find an image based on geographical knowledge. Context about the environment of the user, such as time, place and weather, could also influence the user intent (Kofler, Larson, and Hanjalic, 2016).

2.2 Virtual reality

One of the key features of virtual reality is enhanced immersiveness (Burdea and Coiffet, 2003). The following sections discuss the topic of immersion in virtual reality. In particular, the next section discusses human senses and depth perception and the succeeding section reflects on the ability of virtual reality devices to stimulate depth perception. This is followed by a section about interaction between a user and the virtual reality and a section specific to viewpoint-controlled interaction.

2.2.1 Immersion in virtual reality

All senses can be stimulated to enhance the experience of a virtual reality. Displays enable you to *see* the virtual environment, audio systems make it possible to *hear* sounds coming from a virtual direction, odors and flavors create the illusion of *smell* and *taste*, and haptic devices allow a person to *feel* objects in the environment. In this thesis the assumption is that it only makes sense to stimulate the senses that are affected by the chosen environment. If such an environment consists of images, these images only affect the visual sense and thus we look into displays to show the virtual environment.

In this and the succeeding section, we regard three types of displays: Monitors, head-mounted displays and surround-screen environments. Monitors are defined as a fixed window into the virtual environment, head-mounted displays as a window into the virtual environment that rotates along with the rotation of the user's head, and surround-screen environments as projections of the virtual environment all around the user. According to Cruz-Neira, Sandin, and DeFanti (1993), virtual reality displays can be distinguished from conventional workstation graphics on the basis of the depth information that they support. Cutting (1997) defined nine sources of information for the perception of depth. Their lists differ slightly on the definitions of perspective projection and atmospheric perspective. Cruz-Neira, Sandin, and DeFanti (1993) considered lightning and shadows as separate sources of information, and Cutting (1997) split perspective projection into three sources: Relative size, height in the picture plane, and relative density. Furthermore, they agree on the same sources. The sources of depth information associated with virtual reality are:

- Binocular disparity
- Motion parallax
- Convergence
- Accommodation

Binocular disparity involves depth information caused by the difference in projections onto the left and right eye, motion parallax introduces depth information that is retrieved from looking at an object from different angles by head motion, convergence results in depth information by rotating the eyes inwards, and accommodation is the phenomenon of focusing the eyes to gain information about the depth. Accommodation is closely related to convergence. When a person looks at an object nearby, the lenses of the eyes become more convex to focus and the eyes converge to keep both eyes on the object.

The sources of depth associated with virtual reality are distinctive from the other sources, because they depend on actions in the physical space to interpret the informational space (Jensen et al., 2002). In other words, the user performs physical actions to perceive the virtual information space. This may lead to higher immersion, as Slater and Wilbur (1997) identified that a match between the proprioceptive feedback about the body movements and the information about the virtual environment on the display is an aspect of immersion.

2.2.2 Sources of depth information provided by virtual reality displays

Monitors do not support any of the sources of depth information associated with virtual reality by default. However, the monitors can be extended with additional tools to provide these depth sources. Stereoscopy can be used to generate binocular disparity, rotating (the perspective of) the monitor along with the head movements creates motion parallax, and moving the monitor back and forth along with the focal point induces convergence and accommodation.

One extension of a monitor is the fish-tank virtual reality system (Ware, Arthur, and Booth, 1993), which is characterized by the stereo image of a three-dimensional environment viewed on a monitor that uses the perspective of the user. Head-mounted displays apply stereoscopy by projecting the views of a virtual left and right eye on two separate screens that cannot be seen by the other eye. The orientation of the virtual eyes is updated along with the rotations of the head. This covers the motion parallax. Surround-screen environments cover the motion parallax by updating the perspective based on the distance from the user to the walls. In the CAVE, stereoscopy was achieved by mixing the two images on a screen and using glasses to filter the correct image per eye. This is done for all screens in the CAVE.

The virtual reality displays mentioned in the preceding paragraph utilize binocular disparity and motion parallax for enhanced depth information. Convergence and accommodation are typically not dynamically utilized by monitors, head-mounted displays or surround-screen systems. Head-mounted displays use lenses that can be replaced by less or more convex lenses, but the shape of the lenses cannot be changed while using the device. The shape of the lens determines the focal point and thus the accommodation of the eyes. Another measure to estimate the accommodation is the zonular tension, which is induced by the muscles of the eye (Toates, 1972). Future virtual reality displays may couple the zonular tension to the projection onto the display, to enhance the immersion by matching the accommodation of the eyes with the projection of information about the virtual environment on the display. In case of an abstract layout of images with no or a gradual change of depth, there is not much accommodation. However, you may actually want a virtual accommodation effect to focus on an image of interest. Fisheye views distort the view to magnify the focus (Buering, Gerken, and Reiterer, 2006) and thus create a similar effect as accommodating to an image. Although distortion at the viewpoint (point of focus) is not a natural source of depth information, it could still contribute to the immersion as it is related to the physical head movements to get information about the virtual environment. Additionally, Witmer and Singer (1998) argued that the sense of presence in a virtual environment depends on the ability to focus on selected information that is meaningful to the user. This thesis believes that the immersion and sense of presence in virtual reality are not bounded by the physics of reality. Rather, one should strive to find an optimal balance between physics of reality for natural interaction and artificial physics that suit the user needs. The experiments described in section 4.2 evaluate whether artificial physics of fisheye view approaches improve the user experience. It is expected in this thesis that there is a user need for enhanced focus and that users quickly become familiar with the artificial fisheye view physics. Section 2.3.4 continues on the topic of fisheye views from a spatial navigation perspective.

2.2.3 Performing actions in virtual reality

In order to interact in a virtual reality, users not only need to perceive information about the virtual environment, but also be able to send information to it. Human actions consist of conceptualizing and executing a set of movements (Goldman, 2015). For example, a set of movements can be performed to press a button. The device reports which button was pressed to the system and the corresponding action is performed in the virtual environment.

A motion-tracking system usually detects a set of movements performed by the user and sends this information to the system. Unless all movements are mapped straight to movements in the virtual environment, this requires some conversion from the detected sets of movements towards actions in the virtual environment. Such a set of movements can vary greatly per human and is influenced by the circumstances (Peters, 2015). Hence, the different sets of movements to detect should be simple and not overlap to avoid a mismatch between the user's intention and the system's detection of the movements.

The motion-tracking systems usually require the user to act recognizably rather than naturally, where the systems with buttons as input mechanism take it one step further and neglect natural actions all together. For virtual actions that are hard to simulate with physical actions or in a confined space, pressing a button might be preferred. If the virtual actions are close to natural actions, motion tracking could enhance the ease of learning and immersion.

2.2.4 Viewpoint-controlled actions in virtual reality

This thesis focuses on two types of actions: Looking at images and navigating through a collection of images. Head-tracking, possibly in combination with eye tracking, seems a more intuitive interaction method for the purpose of looking around than using a device with buttons such as thumb-sticks or a mouse that controls the viewpoint. Although users may be more familiar with these devices, they are not necessarily intuitive for this purpose (Raskin, 1994).

Many approaches to navigation are possible, but spatial navigation methods seem most useful for structured layouts. These methods allow the user to navigate between the images in a structured layout according to their spatial location in the layout. According to Epstein (2008), there are two types of spatial perspectives: observer-centered and world-centered. The observer-centered perspective is specific to the viewpoint of the user, where the world-centered perspective is more concerned with the spatial structure of the environment. The observer-centered perspective of navigating according to the viewpoint is natural to people, as they repeatedly look around to navigate in everyday life. The world-centered perspective of the layout representing the collection of images is beneficial to navigate towards unseen goals (Byrne, Becker, and Burgess, 2007). In the case of a layout with images wrapped around the virtual eyes, the spatial location of the viewpoint can be used to interact with the layout. This remains the physical movements of observer-centered perspective of navigation, while matching it to the navigation information of the layout. Therefore, this thesis focuses on viewpoint-controlled interaction.

2.3 Spatial navigation

This section starts with a review of the spatial navigation technique proposed by Igarashi and Hinckley and follow-up work. In relation to this technique, viewpoint control is discussed. Finally, this section proposes ways to point out the spatial location of the viewpoint to the user.

2.3.1 Spatial navigation techniques

Igarashi and Hinckley (2000) showed a new spatial navigation technique for browsing large documents on a regular monitor. They coupled zooming to the scrolling speed. If the user scrolls rapidly, the view zooms out so that the perceptual scrolling speed remains constant. A threshold scroll speed was used to prevent from unexpected zooming at slow scroll speeds. The usability study of Igarashi and Hinckley indicated that this technique worked best for visually distinct data, because these provided scrolling cues at high scroll speeds.

Cockburn and Savage (2004) showed that this technique also works for browsing through a large text document. Cockburn and Savage used a 157 page masters thesis consisting of nine chapters and two appendices. Igarashi and Hinckley used an alphabetically ordered list of words to evaluate text browsing, skipping words for the higher scroll speed levels. Accordingly, it appears that the technique works for text documents as well, provided that it is possible to zoom out without skipping words, as is the case in documents with chapters and sections by enlarged titles. This likens to the conclusion of Igarashi and Hinckley that the technique works better when the zoomed out view provides scrolling cues.

Igarashi and Hinckley also tested their technique for image browsing. In their image browsing system, the static view is a single image. Scrolling faster, the images become smaller and the horizontal neighbors are visible. Although the results were much better than with simple scrolling, they felt a static grid of thumbnails would be superior for browsing many independent images, because spatial abstraction is difficult when the images are not ordered.

When the images are ordered, spatial abstraction may be easier. One option is to use pyramid-style zoom levels. On the lowest level, all images are displayed. Each level above picks one image per region of the level below that is representative for that region. This requires some selection method to pick the representative images and smooth translations to align the representative images while zooming in and to move them back to their original place when zooming out.

Zooming through such a pyramid-style arrangement of zoom levels with a virtual reality display may induce motion sickness, because of the conflicting motions between the fixed position of the user and the visual motion (Hettinger and Riccio, 1992). Therefore, we assume in this thesis that a fixed position of the user is preferred to movements in the virtual environment. Instead, the scroll speed is coupled to the spatial location of the viewpoint in the layout. A follow-up study could evaluate these side effects when zooming is supported and compare the results with this thesis.

2.3.2 Control of spatial location in the layout

Users can look around the virtual environment by rotating their head. The viewpoint of looking can be used as spatial location in the collection of images. Another option is to point with your finger at the spatial location of interest. However, pointing with your finger causes arm fatigue, especially if the user keeps pointing for a while (Pierce et al., 1997). Rotating your head requires less movement than pointing with your finger, because you do not have to lift your arm to a horizontal stance first. On top of that, head movement latency is lower than arm movement latency (Biguer, Jeannerod, and Prablanc, 1982). This may lead to faster response times and thus less corrective movements. A disadvantage of head-tracking is ambiguity of head movements as they are used for both rotating the view and pointing. This may lead to undesired scrolling when the user only wants to look around. However, Chen, Anderson, and Sohn (2001) showed that there is a strong relationship between the gaze and cursor positions and assume that this is even stronger for graphical applications than for text editors. They tracked the viewpoint and the position of the mouse position in a customized web browser. Only 16 percent of the regions visited by the mouse cursor were not visited by the viewpoint. Therefore, this thesis expects that the spatial location of the viewpoint can be used to control scrolling. In addition, looking around is a natural way to see the part of the environment that neighbors the current view. Extending it so that the environment of images continues to scroll while looking to the side seems easy to learn and thus to become familiar quickly.

2.3.3 Visualization of spatial location

Without visual feedback people tend to undershoot the target (Elliott and Allard, 1985). Carlton (1981) showed that movements with visual control usually consists of an initial aiming phase followed by a correction phase. This can be explained by the fact that the EMG discharges for head, eye and arm movement are discharged synchronously (Biguer, Jeannerod, and Prablanc, 1982). Under normal circumstances with visual control, the second phase kicks in when the eye detects that the target is not reached yet. It seems that this second phase does not get activated without visual feedback, since the eye does not detect any information to update the position of the arm.

While scrolling, a visualization of the spatial location of the viewpoint may prevent the user from undershooting the target, because the user sees whether the viewpoint is on the target or not. Also, it may be interesting to highlight the images at the spatial location of the viewpoint. Instead of zooming with the camera position as was done by Igarashi and Hinckley, the image at the spatial location of the viewpoint could come closer. This way the virtual environment remains at the same position except for one image.

Cursor

A cursor can be displayed at the spatial location of the viewpoint. One of the issues with cursors is that they may have the same color as the background of an image. This can be solved with tools to adjust the color contrast automatically (Bates and Day, 2004). In this thesis, the cursor is meant to show the spatial location and thus the image behind the cursor. Hence, the user should be able to know the position of the cursor, but the cursor itself should not be distracting. Moreover, the cursor partly obscures the image which could reduce the experience of beholding. A benefit of having a cursor is that can be used for other feedback as well. For instance, a progress circle that shows how long the user has been looking uninterrupted at an image. This can be used to visualize delay before moving the image closer.

2.3.4 Fisheye

According to Furnas (1986), fisheye views provide a balance of local detail and global context. This seems useful for a global context-based structured collection of detailed images. Schaffer et al. (1996) compared fisheye views against traditional zoom views for hierarchically clustered networks. Their results showed that the users were able to navigate more quickly and needed less exploration to traverse the structure. Hollands et al. (1989) compared graphical fisheye views with scrolling views of networks and found only a slight favor of fisheye views. Sarkar and Brown (1994) pointed out that the fisheye view of Hollands et al. moved the focal point directly to the center causing disorientation. This diminishes the benefit of keeping the global context in the background, because the context changes abruptly on choosing a new focal point.

Liu et al. (2004) compared a slider with a fisheye view for image browsing through a set of images ordered on similarity. The slider ranged from 0 to 1, where the images were placed at their exact position in the similarity space without regard for overlap when the slider was set to 1 and placed in a grid without overlap when the slider was set to 0. The fisheye view used a distorted polar coordinate system with a distortion rate of 0.5. Liu et al. included a figure that shows the fisheye view next to the grid without overlap. The center image is enlarged and the images outside the fisheye remain the same. This resulted in quite a lot of empty space around the center and overlap around the edges of the fisheye. For fisheye views of networks, this is desired to highlight the selected node and direct neighbors without to much clutter of arcs, but it seems useless and distracting to create empty space around the images.

2.4 Summary

In summary of related work, we list the sub-questions that are investigated during the experiment in order to answer the research question.

- *How does the layout affect the viewpoint-controlled navigation method and visualizations?*

This thesis conceptualizes three kinds of context arrangement: A layout of ordered numbers and colors, a layout of random images, and a layout of images with the tag 'Fireworks'. This thesis hypothesizes that people prefer the layout with numbers and colors to find a target quickly, because of their knowledge of the ordering. On the other hand, they are expected to enjoy the layouts of images more, because they can discover new images. The layout of images with the tag 'Fireworks' is expected to be appreciated the most, because it allows for new discoveries by looking around, while the fact that they share a tag gives some guidance on what kind of images can be discovered in that area.

- *Does viewpoint-controlled navigation in a virtual environment with a layout surrounding the user feel intuitive?*

Many approaches to navigation are possible, but spatial navigation methods seems most useful for structured layouts. These methods allow the user to navigate between the images in a structured layout according to their spatial location in the layout. According to Epstein (2008), there are two types of spatial perspectives: observer-centered and world-centered. The observer-centered perspective is specific to the viewpoint of the user, where the world-centered perspective is more concerned with the spatial structure of the environment. The observer-centered perspective of navigating according to the viewpoint is natural to people, as they repeatedly look around to navigate in everyday life. The world-centered perspective of the layout representing the collection of images is beneficial to navigate towards unseen goals (Byrne, Becker, and Burgess, 2007). In the case of a layout with images wrapped around the virtual eyes, the spatial location of the viewpoint can be used to interact with the layout. This remains the physical movements of observer-centered perspective of navigation, while matching it to the navigation information of the layout. Therefore, this thesis focuses on viewpoint-controlled interaction.

- *What is the effect of visualization of the viewpoint on user experience?*

This thesis considers a fisheye view that uses depth to move the center image closer instead of enlarging it and avoids empty spaces by computing the distances between the center image and the edges of fisheye view. Section 3.2.2 explains the implementation of this visualization of the fisheye view. A similar effect can also be achieved with a fisheye view approach that is closer to the physics of reality. This second fisheye view approach consists of a sphere with the same size and height as the first approach. Instead of moving the images in the fisheye view closer, they are projected onto the sphere as if it was a magnifying glass. Both are evaluated and compared to the absence of a fisheye view to find out whether artificial physics are desired for spatial navigation in virtual reality.

Chapter 3

Implementation

This chapter addresses the decisions made regarding the implementation of the virtual environment and the materials used to evaluate spatial navigation in virtual reality. The first section treats the choice of hardware, selected image collection and access to the images. The second section is concerned with the construction of the environment and visualization of the viewpoint.

3.1 Materials

The materials required for this thesis are a virtual reality display, an image collection, and the tools to create an interface for this collection. The following subsections discuss the considerations with regard to these materials.

3.1.1 Hardware and development tools

The head-mounted display used for the experiments is the Oculus Rift Development Kit 2 (with SDK version 1.3). The resolution of this device is 960 by 1080 pixels per eye. The two virtual reality displays that were available for this thesis are the Oculus Rift Development Kit 2 and a Samsung Galaxy s7 edge placed in a shell that can be mounted to the head. Despite the higher resolution of the Galaxy s7 (1280 by 1440 pixels per eye), the Development Kit 2 was used during the experiments for performance reasons. An HP Z440 desktop (Intel Xeon E5-1620v3 @ 3.5 GHz; 16GB RAM; Nvidia Geforce GTX 970) was used during the experiments with the Oculus Rift Development Kit 2.

The application is written in C# with Microsoft Visual Studio Enterprise 2015 (version 14.0.23107.0) as editor and Unity (version 5.4.1) for creating the virtual environment and managing the head-mounted display. It was estimated that 53 percent of games for the Oculus Rift was made with Unity when it launched (Unity Technologies, 2016). From the top 1000 of free mobile games, 41 percent was built in-house and 34 percent was made with Unity. The large market share is accompanied by a large community with over five million registered developers and 222,000 questions of which 147,000 were answered

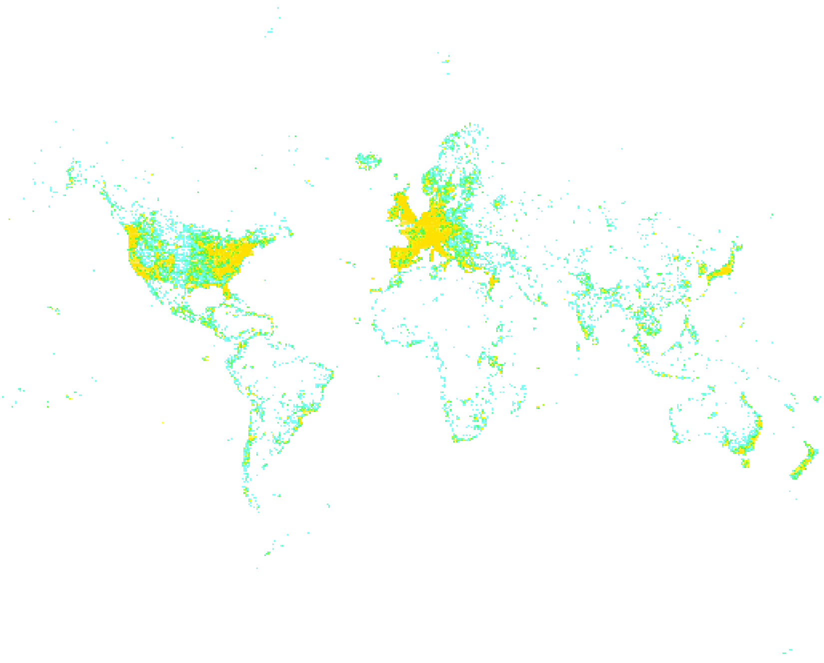


FIGURE 3.1: Heat map of the YFCC100M subset with geo-location. The density increases along with the color from light blue to yellow. Mercator projection is used to map the coordinates in 2D.

(Graphics_Dev, 2016). Visual Studio is used as text editor to write C# code because of personal preference.

3.1.2 Image collection

The images are retrieved from the Yahoo Flickr Creative Commons 100 Million (YFCC100M) data set. In 2015 it was the largest open-access collection of photos and videos (Bernd et al., 2015). At the default pixel resolution used on Flickr, the photos take up 13.5TB and the videos 3.0TB (Thomee et al., 2016). The subset of images used in this thesis is limited to only those with a specified geo-location. The location attribute of the images was considered as a way to visualize the image collection, but this visualization was put on hold for this thesis.

The resulting subset contains 18,662,919 images, most of which were taken in North America, Europe and coastal areas. Figure 3.1 shows a heat map of these locations. The size of the database of Flickr farm URLs and corresponding meta-data is slightly over ten gigabytes. The images itself are not stored locally. Section 3.1.3 dwells on the storage and access of the images.

3.1.3 Accessing the images from the collection

The subset of nearly 18.7 million images composes almost one fifth of the complete data set. Therefore, storing the complete subset locally would take up approximately 2.7TB. Although this storage requirement is manageable, the number of images could be problematic, especially when the image collection is dynamic and thus supports modification, insertion and deletion of images. Cloud storage systems are designed to be extremely scalable and easy to manage (Wu et al., 2010) and thus often used to process very large data sets (Nurmi et al., 2009) (Gu and Grossman, 2009) (Alonso-Calvo et al., 2010). Therefore, these systems seem more suitable to store a dynamic collection of images than local storage systems. Undoubtedly, the Flickr servers are also optimized to store a dynamic collection of images. Using these servers for the experiments in this thesis has two main advantages: it saves the time and resources to manage cloud storage and second, it models a practical scenario of having a large image collection better than storing a static data set locally.

A small performance test was done to evaluate the speed of local and remote access of images. Figure 3.2 shows that accessing the images from the Flickr servers is not significantly slower than loading them from an Solid State Drive (SSD). The experiment was performed with a Samsung 850 EVO 250GB and 'fast ethernet' (100Mb) network. The set of local images was a collection of 10,000 images in a single folder on the SSD, which was retrieved by querying the first 10,000 images from the local database. The same part of the local database was used to query the images from the Flickr servers. According to a CrystalDiskMark test with 5 runs of 500MB with block sizes of 128KB, the sequential reading speed of the SSD was 565.6MBps. Most of the images were in the range of 50KB to 200KB. The internet speed was tested with 5 runs of Ookla Speedtest and showed an average of 9.89MBps. The similar loading times indicate that the storage location is not the bottleneck of loading many images at once. A possible explanation is that the images are loaded separately and thus the amount of images that can be loaded per second may have a lower total size than both transmission media support.

The local database contained the meta data and location on the Flickr servers of the images. It was only used for simple queries on one facet of meta data at a time and thus there was no need for a comprehensive database system. SQLite (version 3.10.2.0) was used as it is lightweight and convenient to construct an embedded database system (Bi, 2009).

3.2 Application

An overview of the system is illustrated by Figure 3.3. The input controller updates the spatial location of the viewpoint and writes

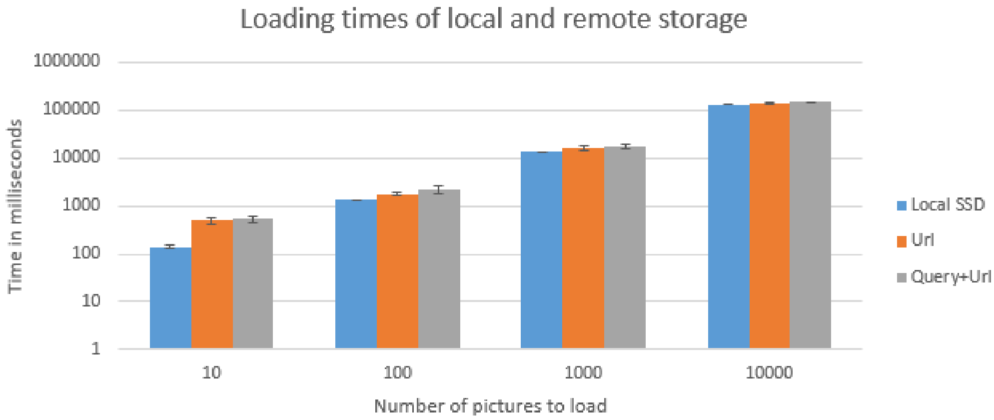


FIGURE 3.2: Loading times in milliseconds of local and remote storage on a logarithmic scale.

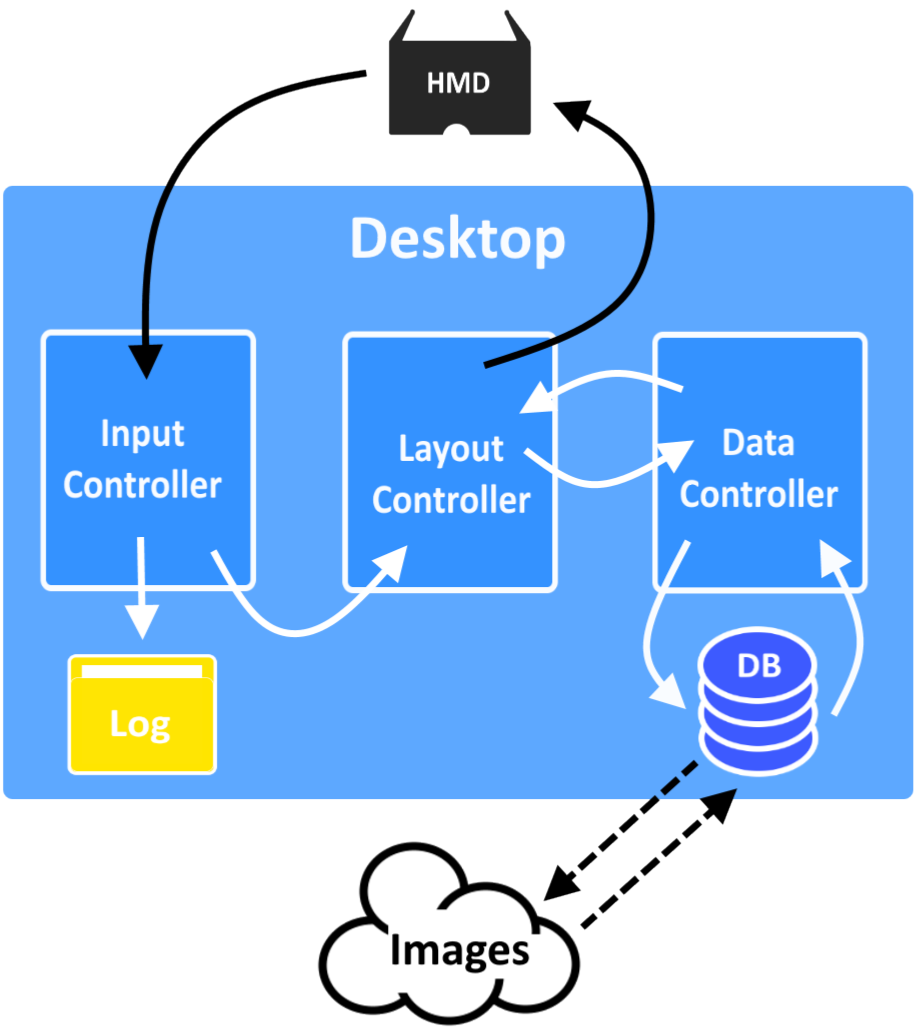


FIGURE 3.3: Overview of the system.

these spatial locations to a log file. The data controller connects to the local database and is responsible for executing the queries and loading the images. The layout controller constructs the layout in the environment. This section describes the implementation of the layout, the scope of the implementation in this thesis and alternative implementations that could be interesting for future work, and the visualization of spatial location of the viewpoint.

3.2.1 Construction of the layout

In this thesis the layout is wrapped around the user's eyes. Because a curve in both horizontal and vertical direction is not possible without skewing the shape of the images, the layout is wrapped around the user horizontally. Horizontal head rotation felt more comfortable than vertical rotations, according to personal experience. An explanation for this preference could be that most objects of interest in everyday life are at eye level and thus humans are less accustomed to rotating the head vertically. To create the illusion of an infinite layout, the sides are extended in the same direction as the last image from the curved area.

Figure 3.4 shows an overview of the layout with numbers and colors. The numbers are ordered from left to right with a step size of 1 and the colors from the top to the bottom based on hue. Saturation and value are set to 1. The step size of hue was determined by the number of columns in the layout. The layout of images with the tag 'Fireworks' is presented in Figure 3.5 and retrieved by the query $select\ *\ from\ YFCC100M\ where\ tags\ like\ '\%fireworks\%'$. The layout with random images was retrieved by a query without the where clause and is displayed in figure 3.6.

Parameters of the layout

According to the best practice guide of Oculus VR (2016), one unit in Unity corresponds roughly to one meter in the real-world. From now, this thesis expresses the virtual distances in meters as well. The virtual eyes of the user are placed at 9 meter away from the images with a size of 1 by 1 meter. The layout has a height of 60 meter and width of slightly more than 88 meter (60 meter extended to the side and 9π meter circumference of the curved area). The angle of the curve is 180 degrees, to keep a constant radius of 9 meter. The radius and distance values were set by personal preference. Future work could be directed at evaluating the effect of other values, possibly in combination with other angles of the curve. With the same distance of the virtual eyes and a larger radius or wider angle of the curve, the images to the sides are further away than those in the center. Increasing both the distance of the virtual eyes and radius results in a larger overview of the images. Decreasing these values results in images

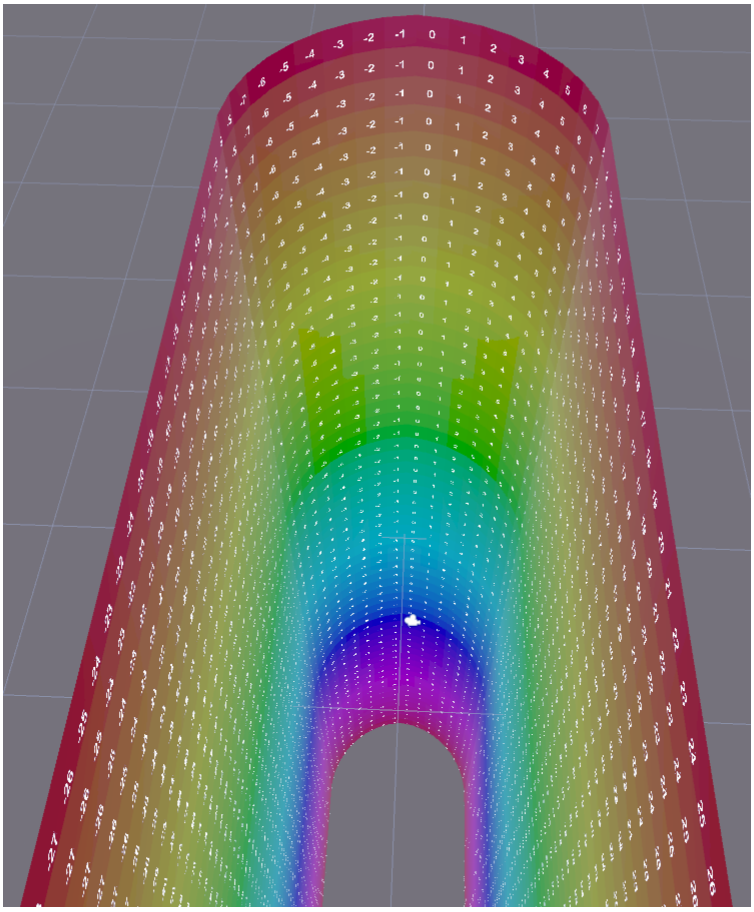


FIGURE 3.4: Layout of numbers and colors.

FIGURE 3.5: Layout of images with the tag 'Fireworks'.

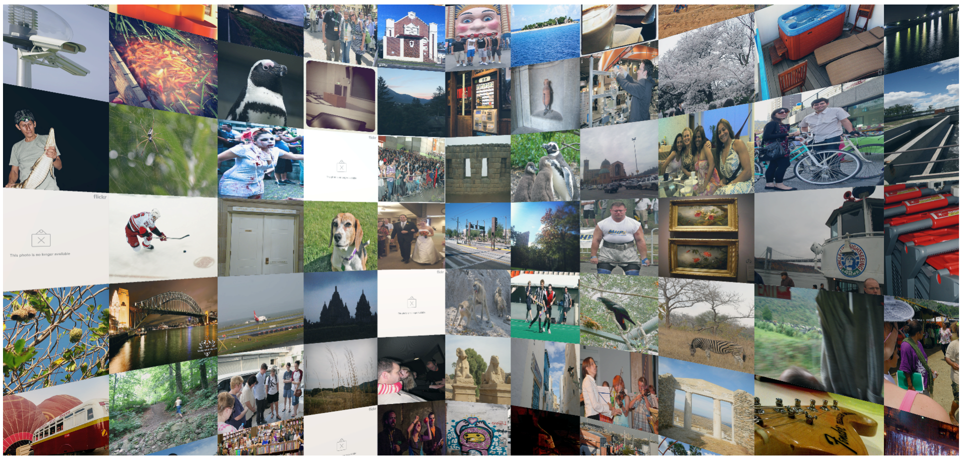


FIGURE 3.6: View of the virtual eyes on the layout of random images.

that are closer to the eyes and thus perhaps a more detailed view. A variant of the spatial navigation technique proposed by Igarashi and Hinckley with a zooming camera could take advantage of these scenarios, but at the cost of zooming motion. It would be interesting to evaluate such an automatic zooming technique in a future study.

Construction of the curve

The curve is constructed in a few steps. First the number of columns in the layout that are curved is computed by: $number\ of\ columns = \frac{camera\ distance \times \pi}{image\ size}$ and the angle of the curve is divided by the number of columns to get the additional angle per image defined as α in this thesis. Second the images are positioned from center to the sides. We define x as the number of images from the center. If the number of columns is an odd number x starts at 0 and is used as starting point for both the left and right side of the curve, otherwise there is no center image, so the left and right images closest by the center start at x is 1. The rotation of the individual images is computed by $rotation = x \times \alpha$.

Figure 3.7 illustrates the geometry behind the computation of depth and width. The depth of image $k + 1$ is the depth of image k minus $\sin(\alpha') + \sin(\alpha'') \times \frac{image\ size}{2}$, because γ' equals α', β'' equals α'', and the lengths of the hypotenuses are both half the image size. In both cases, depth is the opposite side of a given angle and thus computed by the multiple of the sine of that angle and the known hypotenuse. The width is computed in a similar manner, but with the cosine to get the length of the adjacent side. Figure 3.8 displays the resulting curved layout.

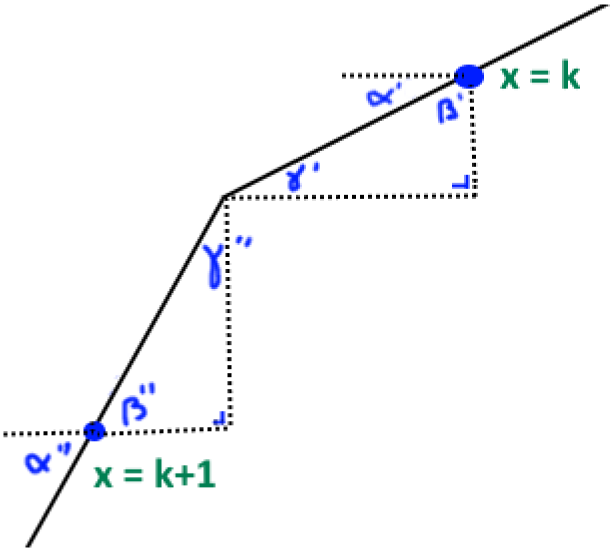


FIGURE 3.7: The geometry behind the construction of the curve.

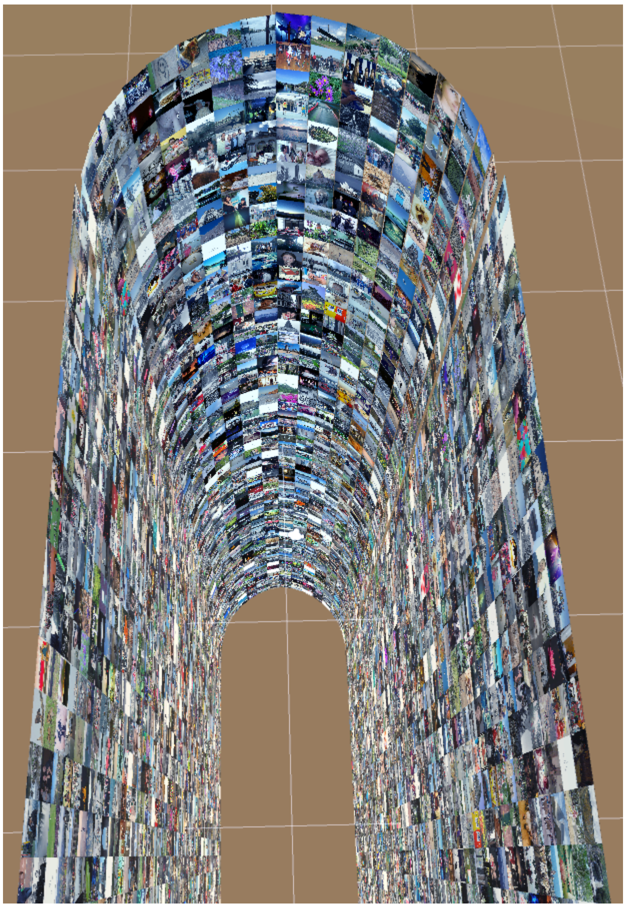


FIGURE 3.8: Overview of the curved layout.

3.2.2 Spatial navigation

To reach all images, one should be able to navigate in all dimensions of this image space. Many approaches to navigation are possible, but we considered spatial navigation methods most suitable for viewpoint interaction (section 1.2.2). These methods allow the user to navigate between the images in a structured layout according to their spatial location in the layout. To identify the viewpoint, a ray is cast from the virtual eyes into the direction of the view. In this thesis the viewpoint is defined as the intersection between the ray and an image. From now on, the image that is intersected by the ray will be called the center image. A cursor and a fisheye view are implemented to visualize the viewpoint.

Cursor

If someone looks at a particular image for a while, we assume that this person is interested in that particular image and thus it might be appreciated if it is enlarged, that is moved towards the user. Knowledge about the spatial location of the viewpoint could be helpful to be sure that the user and the system are considering the same point of interest. Additionally, a progress indicator could be beneficial to ensure the user that the system is waiting for confirmation instead of being unresponsive. Therefore, this thesis first tries to identify what configuration of required time to look at a particular image, knowledge about the location, and progress indicator is preferred, before comparing this visualization with the fisheye view.

The required time period of looking at an image before moving the image towards the user is handled by the input controller using the c# Stopwatch class. If the stopwatch is not running yet or the viewpoint changes from one image onto another, the stopwatch is reset and started over. The input controller request the ElapsedMilliseconds property of an instance of the Stopwatch class to see if the timer has reached a specified confirmation time. ElapsedMilliseconds is ant 64-bit integer and thus rounds the elapsed time down to the nearest whole millisecond value (Microsoft, 2016). Given that the time periods are also defined in whole milliseconds, this should not be a problem.

While the stopwatch has not reached the time period appointed by the configuration and a progress indicator is present, the progress indicator is updated, otherwise nothing happens. In this thesis the progress indicator is combined with a circle shaped cursor to visualize the spatial location. The cursor has an open center and consists of a foreground and background circle. The background circle is half transparent. The foreground circle starts at the top with full transparency which gradually increases in a clockwise direction to reach no transparency back at the top. This is used to create the effect of a progress circle by changing the transparency threshold. In figure 3.9

the threshold is just over half transparency of the foreground circle. Many other color schemes and shapes are possible, but the influence of color and shape on user experience was expected to be meager. Results of the user study may indicate otherwise. In that case, it may be worthwhile for future work to evaluate the effect of shapes and color schemes.

After the stopwatch has reached the appointed time period, the image is rotated and moved in the opposite direction of the ray from the virtual eyes into the direction of the view. This movement occurs with a distance of 0.15 meter in each update step, until it reaches a maximum distance from its original position of 1.5 meter. When the user looks away, the previously selected images are moved back with the same speed and their original rotation is restored. A new image can be selected while there are still images moving back. The rotation in the direction of the virtual eyes is implemented for the images that are not facing the virtual eyes directly, such as the images at the top and bottom of the view or, in case the radius of the curve does not match the distance of the virtual eyes or has another angle than 180 degrees, all images except for the center image. The movement speed was set to 0.15 meter per update to create a fast, but visible motion effect. The exact value is based on personal preference, as the purpose of this movement is just to avoid an abrupt change of position of the image. The maximum distance from the original position was set to 1.5 meters, to avoid that the image of interest overlaps its neighbors. A follow-up study could evaluate alternative solutions such as a dependence between the proximity of the image to the user and the proximity of the viewpoint to the center of that image. In other words, when the viewpoint is close to the center, the image is moved towards the user and when the viewpoint drifts to the side, the image is automatically moved backwards according to the deviation from the center.

Fisheye view

The fisheye view is designed to balance the local detail of the images and the global context. The images around the viewpoint are moved closer to show more detail while the global context remains. In this thesis, the images outside the fisheye view are unaffected and depth is used to increase local detail instead of a distorted two-dimensional coordinate system with large areas between consecutive whole coordinates in the center and smaller areas between the coordinates near the edges. Using depth to increase local detail seems more natural to enlarge the images in that area and avoids the need to make room for the larger center images as the images are not enlarged and do not overlap that much, unless the depth is high compared to the size of the fisheye and thus the difference in depth between the center image and its direct neighbors is large.

FIGURE 3.9: The cursor with progress circle.

The previous section discussed the maximum distance of 1.5 meter from the original position to avoid overlap. To maintain equality, the center image of the fisheye view is also positioned 1.5 meter closer to the virtual eyes. In contrast to the cursor configurations, the neighbors are moved closer as well and there is no time period before the fisheye view is realized. Figure 3.10 illustrates how the neighbors are partitioned. The center image is marked green and has a distance from its original position of 1.5 meter. The neighbors that are horizontally or vertically adjacent to the center are marked yellow and have a distance of 1.0 meter. The values of x and y are non-negative, because a single pass over half the stack size is used for both directions. The diagonally adjacent neighbors are marked red and also have a distance of 1.0 meter, in order to create a pyramid-style arrangement of depth levels as displayed in figure 3.11. The stack size was set to 6 by 6 images, to cover slightly less than the vision of the fovea and remaining a roughly equal area of unaffected images in the peripheral vision when looking directly at the center image. The unaffected images can also be seen with the vision on the fovea by looking around with the eyes while keeping the head steady. Figure 3.12 shows the fisheye view with a layout of images from the perspective of the user.

3.2.3 Scrolling

It is our assumption that scrolling is a very beneficial component of viewpoint-controlled interaction with a virtual environment, because scrolling actions can be performed while keeping track of the objects. For example, by looking to the right, the images at the right side start to move towards the middle and thus the user interacts with the environment by physical motion. This has the potential to

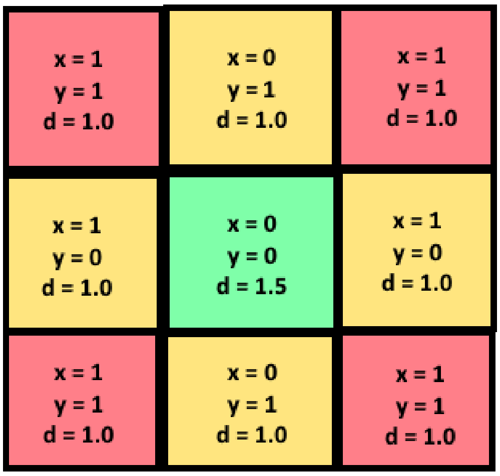


FIGURE 3.10: Illustration of the partitions inside the fisheye view with a maximum depth of 1.5 meter and stack size of 6. Only the center image and its adjacent neighbors are displayed.

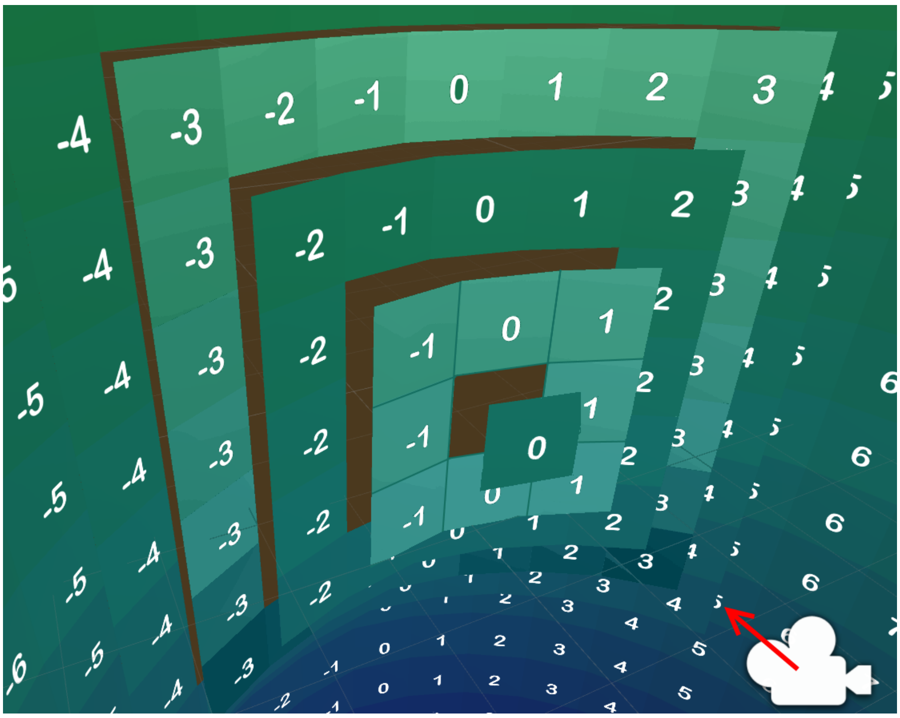


FIGURE 3.11: Fisheye view with a layout of colored squares with numbers.

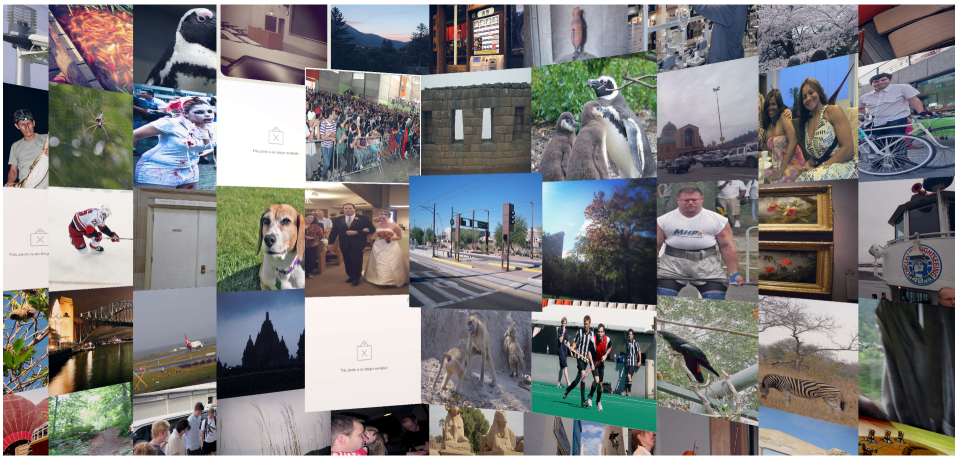


FIGURE 3.12: Fisheye view with a layout of images.

enhance the immersion significantly. However, this effect may carry too far, resulting in visually induced motion sickness (Hettinger and Riccio, 1992) and vection (Keshavarz et al., 2015). Especially vection is quite likely due to the opposite movements of the head rotation and the environment. Although vection is often a desired phenomenon in realistic environments, it seems unlikely that it is desired in an abstract environment, because there is no real-world knowledge to explain the vection.

To avoid vection, both the user's virtual standpoint and the position of the images are not changed. Instead, the images are replaced by the image of a neighbor. To ensure that the layout is not constantly updating, a threshold value of 0.25 was chosen in both horizontal and vertical direction. The rotation values range from 0, directly in front of the user, to 1 and -1 at 90 degrees from the forward direction of the user. The y value of the normalized vector in the direction of the view was used as rotation value. The threshold value was subtracted from the rotation value and the result was divided by 1 minus the threshold value, to map the values between threshold value and 1 to the range of 0 to 1. The resulting value was squared. In case of the horizontal scrolling and looking down it was divided by 2, because it is harder to look up than in the other directions. The results were added to a sum variable and the update was executed when the sum variable exceeded 10. These values were set based on personal preference. The quadratic equation was chosen to start with slow scroll speed for small rotations and rapidly increasing scroll speeds when you look more to the side.

Chapter 4

Evaluation

As stated in the introduction, the research purpose of this thesis is to explore spatial navigation methods in virtual reality for exploratory image browsing. To do so, we evaluate the user experience of viewpoint visualizations, how the layout affects the user experience of viewpoint-controlled navigation with these visualizations, and whether viewpoint-controlled navigation in a virtual environment with a layout surrounding the user feels intuitive. Thus, our dependent variables are user experience and intuitiveness. Section 4.1 assesses user experience measures for both virtual reality and image browsing experiences in order to find a good balance to evaluate the user experience in this thesis.

Furthermore, this chapter covers the research methods and procedure of the experiments. After stating the measures of user experience, the procedure of the experiment is described with the tasks to be completed for each configuration. The term configuration is used to describe a single iteration of the experiment with a specified set of parameters that belong to that configuration. Furthermore, the procedure contains a section for each experiment to discuss the execution of the informal interviews during the experiments.

4.1 User experience measures

Law et al. (2009) performed a survey on user experience and found that researchers and practitioners agree on a concept of user experience as dynamic, context-dependent and subjective due to a broad range of potential benefits users may derive from the system. In a follow-up survey, Vermeeren et al. (2010) collected 96 methods to identify the needs and further research questions on user experience evaluation methods. They believe that usability is subsumed by user experience, where motivation and expectation of the user play a stronger role in user experience than in usability. Therefore, this thesis aims to take the context-dependent motivation and expectation into account during the evaluation.

4.1.1 User experience in virtual reality

Virtual reality may have a significant impact on the *motivation* and *expectation* of the user and thus it should be considered in the evaluation of user experience. Witmer and Singer (1998) proposed a Presence Questionnaire (PQ) and an Immersive Tendencies Questionnaire (ITQ) to evaluate presence in virtual environments. Slater (1999) argued that the PQ does not measure presence, but the response of people to various aspects of a system that are correlated with other measures of presence such as involvement and immersion. Additionally, research on presence is often designed to identify the sense or cause of presence, rather than interaction and interactivity with the virtual environment (Schuemie et al., 2001) Kober and Neuper (2013). An explanation for the lack of measures for interaction and interactivity in virtual environments could be the need for artificial physics, which do not cohere with the sense of presence as defined in current presence questionnaires. It is doubtful that a virtual environment with only real-world physics is desired, because augmented reality seems more suitable for such an occasion. Instead of presence, the evaluation in this thesis aims to identify how the artificial physics of the fisheye view affect the motivation and cohere with the expectation of the user. The ITQ was developed to measure the capability or tendency of individuals to be involved or immersed (Witmer and Singer, 1998) and thus it is targeted at the psychological characteristics of the user. Knowledge about these characteristics could be useful to identify the motivation and expectation of the participants.

4.1.2 User experience of exploratory image browsing

Khanwalkar, Balakrishna, and Jain (2016) evaluated the potential of image exploration in virtual reality by asking the participants to rate the satisfaction of legacy image search using keywords in text boxes, the experience and intuitiveness of their navigation approach and the infinite virtual space, and whether the participants see themselves use a virtual reality application for image browsing in the future. Although these questions gave some insight in the user experience of the system, the answers depended on the user's interpretation of user experience and did not clarify why the experience was positive or negative. Furthermore, they did not look into the motivation and expectation of the user. Instead, the participants were asked to explore images on Flickr, Instagram and Pinterest on a computer or smart phone, followed by a 10 minute section of exploring their system for image browsing in virtual reality. Marchionini (2006) consider learning and investigation as motivation for exploratory search and argues that the web has legitimized browsing strategies that depend on selection, navigation and trial-and-error tactics resulting in a trend towards more active engagement in the search process.

4.1.3 Usability of viewpoint-controlled navigation

(Kaufmann and Dünser, 2007) used a selection of questions from the web-based usability questionnaires arranged by Perlman (2015) for the evaluation of geometry education in augmented reality with a head-mounted display. Jeon, Shim, and Kim (2006) evaluated the usability of viewpoint interaction for augmented reality on a desktop with four general questions about convenience, task difficulty, fatigue and overall experience. A selection of web-based usability questionnaires may lack the usability with regard to the exploratory browsing intentions. A few general questions on the other hand may cover these intentions, but depend on the user's interpretation and does not explore what these intentions are. Hence, an extended usability questionnaire with questions about the exploratory browsing intentions seems most suitable to evaluate our viewpoint-controlled navigation method for exploratory image browsing. Lund (2001) developed the USE questionnaire to measure subjective aspects of the usability rather than performance measures. USE stands for Usefulness, Satisfaction and Ease of use. These aspects cover a large part of the user experience of viewpoint-controlled navigation. However, the questions to identify the usefulness are still too general to figure out which characteristics of viewpoint-controlled navigation are useful and for which intent.

The first four statements about Usefulness were removed, because they depended on the user interpretation of effectivity, productivity, usefulness and control over the activities in one's life. The remaining statements are related to the expectation of the user. The removed statements are replaced by the following statements:

- *It is useful to scroll through a set of images*
- *It is useful to find an image quickly*
- *It is useful to explore many images*
- *It is useful to discover new images*

The satisfaction part of the questionnaire is extended with the statements:

- *I feel more engaged with the collection of images*
- *It stays interesting over time*
- *I was more aware of the images around me*

The new usefulness statements aim to determine the usefulness for different search intents. The new satisfaction statements are presented to the participants to evaluate the claims about peripheral awareness and enhanced engagement of virtual reality that were made in the introduction. The complete set of usability questions is included in appendix section A.2.

4.2 Procedure of the experiment

Each participant was seated in a rotatable chair and asked to fill out the Immersive Tendencies Questionnaire. On completion, the participant was asked to face the head-tracking sensor and put on the head-mounted display when ready. The examiner made sure that the wires of the head-mounted display did not tangle up and helped to adjust the distance of the display and tightness of the straps, especially for the participants with glasses. All participants started with a small test to get familiar with viewpoint-controlled navigation. The participants were asked to look around and express their initial reactions. After the participants figured out that they can scroll by rotating their head, they were asked to navigate to a couple of numbers and colors chosen by the examiner. The participants were requested to express any unexpected or undesired behavior during the navigation and afterwards they were asked to explain their search strategies. The first part of section 4.2.2 describes these navigation tasks. In the meantime of these first tasks, the examiner inquired about the shape of the layout and whether or not they found it practical and intuitive to navigate with head movements.

After completing the first part of the test, the participants were asked whether they experience motion sickness or any other discomfort that required a break. This was followed by three configurations with a different visualization of the viewpoint. Namely a cursor, a fisheye view created by stacking the images in depth, and a fisheye view conceived by projecting the viewpoint on a sphere. The order of these three configurations was balanced to create three groups of participants with a different order. Each configuration occurs once in each position. Table 4.1 shows the arrangement of these configurations along with the arrangement of the layouts over the participants. The configurations existed of three different visualizations of the spatial location of the viewpoint. For each configuration, the participants were again requested to express any unexpected or undesired behavior and if they noticed any differences in their search experience.

Upon the completion of the viewpoint experiment with numbers and colors, the participants were asked again whether they experience motion sickness or any other discomfort that required a break. Before continuing to the next experiment, the examiner inquired the participants about their current usage of image browsing systems, their needs to use image browsing systems, the limitations of those systems to satisfy their needs, and whether they could envision an image browsing system that caters to their needs if there were no technical limitations. The term image browsing system was clarified by stating examples such as Google Images, Flickr, and (personal) photo albums.

	Navigation			Layout	
Participants	Cursor	Stack	Lens	Random	Fireworks
A	1	2	3	1	2
B	1	2	3	2	1
C	1	2	3	2	1
D	1	2	3	1	2
E	2	3	1	1	2
F	2	3	1	2	1
G	2	3	1	2	1
H	2	3	1	1	2
I	3	1	2	1	2
J	3	1	2	2	1
K	3	1	2	2	1
L	3	1	2	1	2

FIGURE 4.1: Participant design over the viewpoint visualization and layout configurations.

Next, the viewpoint experiment was repeated, this time with images. In one configuration it was a random subset of images from the collection and in the other configuration it was a random set of images with the tag 'Fireworks'. These two configurations were balanced to distribute the possible learning and repetition effects of the simulated exploratory search tasks to be performed over both configurations. The simulated exploratory search tasks are described in the second part of section 4.2.2. Again, the participants were requested to express any unexpected or undesired behavior and if they noticed any differences in their search experience. Additionally, they were asked whether the contrast and detail of images affected the experience of the visualizations and navigation in general.

After the experiment was completed, the participants were asked to take off the head-mounted display and fill out the adapted USE questionnaire. They were urged to motivate their agreement with a statement if it was substantially different for at least one of the viewpoint or layout configurations.

4.2.1 Participants

The interviews of eleven participants were conducted in Dutch, because Dutch was the first language of these participants and the examiner. One interview was conducted in English, because the participant did not speak Dutch and both the participant and the examiner spoke English as second language. The sample consisted of nine male and three female participants. Four of the participants wore glasses, two wore contacts, two had laser eye surgery, and four had normal vision from birth. Nine participants followed a master track of which five in computing science, two in medical science, one in chemistry, and one in art history. Of the remaining three participants, one was enrolled in the bachelor program nutrition and health, one worked as a journalist at a local newspaper, and one participant worked in (special) education. The latter had been diagnosed with epilepsy. This required a more careful conduction of the experiment and the participant was asked about motion sickness or other discomfort more frequently than the other participants.

4.2.2 Tasks

In all configurations, the participants were instructed to find an object that met their interpretation of the task given by the examiner. This could be a concrete objective such as "Find a blue square with the number 50" or more exploratory as in "Look for a picture of a park". The goal of these tasks was to simulate the need for navigation through the collection and gain an insight in the experience of this type of navigation.

Numbers and colors

The participants were instructed to find a number in the range of 20 to 50 away from the current number in the center and a distinct color (red, green, blue). Since completion time and spatial distance were not evaluated, the examiner picked the numbers and colors arbitrarily based on the intention that the participants needed scanning movements to find the right direction and more precise movements to find the assigned number and color.

Images

The image searching tasks were more open. In case of the fireworks images, the participants were asked to find images with a single color (red, green, blue, gold) or multi-color. Other characteristics such as special effects (fish, mine, spinner, palm) were also pursued, but these were hard to explain to most participants and did not occur that often in the subset of the image collection, because it mainly consisted of images taken at professional firework displays where they usually fire shells with brocade effects.

The random images were quite diverse, but there were a couple of reoccurring themes, such as parks, buildings, lakes/beaches, people and animals. These themes were used to instruct the participants with a clear but broad objective, where the target image was chosen by the participants themselves by exploring the collection and selecting the image that fulfilled their interpretation of the objective the most.

4.3 Analysis of the experiment

During the experiments, the examiner wrote down the keywords from the comments of the participants and recorded the interviews in case the keywords were not sufficient to retrieve the intention of the comments. The purpose of the interview was to identify user experience of viewpoint-controlled interaction methods in virtual reality and how these methods affect the user experience of exploratory image browsing. A mixture of directed questions such as "Is the [viewpoint visualization] helping you to focus?" and undirected questions such as "What do you think about the images surrounding you?" were asked. Instead of [viewpoint visualization], the terms cursor, lens, and stack were used. Some participants preferred the term sphere over lens or pyramid instead of stack and in these cases their preferred terms were used in the interview. The participants were asked for each of these different types of visualizations to compare it with the scenario without viewpoint visualization and the visualizations that they had experienced before.

Furthermore, the interview was also used to identify the browsing strategies of the participants and whether they adapted their strategies according to the layout. Questions such as "What is in your current focus?", "In which direction do you look during the navigation?", and "How do you navigate? Do you mainly use knowledge of the structure or do you mainly use the passing numbers and colors to determine the direction of navigation?" were used to identify the browsing strategies in the environment with numbers and colors. In the environments with images, the participants were asked how the lack of structure affected their browsing strategies and whether the higher similarity between the images in the fireworks collection had an impact. Participants who started with the random collection of images were asked to compare their browsing experiences in the fireworks collection with the random collection and (if they did not already mention it) whether the lower contrast between the images affected navigation in general and the visualizations. The other half of the participants started with the fireworks collection and thus were asked to describe their experience of the low contrast between the images. When these participants were browsing the random collection of images, they were asked whether they felt that the arbitrariness of the collection influenced their browsing strategies.

Data of the questionnaires was checked for normality by Sharpido-Wilk tests and ANOVA was used to determine significant effects between subgroups and factors. Correlation was computed between the categories within the questionnaires and frequency bar charts were used to visualize the distribution of the questions and factors with a not normal distribution. Chapter 5 shows the results of these analyses and discusses the realization of the subgroups and factors to analyze.

Chapter 5

Results

This chapter presents the results of the questionnaires and interviews. The Sharpido-Wilk test was used to find out which questions and sub-scales led to a normal distribution of the scores and which did not. One- and two-way ANOVA was used on the questions and sub-scales to spot significant differences between subgroups with a certain characteristic and a normal distribution within the subgroup. The commonly used α-value of 0.05 was chosen to determine whether the null hypotheses should have been rejected. The exact levels of significance (the p-values) are also included in this chapter, in support of the idea of Fisher (1956) that the level of significance is a property of the data, rather than a property of the test (Gigerenzer, 2004). Furthermore, the data in this thesis is primarily subjective and thus it seems wise to interpret the results in their context, rather than blindly trusting the test.

5.1 Immersive tendency scores

Figure 5.1 displays the average scores per sub-scale and the total average of ITQ scores. Participants rated the focus questions on average with a 4.8 and the involvement questions with a 4.1. The virtual reality question (4.3) scored higher than the two gaming questions on average (3.8), although it should be noted that the virtual reality question was expressed to identify frequency of use whereas one of the gaming questions was concerned with immersion in games. The second question about gaming involved the frequency and stated that often should be taken to mean every day or every two days on average. However, some participants expressed that they did not consider playing games on a daily basis as often.

The deviation of the total ITQ score and focus sub-scale were relatively small with a population standard deviation of respectively 0.49 and 0.40. Involvement scores were already somewhat more diverse with a standard deviation of 0.73, but most variety in scores can be found in the gaming and virtual reality sub-scales with standard deviations of respectively 1.7 and 1.6. Notwithstanding the diversity of the participant sample, the small number of questions of these latter two sub-scales should be taken into account.

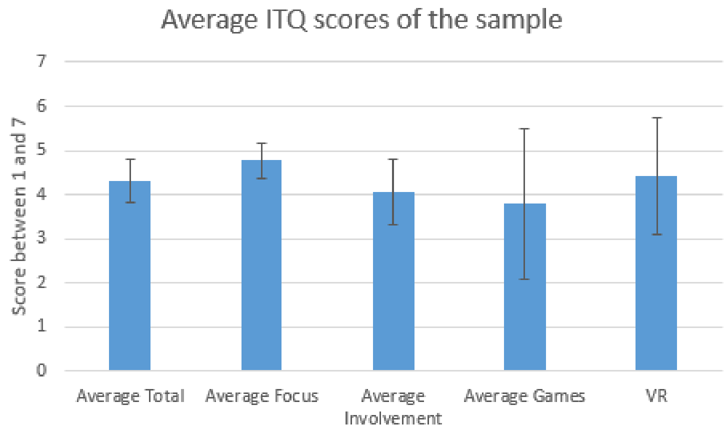


FIGURE 5.1: Average scores for each sub-scale and the total average. The error bars range from -1 to 1 standard deviation.

5.1.1 Score distributions of the immersive tendency sub-scales

The ITQ questionnaire contained sub-scales for focus, involvement and gaming. Focus and involvement each had 7 questions, gaming only 2. In this thesis it was extended with a single question about VR experience. The Sharpido-Wilk test was carried out for each sub-scale with the average scores over the questions that belong to that sub-scale of the participants. The null hypothesis for each test was a normal distribution of the data and the alternative hypothesis was a not normal distribution of the data. The interpolated p values for the total score ($p = 0.437$), focus sub-scale ($p = 0.175$), involvement sub-scale ($p = 0.341$), and gaming sub-scale ($p = 0.082$) were above 0.05 and the value for the VR sub-scale ($p = 0.012$) was below 0.05. Therefore, the null hypotheses for the total score, focus, involvement, and gaming sub-scale are retained and the null hypothesis of the VR sub-scale is rejected. The chart in figure 5.2 shows three bars of equal size with empty intervals in between and thus similar to an 'alternating uniform' distribution.

5.1.2 Score distributions of the immersive tendency questions

The questions and responses of the ITQ are included in appendix section A.1. For 3 questions of the involvement sub-scale (question 4, 10 and 15 with p-values 0.089, 0.444 and 0.316), 3 of the focus sub-scale (question 7, 11 and 16 with p-values 0.054, 0.508 and 0.122), and 2 of the gaming sub-scale (question 6 and 12 with p-values 0.189 and 0.082), the scores of the participants were normally distributed.

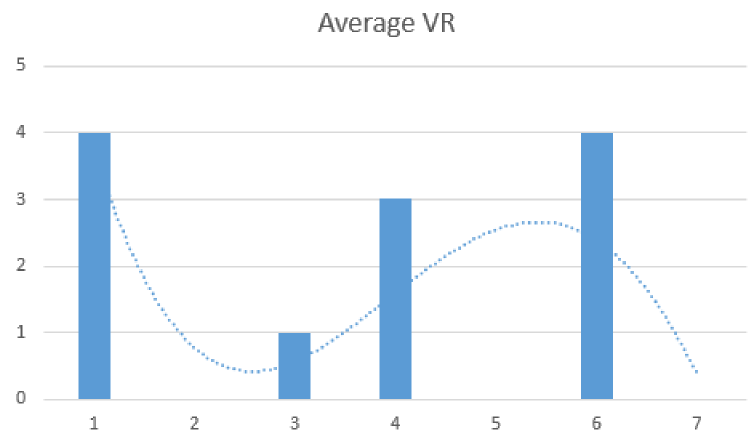


FIGURE 5.2: Frequency of the VR scores with an interval of 1. The trend-line is a fourth order polynomial.

5.1.3 Results of ANOVA within the immersive tendency scores

The VR experience sub-scale was left out the ANOVA, because it violated the assumption that the residuals of the sub-scale were normally distributed. First, three one-way ANOVA were performed between the participants, respectively within the total score, focus, involvement and gaming sub-scales. A two-way ANOVA with replication was performed between the participants within the focus and involvement sub-scale. The total score and gaming sub-scale were not included in the two-way ANOVA, because the number of questions per scale, the sample size, was different from the number of focus and involvement questions. The correlation between the sub-scales, including total score, gaming and VR experience, was analyzed separately.

Analyses of the sub-scales

In all analyses of variance, the null hypothesis was that the groups were randomly sampled from the same population and the alternative hypothesis was a significant difference between the groups. One-way ANOVA within the total score resulted in a significant difference between the participants ($F = 6.452$ and $p = 4.352 \times 10^{-9}$). One-way ANOVA within focus, involvement and gaming returned not significant differences (respectively $F = 0.736$ and $p = 0.701$, $F = 1.492$ and $p = 0.153$, and $F = 2.306$ and $p = 0.083$). The two-way ANOVA between the participants within the samples of focus and involvement scores showed a significant difference between the samples ($F = 13.081; p = 4.12 \times 10^{-4}$), but no significant differences between the participants ($F = 1.578; p = 0.111$) or interaction between the participants and the samples ($F = 0.783; p = 0.657$).

	Average Total	Average Focus	Average Involvement	Average Games	VR
Average Total	1				
Average Focus	0.766121061	1			
Average Involvement	0.662103527	0.20930359	1		
Average Games	0.672463502	0.57285644	-0.013900713	1	
VR	0.23535885	0.37252687	-0.27831158	0.48207006	1

TABLE 5.1: Correlation of the ITQ sub-scales.

The correlation between the sub-scales is presented in table 5.1. The correlation is computed using the formula $r(X,Y) = \frac{\sum(x-\mu_x)(y-\mu_y)}{\sqrt{\sum(x-\mu_x)^2\sum(y-\mu_y)^2}}$ where μ_x and μ_y are the sample means of sub-scale X and Y. The correlation between VR experience and the other sub-scales does not exceed 0.4. The other sub-scales have a correlation above 0.7 to the total score and the gaming scores correlate to the focus scores with a value of 0.57.

Analyses of subgroups

Three potentially interesting factors arose from the data and the execution of the experiment. The VR experience responses indicated three subgroups: those who never experienced VR before, users that experienced VR once or twice, and regular users. Gaming had a relatively high correlation with the total score given that it only contributed to the total score with 2 questions. Therefore, two subgroups were defined based on their gaming score: half of the participants had a score of 4 or below and the other half had a score of 4.5 or above. The third subgroup was derived from the experiments because of the noticeable issues of the participants who wore glasses. The chosen subgroups for this factor were participants who wore glasses, those who had corrected vision (contacts or laser eye treatment) and people with naturally normal vision.

One-way ANOVA between these groups within the total score did not result in significant differences (see table 5.2). Two-way ANOVA within the sub-scales focus and involvement returned a significant difference between the samples, but no significant difference or interaction between the columns and the samples (see table 5.3.

Lastly, one-way ANOVA was used to determine the impact of the participant design. This resulted in a F-value of 0.045 and a p-value of 0.956 for the three groups that started with a different viewpoint visualization and a F-value of 0.296 and a p-value of 0.590 for the two groups that started with a different layout of images. Hence, there were no significant differences between the groups that started with a different configuration.

	F-value	p-value
Gaming	0.214	0.647
Vision	0.822	0.446
VR	2.494	0.093

TABLE 5.2: F- and p-values of the one-way ANOVA between subgroups within the total score.

	Gaming		Vision		VR	
	F-value	p-value	F-value	p-value	F-value	p-value
Sample	8.334	0.008	11.093	0.002	10.471	0.003
Columns	0.490	0.491	0.438	0.649	0.839	0.440
Interaction	0.765	0.390	0.581	0.565	0.022	0.979

TABLE 5.3: F- and p-values of the two-way ANOVA between subgroups within the samples focus and involvement.

5.2 Usability scores

Figure 5.3 shows the average scores per usability aspect and the total average of usability scores. The ratings of the aspects usefulness, ease of use, and satisfaction were in the range of 5.1 and 5.4 whereas the ease of learning score was substantially higher with an average of 6.4.

This emerges again in the population standard deviation of these aspects. The scores of the participants for the usefulness, ease of use, and satisfaction aspects deviated respectively 1.0, 1.0 and 0.96 from the average, while the diversity of the average ease of learning scores per participant is merely 0.76.

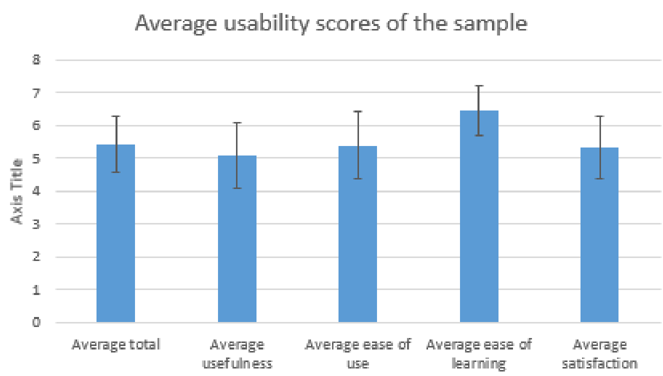


FIGURE 5.3: Average scores for each usability aspect and the total average. The error bars range from -1 to 1 standard deviation.

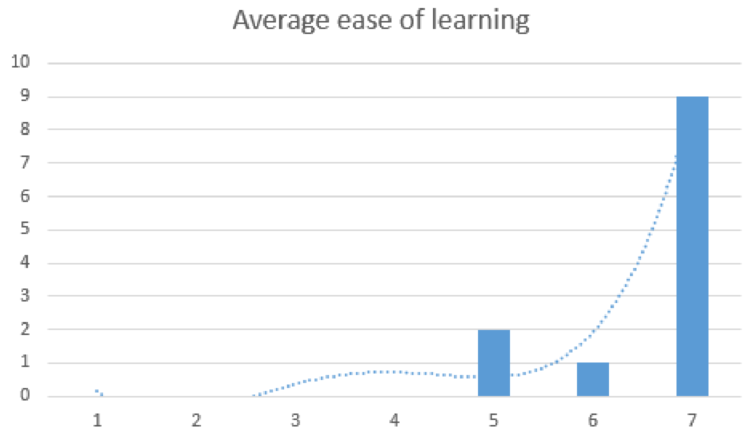


FIGURE 5.4: Frequency of the ease of learning scores with an interval of 1. The trend-line is a fourth order polynomial.

5.2.1 Score distributions of the usability aspects

The usability questionnaire can be categorized into four usability aspects. The first 8 questions were concerned with the usefulness aspect, followed by 11 questions about the ease of use, 4 about the ease learning, and 10 about satisfaction. According to the results of the Sharpido-Wilk test, the distribution of the scores that belong to usefulness ($p = 0.533$), ease of use ($p = 0.501$), satisfaction ($p = 0.488$), and the complete set of usability questions ($p = 0.601$) was normal. Only the distribution of ease of learning ($p = 0.010$) resulted in a p-value below 0.05 and thus the null hypothesis of a normal distribution was rejected with 95% confidence. The chart in figure 5.4 shows one large bar of 8 participants and two small bars of 3 and 1 participant next to it, similar to an exponential distribution.

The score distributions of usability aspects usefulness, ease of use, and satisfaction were also evaluated per subgroup. Only the test with usefulness scores from the group with lowest VR experience resulted in a rejection of the null hypothesis. This subgroup rated 5 usefulness questions with a 5.5 on average, 2 with a 6 and 1 with a 4.5. Hence, the distribution of the scores was too skewed to retain the null hypothesis of a normal distribution.

5.2.2 Score distributions of the usability questions

The questions and responses of the usability questionnaire are included in appendix section A.2. All usefulness questions resulted in a normal distribution of the scores by the participants. The question "I quickly became skillful with it" ($p = 0.087$) was the only question of the ease of learning aspect that retained the null hypothesis of a normal distribution. Question 9, 10, 15 and 17 from the ease of use aspect and question 24, 28, 29 and 30 of the satisfaction aspect resulted in a rejection of the null hypothesis. The charts in appendix section A.2 show that the distributions of the scores in all but one of these cases contained either a single bar or two not adjacent bars that were much larger than the others. The exception is the question "Both occasional and regular users would like it" ($p = 0.047$). From the 12 participants, none rated the question with a score of 4, 7 rated it with a score of 2 or 3, and the remaining 4 participants spread their ratings over the scores 7, 3, 2 and 1. Therefore, even though the peak was spread over two scores, the range of scores outside of the peak was too large for a normal distribution. Instead, the distribution was skewed. A repetition of the experiments with more participants may educate whether the range was too large because of outliers or representative for the total population.

5.2.3 Results of ANOVA within the usability scores

The ease of learning aspect was left out the ANOVA, due to the violation of the normality assumption. Parallel to the ITQ questionnaire, ANOVA was used to determine significant differences between the sub-scales derived from the usability aspects and subgroups. Additionally, the results were compared with ANOVA between the subgroups within the ITQ scores and a new type of subgroups was introduced: Half of the participant sample with below average ITQ scores and half of the participant sample with above average ITQ scores.

Analyses of the sub-scales

In all analyses of variance, the null hypothesis was that the groups were randomly sampled from the same population and the alternative hypothesis was a significant difference between the groups. First, one-way ANOVA was used to figure out whether the new usefulness and satisfaction questions affected the scores of the usefulness and satisfaction aspects. In both cases, the null hypothesis of no significant differences was retained. ANOVA between the participants within the new and old usefulness questions resulted in an F-value of 0.366 and p-value of 0.551, and the new and old satisfaction questions in an F-value of 0.110 and p-value of 0.743. One-way ANOVA between the participants within the total scores returned an F-value of 23.469 and a p-value of 8.758×10^{-37} and thus the null hypothesis was rejected with at least 95% confidence.

One-way ANOVA between the participants for the usability aspects usefulness ($F = 10.073$ and $p = 1.988 \times 10^{-11}$), ease of use ($F = 11.377$ and $p = 2.971 \times 10^{-14}$), and satisfaction ($F = 9.547$ and $p = 7.928 \times 10^{-12}$) all resulted in a rejection of the null hypothesis and thus significant differences in the scores between the participants.

	Average total	Average usefulness	Average ease of use	Average ease of learning	Average satisfaction
Average total	1				
Average usefulness	0.918018	1			
Average ease of use	0.958846	0.801221644	1		
Average ease of learning	0.583792	0.470063803	0.612572539	1	
Average satisfaction	0.930111	0.834697543	0.847685346	0.328933455	1

TABLE 5.4: Correlation of the usability aspects.

	Gaming		Vision		VR	
	F-value	p-value	F-value	p-value	F-value	p-value
Sample	1.009	0.371	1.220	0.301	1.134	0.327
Columns	1.368	0.247	23.663	8.07×10^{-9}	21.92	2.46×10^{-8}
Interaction	1.756	0.182	2.048	0.095	1.685	0.161

TABLE 5.5: F- and p-values of the two-way ANOVA between subgroups within the samples usefulness, ease of use and satisfaction.

Two-way ANOVA between the participants within the samples usefulness, ease of use, and satisfaction also showed a significant difference between the participants ($F = 32.722$ and $p = 2.805 \times 10^{-46}$), but no significant difference between the samples ($F = 1.688$ and $p = 0.186$). The interaction between the participants and the usability aspects was significant ($F = 2.077$ and $p = 0.0035$).

The correlation between the usability aspects is displayed in table 5.4. All aspects, except for ease of learning, had a correlation larger than 0.92 with the total average and larger than 0.82 with each other. The highest correlation of ease of learning was with ease of use (0.56).

Analyses of the subgroups

Two-way ANOVA was used between the subgroups with the usability aspects as sample. The results are displayed in table 5.5. The null hypotheses of the samples and interaction were retained. The scores of the two groups related to gaming experience also resulted in preservation of the null hypothesis. Between the scores of the vision subgroups and between the scores of the VR experience subgroups there were significant differences which led to a rejection of the corresponding null hypotheses. The average usability score of the participants with glasses was 4.640, while the average usability scores of participants with normal and corrected-to-normal vision were respectively 5.325 and 5.571. The average usability score of participants with average VR experience was 4.607 and the average scores of the participants that never or regularly used a VR device were respectively 5.329 and 5.578.

Lastly, one-way ANOVA was used between the subgroups that started with different visualizations and layouts. The ANOVA between the layout subgroups resulted in an F-value of 0.0183 and p-value of 0.896 and thus no significant difference between the participants who started with the random and the firework collection. Between the viewpoint visualization subgroups, ANOVA returned an F-value of 11.712 and a p-value of 0.00151 and thus a rejection of the null hypothesis. The average usability score of the participants who started with the lens (4.530) was slightly less than 1.2 lower than the score given by the participants who started with the stack and slightly less than 1.5 lower than the score from the participants who started with the cursor (6.068). The ITQ scores of these groups were not significantly different.

5.3 Interviews

Six of the participants mentioned that the environment of surrounding numbers and colors felt natural. Three stated that it was practical and two reported that it both felt natural and practical to use. One participant commented that he found it unnatural and cumbersome to use head rotations instead of much smaller movements such as a joystick or mouse.

Seven participants mentioned that they looked at the center during the navigation. Four participants looked at the edge of the view-port. One of the participants that looked at the edge of the view-port mainly used the structure of the layout to navigate and the other three to look at the approaching images. Two of the seven participants looking at the center of the view mainly used the structure to find the desired number and color. The other five participants looked at the numbers and colors that were passing by. Nine of the participants preferred viewpoint visualization for enhanced focus on the center. Two desired the viewpoint visualization to be aware of the exact center of the view. None fancied the lack of viewpoint visualization.

Eight participants preferred the lens, three preferred the cursor, and one preferred the stack. The stack was favored by one participant because it was easy to switch the focus to the neighbors. Seven of the other participants strongly disliked the stack, because it did not work as expected. The four participants that looked at the edges of the view all remarked that the stack was too distracting from their intention to look at the edge. The stack drew their focus to the center even though this was not desired. The other three participants mentioned that the stack made it hard to distinguish the center of the view from its neighbors. The three participants who preferred the cursor all looked at the edges of the view while navigating. From the six participants that preferred the lens, all used the passing numbers and colors more than knowledge of the structure. From each of the groups that preferred a visualization type, one participant mentioned the distortion of the lens in the environment with numbers and colors. All participants noticed it in the environment of images.

The remarks concerning the layouts of images were less frequent and participants seemed to have more trouble to express different experiences than before when they were requested to express the different experiences with the viewpoint visualizations. Three participants mentioned that the diversity of the random collection was too overwhelming. Two of them also mentioned that it was harder to focus because of it. Four participants noticed that the levels of the stack were more clear in the environment with the random collection. One of the participants mentioned that the collection of firework images allowed him to search more globally on contrast and colors within the images. Lastly, two participants remarked that the visual element of the images was more useful and interesting to explore.

Chapter 6

Findings

This chapter elaborates on the results with respect to the research questions presented in section 2.4 and research purpose defined in the Introduction. The following four sections discuss these questions and the final section covers the research purpose. Each of the sections is introduced by the question that is addressed.

6.1 Viewpoint-controlled navigation

Does viewpoint-controlled navigation in a virtual environment with a layout surrounding the user feel intuitive?

To answer this question, we first need to clarify the definition of intuitive in this context. Earlier on (section 2.2.4), Raskin (1994) was cited on his claim that the term intuitive is often (mis)used in occasions where familiar is more appropriate. In this situation, virtual reality interfaces are relatively new and a third of the participants never experienced virtual reality before and it was hypothesized in this thesis that looking around is a natural way to scan the environment (section 2.3.2). Hence, viewpoint-controlled navigation may be closer to intuitive interaction than familiar interaction. In accordance with the impression of intuitiveness described by Raskin, intuitive interaction should feel naturally and human intuition should suffice to use it, without training or rational thinking. The latter could not be tested completely, because all participants had heard about head-mounted displays and thus had prior knowledge about the fact that the device was meant to be mounted on the head. Ignoring the intuitiveness of the device, we look into the intuitiveness of viewpoint-controlled navigation when already immersed in the virtual world. The next paragraphs address to what extend the viewpoint-controlled navigation felt natural according to the users, if it was intuitive to learn to use it, and if the observations correspond with a natural way to process the eye movements.

6.1.1 Does it feel natural?

When asked to express their initial experience with the layout of images that was wrapped around them, eight of the participants mentioned that it felt natural. One of the participants that never experienced virtual reality before expressed that she felt so immersed in the environment that she tried to point in physical space while describing what she saw in the layout. Hence, it should be noted that participants may have used the phrasing natural for the lack of a word to describe the immersion.

6.1.2 Is it intuitive to learn?

The ease of learning questions from the usability questionnaire were rated very positive. None of the participants gave any of the questions a score below 4 and the peak was at the maximum score. This was also the case for ease of use question "I can use it without written instructions". A few of the participants asked about the point of these questions, because they found it too obvious. One of these participants mentioned: "Of course it is easy to learn, there was nothing to be learned. All I had to do was look around.". However, one participant did not understand the scrolling effect without any explanation and four participants were unaware of the ability to scroll up and down through the colors. Whether this was caused by the preference for a horizontal visual axis (Moorrees and Kean, 1958) or the fact that humans are more familiar with scrolling through numbers than colors is an interesting research question for future work.

6.1.3 Do the observations correspond with observations of natural eye movements?

The second paragraph of section 5.3 describes the observations related to eye movement and focus during the navigation tasks. The results are for a large part in line with the observations discussed by Kahneman (1973) in relation to eye movements and the spatial orientation of thought. Master chess players were allowed to study a complex chess situation for five seconds and were able to perceive the best possible moves for both opponents, even though their eye movements did not correspond with fixations on the pieces of interest. However, when the masters were allowed to study the board for 10 and 15 seconds, the eye movements correlated more with the important pieces and moves to perform. Kahneman derived from this the suggestion that the correlation between the physical locus of the eye and the perceived locus is optional rather than obligatory. This was also observed in the study of Kaplan and Schoenfeld (1966), in which participants were asked to solve anagrams that all followed the same transposition. Participants that were aware of this transposition fixated on each of the characters exactly once in the reversed sequence of the transpositions. Those who were not aware of the pattern did not follow the same sequence of transpositions, but were able to solve the anagrams in not significantly different periods of time.

The observations from the experiments in this thesis also indicate an optional correlation between the eye movements and spatial orientation of thought. In the interview the term focus was used to identify the spatial orientation of thought. Although physical eye movements were not measured, participants were confident to answer how they used their eyes to look around during the navigation. Eight participants fixated on the center of the view and four looked at the edges in the scrolling direction. Two of the participants looked toward the edges to see what numbers and colors were coming and the other two argued that they looked to the edge, because it was in the direction of their target. Seven of the participants fixated on the center to assimilate the numbers and colors passing by. One participant explained that she fixated on the center to avoid being overwhelmed by the addition of motions and that her glasses made it harder to look near the edges of the view.

In general, we see that most of the participants preferred to fixate on the center, while they did capture the information around the center as well. A smaller group used a similar approach, but fixated on the edge of the view and another small group correlated their gaze with the spatial orientation of thought. The preference for fixation on the center of the view may be explained by the movements and thus relatively short time to process the numbers and colors, as was the case by the one participant that fixated on the center to avoid being overwhelmed and the master chess players with little time to process the situation.

6.1.4 Is it appreciated?

The average scores per usability aspect are slightly higher than 5 on a scale from 1 to 7. The usefulness aspect was mostly task-oriented, except for the question "It does everything I would expect it to do" which was rated very positive. The ease of use questions "I don't notice any inconsistencies as I use it" and "Both occasional and regular users would like it" resulted in mixed scores, but some of the participants remarked that they based their ratings mainly on respectively visualization and usefulness. The questions "It is flexible" and "Using it is effortless" were also rated differently by the participants. There seemed to be a consensus that viewpoint-controlled navigation alone is not sufficient. Either a smart recommendation system or additional filter and search methods are needed. Despite the consensus, the question appeared not to be interpreted in a similar fashion by the participants of which some based their score on the navigation technique and others on the effectivity of the image browsing tasks.

Glasses

The variety of scores within the results of the question about effort consumption seems correlated with the factor vision. The participants with glasses required more time to put on the head-mounted display and two of the participants hold the device with their hands for better grip and less pressure on the glasses. Those participants rated the question about effortless the lowest with 3.75 points on average. Remarkably, the difference between the participants with normal vision (4.75) and corrected-to-normal vision (6.0) was even larger. Given the normal distribution of scores on the question and a relatively small participant sample, it could be a coincidence that the participants with the highest ratings happened to have corrected-to-normal vision. On the other hand, participants with corrected-to-normal vision had worn glasses for a substantial period of time before switching to contacts and thus may have had less issues with the heavy head-mounted display. Unfortunately, the author of this thesis could not find any literature related to this issue.

Although it may be hard to evaluate the time it takes to get accustomed to head-mounted displays independent of the chosen device, it seems at least wise to consider this effect instead of waiting for the hardware to improve. Therefore, research is desired on acceptance and willingness of people to wear virtual reality glasses and adapt their eyes to the lenses. Since current technology of head-mounted displays requires a lens for each eye and a split screen, devices similar to regular glasses seem to be most convenient as these regular glasses are already optimized for wearing lenses in front of the eyes. In an extreme stadium contacts or implants could also be realized, but people will likely be hesitant for such devices until they are familiar with virtual reality glasses and the benefits of contacts or implants should outweigh those of glasses. Li et al. (2008) contacted 597 middle school students with reduced sight in the rural town of Xhichang in China. More than half of the participants (56.8%) did not have glasses before and only 30.7% of them bought glasses when offered. Nearly half of them were satisfied with their current vision (48.7%), 17.6% was concerned over the expense, and 12.8% feared that glasses would weaken their eyes. A similar study was performed by Keay et al. (2010) with a sample of 428 junior high school students with reduced sight in the urban area of Guangzhou. The participants were provided with free glasses and visited unannounced one month later. Half of the participants did not wear the glasses anymore, for a large part because of appearance and anxiety of being teased.

Virtual reality glasses

Although the participant samples from the previous paragraph are not representative for the early adopters of virtual reality, the concerns may very well exist regarding virtual reality glasses as well. According to a survey in the United States participants were asked for the amount of money they were willing to spend on a virtual reality headset. The most selected options were less than $150 (22%), between $150 and $250 (33%), and between $250 and $400 (28%) (Statista, 2016a). When asked if they would buy such a headset given that the price is acceptable, 16% rejected the statement (Statista, 2016b). Unfortunately, explanations were not provided. Equal to the reasons not to buy glasses, lack of interest and fear for eye damage may well be significant factors in the decision to dismiss virtual reality glasses. Skeptics are often concerned with the current direction of virtual reality. They are not convinced that virtual reality is suitable for mainstream gaming (Nace, 2016) (Bailey, 2016) and requires more extensive productions for virtual reality films to exceed basic three-dimensional films (Gallaga, 2015). Health and safety issues could be of concern as well. Cobb et al. (1999) assessed the effects of participating in virtual environments with head-mounted displays. Most problems were relatively minor and many of the complications could be reduced by external improvements, still 5% of the participants experienced serious negative effects.

The participants of the experiment described in this thesis rated the usability aspect quite positive with a 5.2 on average, although the questions, except for "It is fun to use" and "I am satisfied with it", also got a few negative scores. Especially the question "I feel I need to have it" was rated neutral by five participants and negative by three. This corresponds to the reluctance of the participants when asked if they had browsed through image collections before and whether they could see themselves get more involved with image browsing if they were not bounded by any limitations. These responses are further discussed in section 6.2. All three participants with a negative rating remarked that viewpoint-controlled navigation in virtual reality felt more like a gimmick than practical for image browsing. On the follow-up questions whether they preferred speed or experience, they all replied speed. None of the participants mentioned any concern for health and safety issues or worry about their appearance. Comfort was only an issue for the participants with glasses.

6.1.5 Conclusion about viewpoint-controlled navigation

In conclusion, our sample of participants required hardly any feedback to learn to use viewpoint-controlled navigation in virtual reality and found the interaction with a layout wrapped around the virtual eyes natural. The participants used eye movement strategies similar to those of master chess players with little time to observe the situation and participants who were asked to solve anagrams. The intuitiveness of the interaction was appreciated by all participants, but some were apprehensive of its applicability to their image search needs and people who wore glasses had difficulties with the effort consumption.

6.2 Exploratory image browsing

What are the needs for exploratory image browsing?

Before the second part of the experiment with images instead of numbers and colors started, the participants were asked about their usage of image browsing systems and their desires. In the previous section it was already mentioned that most of the participants were reluctant to answer these questions. Only two participants used an image platform other than Google Images, Facebook and Instagram and only five of the participants occasionally browsed through personal photo albums. This complicated the follow-up question, because the participants seemed to look for extensions of their current usage of image browsing systems rather than imagining new ways to interact with images. When the examiner steered the participant into such a direction, some became enthusiastic. One of the participants was involved with brain tumor research and used Google Images to look up images of a certain spot in the brain. When the examiner proposed a system where the images are associated with their corresponding location on a map of the brain, the participant responded that he would definitely use it and mentioned that the viewpoint visualization could be a nice addition to have a closer look at an area. The journalist needed images to suit the content of the articles and attract readers to buy the newspaper. He was pleased with the idea of a system that uses semantic analysis to recommend images that match the content of the article. Furthermore, a special education teacher was interested in using content-based interfaces to browse through images related to the content of the lectures. Lastly, an art student had to use an outdated interface to browse through the art collection. It was a direct port of the physical collection, where users could browse the collection according to the categorical indices of the images. She stated that even the slightest improvement of search functionality would be appreciated. The proposal of a content-based system sounded interesting to her, but a keyword-based system with filters would suffice as most of her needs were concerned with direct searches on descriptive features such as the painter, time period and style.

Contrary to section 2.1.1 and 2.1.2, the participants saw more potential in exploratory image browsing systems for professional tasks.

Possibly, because they could apply their knowledge of the domain or were confident that a recommendation system could posses enough knowledge about the domain to make educated decisions. The participants who browsed through personal photo albums and photo collections on Facebook and Instagram did not see any need to improve the browsing experience. Although three participants mentioned that it would be fascinating to experience being surrounded by their personal image collections, they did not see themselves use it often, because they were satisfied with their current methods already.

The initial hesitance and skepticism of broad applicability concurs with the statement of Borup et al. (2006) that expectations about specific technologies are less robust than those of more generic application areas and more vulnerable to hype-disappointment. Although images and visualizations are common for many domains, people appear to have a quite specific idea about image browsing, that seems to originate from currently available systems. This is also the case for virtual reality, which has drawn a lot of attention from the media and large companies leading to the impression by many that virtual reality is intended for gaming and virtual presence only. Although all this attention is a huge incentive for the research areas concerned with virtual reality, expectations should be taken under consideration.

In this thesis viewpoint-controlled navigation in virtual reality was evaluated for exploratory image browsing, because it was hypothesized that it allowed users to scan the collection in a quite intuitive manner and as a result was appreciated by users for exploratory image browsing. However, it was not anticipated that the participants had low expectations of the potential of exploratory image browsing. Expectations could be raised when participants were able to express a specific intent. These needs where quite different for most of the participants and depended on domain specific knowledge. The participants could not conceive a system that was able to deal with more diverse user needs and rather relied on their own ability to process information into keyword queries.

Interestingly, this is different for video browsing systems such as YouTube, where about sixty percent of the clicks from the home page and thirty percent of all video clicks were on recommended videos and the click through rate of recommended videos was more than twice the rate of most viewed, top favorited, and top rated videos (Davidson et al., 2010) (Zhou, Khemmarat, and Gao, 2010). Furthermore, the highlight and search sections of YouTube result in more views of the videos with many views already whereas recommendations lead to a better distribution of views and helps users to find niche videos (Zhou et al., 2016). What makes it that people often make use of the video recommendation service while neglecting the desire for better image browsing systems? Perhaps it is related to the amount of tension required. One could actively study an image for a period of time or passively watch a video. The former requires more attention and thus possibly a more deliberate user need before it is pursued. Watching videos is often a more passive occasion where one sits down and wants to be entertained by the video.

To conclude, the findings in this thesis indicate that people are content with current ways to access images, even though they were much more enthusiastic when presented with an image browsing systems that was dedicated to their needs. This is in contrast with the hypothesis that exploratory image browsing is most suitable for broad leisure activities. The expectations of people about image browsing were more specific than anticipated in this thesis and they may have only considered active image browsing tasks.

6.3 Viewpoint visualization

What is the effect of visualization of the viewpoint on user experience?

In this thesis, we reduced the scope of viewpoint visualizations to a cursor and two types of fisheye views. Cursors were familiar to the participants and the fisheye views were hypothesized to enhance detail while remaining global context. The distinction between the fisheye views was made to evaluate a lens with physical properties of a magnifying glass and rearrangement of the images in depth without distortion. Participants were asked to express anything that came up, but in particular about the effect of visualization on the focus, orientation and distraction.

Focus and distraction related questions from the examiner evoked strong opinions. Orientation was only mentioned in comparison to the case without visualization. Two of the participants already mentioned their desire for a cursor in the phase without visualization, because they found it hard to orientate. In general, participants seemed to associate the cursor most with orientation, the fisheye lens with focus, and the fisheye view by stacked images with distraction. Participants remarked that the stack of images changed too much of the environment at relatively small movements. These changes draw their attention away from their point of interest towards the movements of the stack and left little room for exploration of the areas outside the center of the view. The fisheye lens also drew attention to the center of the view, but changes in movement were more gradually and thus it was less distracting to look at the areas around the lens. Least distracting from the edges was the cursor, but at the cost of little distraction at the center because it overlapped the object in the center. One of the participants commented that she looked left to the cursor, to avoid this overlap.

The intent of fisheye views worked best with the fisheye lens. However, all participants eventually expressed some nuisance about the distortion that was caused by the projection onto a sphere, especially in the layouts of images. A suggestion that was proposed by two of the participants was to use a flat circle instead of a sphere, but this would complicate the smooth transition between the global structure and the lens. Nonetheless, it seems wise to look for a fisheye view without distortion for image browsing. Furthermore, three-fourth of the participants preferred a fisheye view visualization over the cursor. These results are in line with the related work discussed in section 2.3.4. The observation of Sarkar and Brown about the fisheye views in the work of Hollands et al. seems also applicable to our fisheye view with stacked images. However, participants expressed that they were unable to deviate their focus from the center rather than being disorientated. Possibly, both disorientation and inability to deviate the focus are symptoms of expectations. Subjects in the study of Hollands et al. may have expected that the fisheye views occurred at the selected locations and participants in our experiment appeared to expect a slight delay or smoother transition when the stack shifted.

In summary, fisheye views are preferred over a cursor considering that the view should behave as expected with smooth transitions and no distortion. Such a fisheye view allows the user to easily switch between scanning the global environment and focusing on local detail. The position of the fisheye visualization also functions as orientation cue, although people that explicitly expressed the need for orientation preferred the cursor.

6.4 Layout of images

How does the layout affect the viewpoint-controlled navigation method and visualizations?

Three types of layouts were used of which two contained images and one a collection of ordered numbers and colors. The firework images had a high contrast within the image (bright colored firework effects against a dark background) and low contrast between the images. The random collection was more diverse and had more contrast between the images than the set of firework images.

Although none of the participants decided to switch their preference for the visualizations based on the layouts of images, four participants remarked that the stacked fisheye view worked much better with the random collection of images. The lens, on the other hand, suffered most from the random collection of images. Participants expressed that they were only able to see the image in the center of the lens properly due to the distortion at the edges of the lens. Ten of the participants responded that this effect was worse with the random collection than with the firework images. Nine of the participants did not notice the distortion with the layout of numbers and colors.

Perhaps this is related to the peripheral awareness briefly introduced in section 1.2.1. The firework images had several characteristic features that were shared between most of the images and the numbers and colors were structured in a logical order. Using knowledge about these characteristics and the structure, the participants may have been able to fill in the information that was lost by the distortion. Ramachandran and Gregory (1991) observed that participants actively filled in the artificially induced scotomas rather than ignoring the absence of information in those spots. Grossberg and Mingolla (1985) theorized that there are two parallel contour sensitive processes of which one prevents the observer from perceiving an overflow of featural quality and the other compensates for missing featural qualities. The former leads to illusions such as neon color spreading (Van Tuijl, 1975) and the latter to filling in information. Since the backgrounds of the numbers consisted of a single color, both these processes could be applied to spread the color of observed qualities and to fill in the information of the distorted qualities. Furthermore, Buffart and Leeuwenberg (1983) claimed that humans process objects as structures and interpretations instead of exhaustive specifications of those objects. Therefore, distortion of the numbers in our environment may have had little interference with the perception of those numbers.

For the three participants that used the the structure to navigate, it seems evident that they processed the structure rather than the specifications of the numbers. However, one of them noticed the distortion quite quickly and argued that distortion was the main reason for him to use the structure to navigate. Interestingly, seven out of nine participants who looked at passing objects did not notice the distortion and thus may have been processing the interpretations of the numbers passing by rather than the specifications of the distorted numbers within their focus.

Lastly, from the comments of the participants it appears that brightness and contrast between the images affects the perception of depth and distortion. Both depth of the stack and distortion of the lens were more clearly perceived by ten of the participants in the environment with random images than in the environment with firework images. The perception of depth of the stack corresponds with the results of Brigner and Gallagher (1974) and O'BRIEN (1958). O'BRIEN proposed that edges may not be perceived if the change in brightness of the surrounding area is gradually and small. One participant explicitly mentioned that the edges of the images in the stack were only visible in the random collection of images and another participant suggested that a small white bar between the images might have been helpful. Brigner and Gallagher was concerned with the relation between brightness and subjective contours. He showed that a high contrast brightness stimulated the perception of subjective contours. Coren (1972) believed that these subjective contours were a result of depth cues that were only present in the eye of the beholder. He argued that subjective contours are used to simplify a chaotic collection of complex two-dimensional elements into a simple and easily coded three-dimensional array of meaningful or symmetrical elements and thus mentally added depth to the two-dimensional information which caused the subjective contours. Brigner and Gallagher showed that these depth cues failed to sustain with reduced brightness contrast. This could explain the difference in depth perception between the relatively dark fireworks collection and more bright collection of random images. However, saturation and value of the numbers and background colors were set to the maximum of 1. This indicates that there can be more factors than lack of brightness alone that lead to reduced perception of depth. First comes to mind that the edges in both the environment of numbers and the firework collection were hardly perceived, whereas they were much more prominent in the collection of random images. Possibly, these prominent edges are perceived as turning points and thus indicators of a change in depth between the objects, whereas the lack of observed edges results in the perception of the stack as a whole. Consequently, the edges provide more local depth cues which seemed to be desired by most of the participants, because the center of the stack was of much more importance than the edges. The edges may have resulted in the mental option to switch between perceiving the structure of the stack and the interpretation of the center image. Therefore, it might be wise to evaluate the stack approach with distinct edges around the center image.

Distortion on the other hand, was more apparent in the collection of random images than the collections of numbers and firework images. In this scenario, perceiving the lens as a whole may have been preferred over perceiving the fragments of the lens for each image separately. Distortion towards the edges of the lens seemed to enable the participants to interpret the center image and the complete lens, while making it harder to identify the structure of the images mapped onto the lens.

In conclusion, stacking images appears to preserve the structure between the images and thus seems most applicable for diverse sets of images with high contrast between the images. Projecting the images in the center of the view onto a sphere enables the user to easily switch between the interpretation of the image in the center and the interpretation of the sphere that includes the images nearby the center. The lens seems most applicable for image collections that are arranged based on similar backgrounds.

6.5 Viewpoint-controlled navigation for exploratory image browsing

Is viewpoint-controlled navigation in virtual reality suitable for exploratory image browsing?

So far, the main research question was split up in four sub-questions about the intuitiveness of viewpoint-controlled navigation, its relevance for exploratory image browsing, the effect of visualizations of the viewpoint, and the influence of the layout. These questions lead to insights about the configurations of viewpoint-controlled navigation in virtual reality for exploratory image browsing.

First of all, it seems acceptable to assume that there is not a general viewpoint-controlled navigation method that is suitable for all types of exploratory image browsing systems. Besides, it appears that people only want to explore the collection if they are convinced that the system is able to understand and cope with the domain better than they are themselves using familiar search methods.

Secondly, there was no universal browsing strategy. Some people looked at the center of the view, while others looked towards the edges and a few used the structure of the numbers and colors to navigate, whereas other people looked at the passing numbers and colors.

thirdly, people appreciated the enhanced focus of fisheye views under the condition that the highlighted area corresponded with their intended region of focus.

fourthly, people were able to learn to control the viewpoint navigation quickly without written instructions and hardly any verbal guidance regardless of their previous experience with virtual reality. Moreover, most people expressed that viewpoint-controlled navigation in a layout surrounding the user felt natural or practical.

These insights provide the initial outlook on demands for viewpoint-controlled navigation methods in exploratory browsing systems. To formalize this prospect, more quantitative research is needed on the factors that correlated with different user experiences and preferences for other visualization methods. Our believe that these factors influence the perception of the users, and as a result the user experience and preferences, are related to the principles of Gestalt. It should be noted that this area of psychology is criticized for the lack of quantitative evidence, although this criticism is not always justified (Jäkel et al., 2016). Unfortunately, the papers providing quantitative evidence are diverse and heterogeneous and thus do not postulate a unified Gestalt theory. Nonetheless, whether it are the principles of Gestalt or any other explanation for differences in perception does not change that there is some form of processing in the brain that should be accounted for. Until there is a unified theory on perception, we may rather focus on the domain specific issues with perception and given that people are most receptive to exploratory image browsing for domain specific tasks, it seems wise to develop and evaluate viewpoint-controlled navigation in virtual reality for exploratory image browsing within a particular domain. In these cases, the visualization should be chosen such that it coheres with the desired perception of the users. For instance, because the virtual map of the brain is designed to visualize structure and brain scientists need to focus on local spots in the brain, a fisheye view with properties of a magnifying glass such as our lens enables the users to easily switch between perceiving the global structure and focusing on local spots. Distortion may actually be desired in this case, as it may enforce the brain to perceive the local spots isolated from their surroundings.

Finally, we conclude that viewpoint-controlled navigation in virtual reality is suitable for exploratory image browsing in cases where there is a desire for intuitive interaction and need to switch between global context and local detail. The capability of fisheye views to assist the user with perceiving part of the layout correctly is contingent upon the spatial location of targeted local detail. If only the global structure and the center of the view are important, a fisheye lens with distortion towards the edges seems suitable, according to our results. In case users intent to look at local detail outside the center of the view, the appropriateness of a fisheye lens diminishes while appreciation of the fisheye view without distortion grows. Our fisheye view without distortion, in which the images were stacked according to a pyramid shape, had too many flaws to be preferred over a lens.

Chapter 7

Impact

In this final chapter, we examine the impact of our method on the results and emerged conclusions. The first section elaborates on the participant sample. In particular, whether the sample was representative of the targeted population. The second section reviews the value of the immersive tendencies questionnaire and the validity of the USE questionnaire extended by custom questions in relation to the user experience. The third section reflects on the interviews, again in the light of user experience. The final section of this thesis summarizes the issues that are relevant to be addressed in follow-up or related work.

7.1 Participant sample

The participant sample was taken from acquaintances and thus consisted mainly of students of which more than average followed a computer science track. These computer science students had more experience with virtual reality. Remarkably, the immersive tendencies and usability scores were only significantly different for the participants that had experienced virtual reality only once or twice. Possibly, this subgroup was influenced by another factor as well due to the size of the subgroups. Since the scores of both the participants with little and a lot of experience with virtual reality did not diverge significantly, there seems to be hardly any advantage for more experienced users or excessive benefit from being introduced with virtual reality. However, it is possible that the effect of both is equal and thus explains the difference of these scores with the participants with occasional experience. Due to the relatively small sample size and only circumstantial evidence, further examination of this effect is needed.

Another interesting characteristic of the participants was their sight. One-third wore glasses, another one-third had normal vision from birth, and the remaining participants had contacts or were treated with laser eye surgery. This roughly corresponds with the statistics for the population of the Netherlands in 2012 (Bruggink, 2013). Unfortunately, three of the participants who wore glasses were assigned to the same order of the visualizations. As a result, the group that started with the lens visualization scored lower than the groups that started with the cursor and stack. Because the lower scores of the participants with glasses can be explained by the blatant discomfort of wearing both the glasses and the head-mounted display, no conclusions could be drawn from the lower score of the participants starting with the lens.

7.2 Questionnaires

The immersive tendencies questionnaire did not provide much insight in the immersive tendencies of the participants. None of the subgroups had significantly different ITQ scores. Participants often mentioned that the questions were too broad to answer them properly. All questions, except "Do you ever become so involved in doing something that you lose all track of time?", scored mixed results while all participants rated the sub-scales involvement and focus between 4 and 5 on average with no apparent difference between certain groups of participants. One of the reasons for the incompetence to answer each question from the heart could be the diversity of entertainment these days of which virtual reality is yet another category. One participant expressed that he was very immersed in games, but not at all when he read a book. In addition, some of the participants expressed that they hardly read any (fictional) books any more.

Since participants regarded the questions often as too vague, it might be better to evaluate immersion, involvement and focus based on an example. For instance, participants could be shown a couple of movie fragments with diverse genres and then being asked the questions about immersive tendencies keeping these fragments in mind.

The usability questionnaire seems to have captured the experienced usability of the participants quite well. Participants that expressed positive comments during the experiment rated especially the usefulness and satisfaction questions higher than average, whereas more skeptic participants rated these aspects below average. Ease of learning was rated very positive by all participants, regardless of their remarks.

One issue observed by the participants was redundancy. They could not distinguish the questions "It is easy to use" "It is simple to use", and "It is user friendly". The questions "I learned to use it quickly", "I easily remember how to use it", and "It is easy to learn to use it" often lead to remarks about their influence on each other. Furthermore, some of the participants required further explanation from the examiner about the questions "It is flexible", "I feel more engaged with the collection of images", and "I was more aware of the images around me". Flexible was considered too broad and thus ambiguous which aspect they were supposed to evaluate. They were told to express whether they considered the viewpoint-controlled navigation method to be suitable on more exhaustive search tasks. The questions about engagement and awareness were clarified by pointing out that they were intended to examine the engagement with the layout of images in a virtual reality and the awareness of the images in a layout that surrounded the user.

7.3 Interviews

The interviews developed over the duration of the experiments. Therefore, the first four participants were asked about their opinion of the layout surrounding the user and whether they looked at the center or to the edges a couple of days after they had participated. Another shortcoming of the interviews was that some of the participants seemed a little hesitant to be too critical. In retrospect, it may have been better to encourage the participants to be critical beforehand. Nonetheless, the examiner was not under the impression that information was withheld because of this, only that it was expressed more carefully.

7.4 Future work

A number of issues and opportunities were raised in the preceding chapters of this thesis. Noteworthy are the lack of quantitative evaluation methods to support the claims, and the interest in exploratory browsing through image collections within a confined domain.

7.4.1 Quantitative follow-up work

Participants did not all use the same browsing strategy and appeared to be more sensitive to distortion when there was no structure or similarity to fall back on. The hypothesis that were conceived are preliminary due to the small size of the participant sample and wide scope of the experiments. Therefore, quantitative research is necessary for a better understanding of these phenomena and, in consequence, facilitate the design of more suitable visualizations of the viewpoint and points of interest.

In particular, eye tracking can be used to identify the physical eye gaze besides the cognitive point of interest, and dependent variables such as distortion and structure should be singled out. For instance, the fisheye lens could be configured with different levels of depth either while remaining the same location or adjusting the position of the sphere according to the depth such that the front of the sphere remains at the same distance of the virtual eyes. The former may be useful if the aim is to find a trade-off between distortion of the edges and magnification of the center, whereas the latter is more appropriate to solely study the effect of distortion. The effect of structure could be evaluated by taking the environment of numbers as well as colors and vary their arrangement with both random and structured assignments of either the numbers, colors or both. A between participants design seems necessary to avoid carry-over effects between the configurations of the same environment with only one variable changed at a time.

Furthermore, glasses had a significant impact on the usability even though the participants with glasses expressed that the setup was tolerable before the experiment started. Observations about usability aspects, especially comfortability, derived from future experiments with head-mounted displays may be more robust if the results of participants with glasses are excluded from the main sample and reflected separately.

Moreover, the fisheye view visualization produced by stacking the images suffered from several side-effects, hypothesized to be caused by the lack of depth cues due to similarly colored backgrounds. Not only would it be interesting to study this phenomenon, a follow-up study is desired to compare the fisheye view that simulated a magnifying glass and a more advanced artificial fisheye view created by stacking the images. This advanced stack should have a clear distinction of the center image and likely benefits from smoother transitions from one center image to the other. Performing a quantitative study with a between participant design allows for a more detailed examination of the strengths and weaknesses of the fisheye views independently instead of exposing the participants to both.

Lastly, the scroll speed was fixed in the experiments of this thesis. The desired scroll speed may very well be dynamic considering the results of spatial navigation methods that were discussed in section 2.3.1. As the speed is expected to depend on the intent of the user, it seems sensible to evaluate dynamic speed protocols after the decision is made for an image collection and visualization. However, a more general experiment could provide interesting insights into the acceptance of users regardless of their willingness to spend more effort to achieve their goals.

7.4.2 Conceptualizing viewpoint-controlled browsing systems

We end this thesis with a prediction of user needs that will benefit from viewpoint-controlled navigation to explore, possibly large, collections. Our preliminary results suggested that a fisheye lens in combination with a structured layout of similar backgrounds helped the user to easily switch their focus between local detail and global context. Many of current image browsing systems provide the user with a set of results and demand the user to click on an image to enlarge it. Only 35% of the queries in the log file studied by André et al. contained a click, but the average number of clicks in those queries with at least one click was 5.4. Thus, users do not always need to switch to local detail, but if they do, there is a good chance that they will look at multiple images in detail. In case users of a particular system intent to view a lot of images in detail, viewpoint-controlled navigation with fisheye view visualizations may very well be an improvement over systems that require a more demanding action to enlarge the images in detail.

Image browsing with map-based interfaces

The sketch of a system that would allow the user to browse through a collection of images of the brain proposed in section 6.2 can be generalized to map-based interfaces where images are associated with spatial locations in the maps. The fisheye lens could show the images that are associated with the spatial location of the viewpoint while remaining the global orientation within the map. Distortion of the fisheye lens might be useful for a smooth transition between the global view of the map and local view of the images. Without any further input, this could be problematic for spatial locations that contain many images, as these will not fit all in the fisheye view. One solution that might retain a relatively high level of immersion is a physical sphere or trackball that represents the fisheye lens. This way, users make use of their head to interact with the global structure and their hand to interact with the images belonging to a local spot in parallel.

We illustrate the generalization with two distinct examples. History education often involves both details of historical events and how these are related. Therefore, a diagram of historical events that shows details about these events without leaving the global scheme would be an excellent tool for history teachers.

Another scenario is designing an architectural plan. Building a three-dimensional model of the plan captures the imagination quite well, but actual photographs of finished projects also contribute to a more vivid understanding of the realization of such a plan. We expect that virtual reality is the next step to explore these three-dimensional models. Hence, it would be great to have a system in which users walk around in a three-dimensional model of the architectural plan with a dedicated area inside the view reserved for photographs of materials and special features. This enables a client to build their own perception of the global plan and compare it with suggestions of the architect, which would help to make sure that both parties are on the same wavelength.

Browsing through a set of thumbnails

In this thesis, a layout of images was displayed and individual images were enlarged either by looking at that image for a short period of time or by hovering the fisheye view over it. Other media, such as videos, music tracks, text documents, and web pages, can be represented by an image, which is often called a thumbnail. Not only could this thumbnail be enlarged, it could also show a preview of the content that it represents. The preview of a video could be a small fragment or multiple thumbnails in sequence, the music track could start playing part of the song, and a short summary of the text document may be a useful preview for a potential reader. Text summaries in combination with a preview of a web page were useful to find previously visited pages, but this preview actually made it harder to find new websites in case the user does not know what kind of layout to expect for a particular topic (Dziadosz and Chandrasekar, 2002). Prior to the experiment of Dziadosz and Chandrasekar, Czerwinski et al. (1999) identified that participants were able to remember the spatial locations of web page thumbnails in a three-dimensional environment that contained one hundred snapshots of web pages.

Thus, we envision great potential for spatially arranged interfaces with thumbnails in virtual reality. Not only for collections of panoramic images and videos, but also in order to design your own multi-media portals. Viewpoint-controlled navigation with a dedicated area to preview the content could be very useful in such a personalized environment, because the thumbnails provide orientation cues to quickly explore the collection, while the preview area allows you to inspect the content behind the thumbnail before switching to the full experience.

Appendix A

Questionnaires

A.1 Immersive Tendencies

The charts of the score distribution per immersive tendency question are displayed from figure A.1 up to figure A.6.

A.1.1 Statistics

The average scores per participant for each of the usability aspects is displayed in figure A.7.

A.2 Usability

The charts of the score distribution per immersive tendency question are displayed from figure A.12 up to figure A.24.

A.2.1 Statistics

The average scores per participant for each of the usability aspects is displayed in figure A.25.

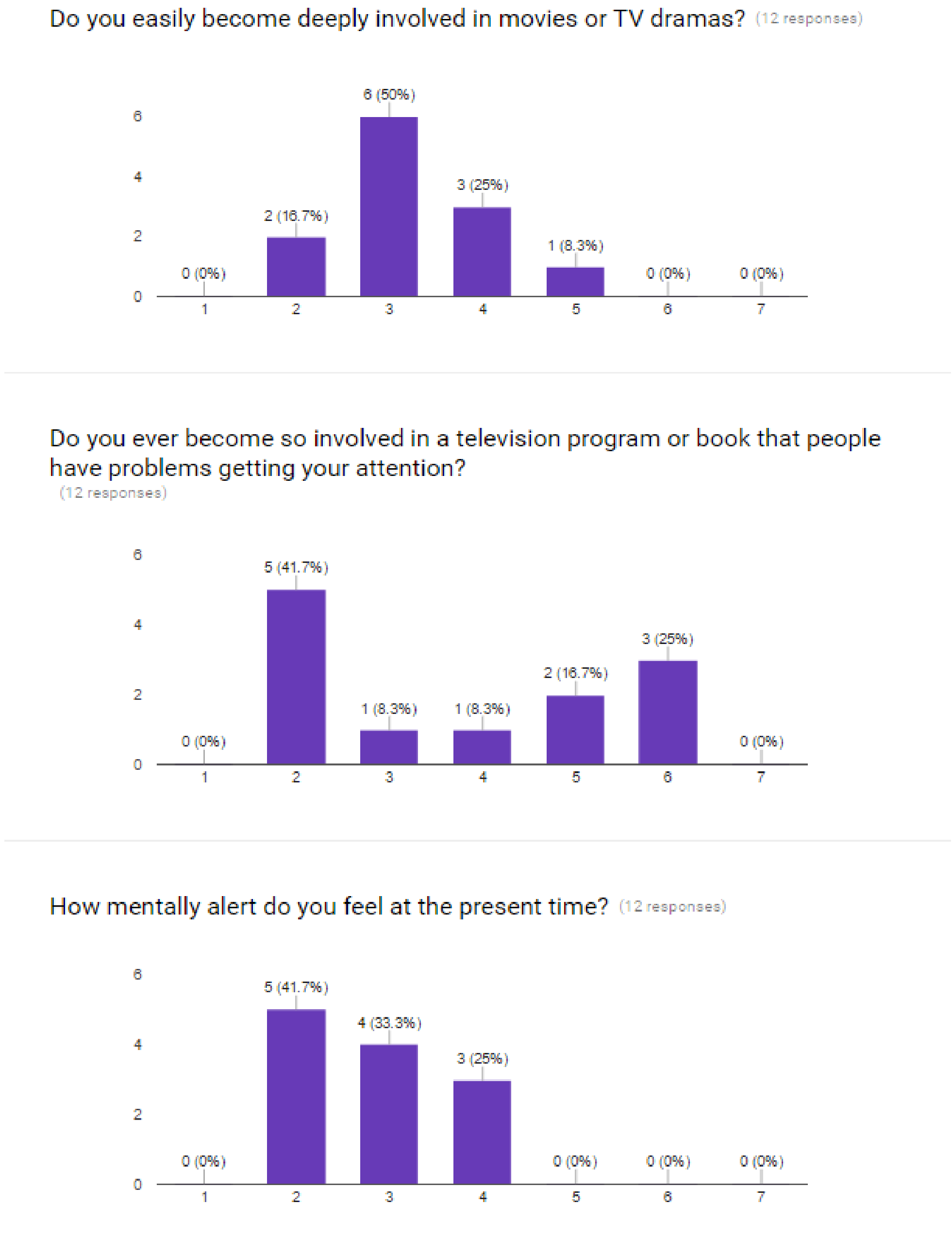


FIGURE A.1: Question 1, 2 and 3.

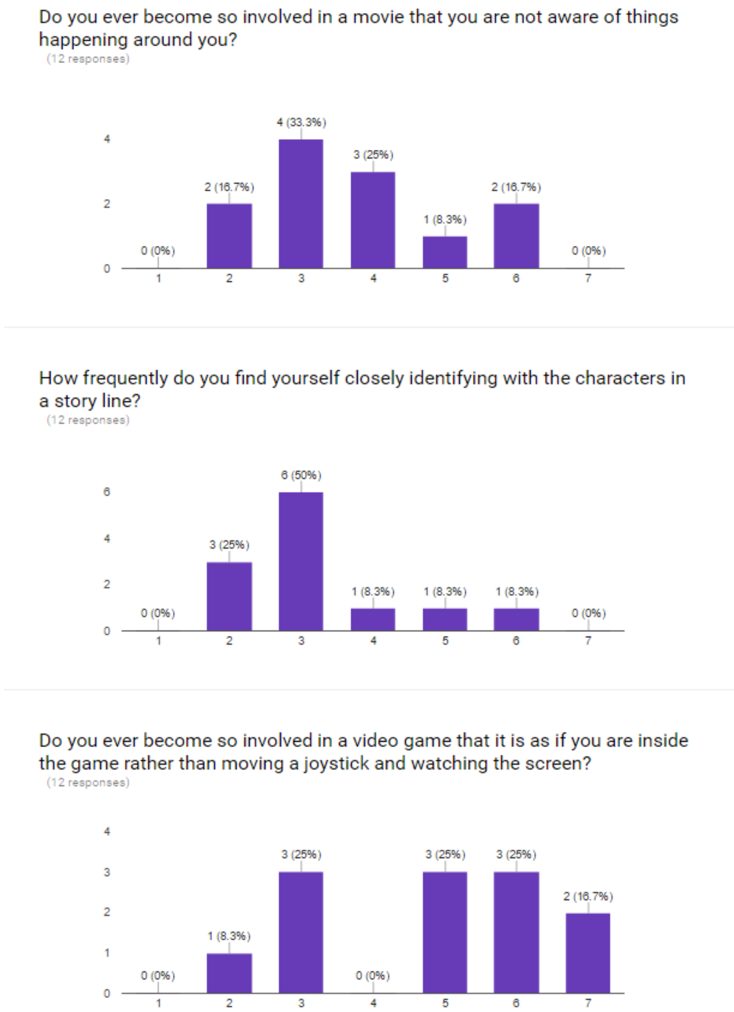


FIGURE A.2: Question 4, 5 and 6.

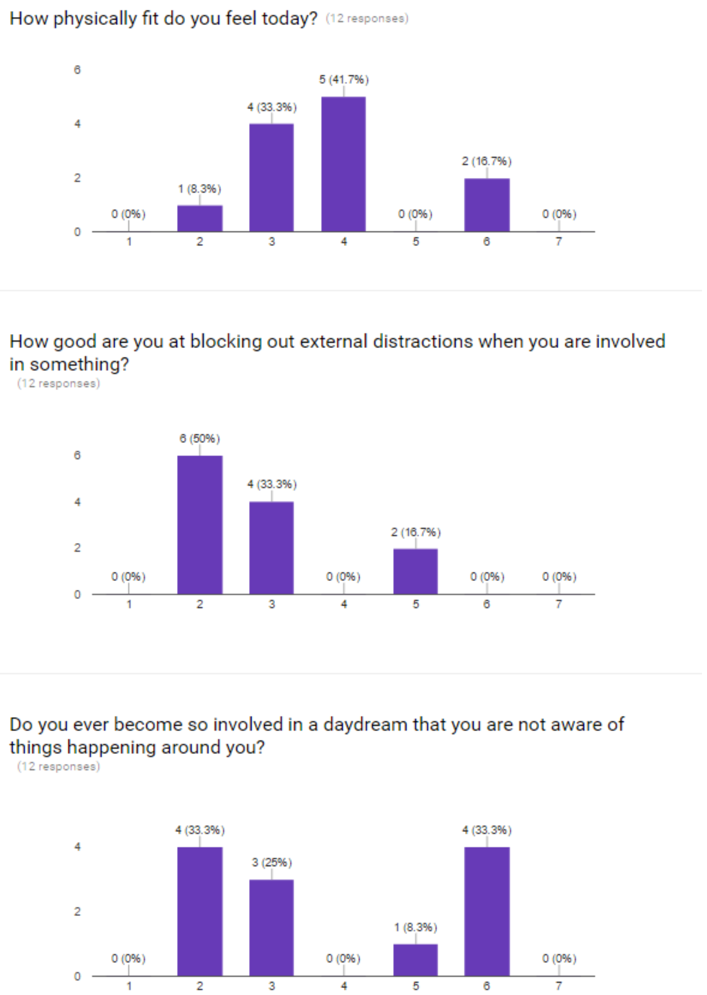


FIGURE A.3: Question 7, 8 and 9.

FIGURE A.4: Question 10, 11 and 12.

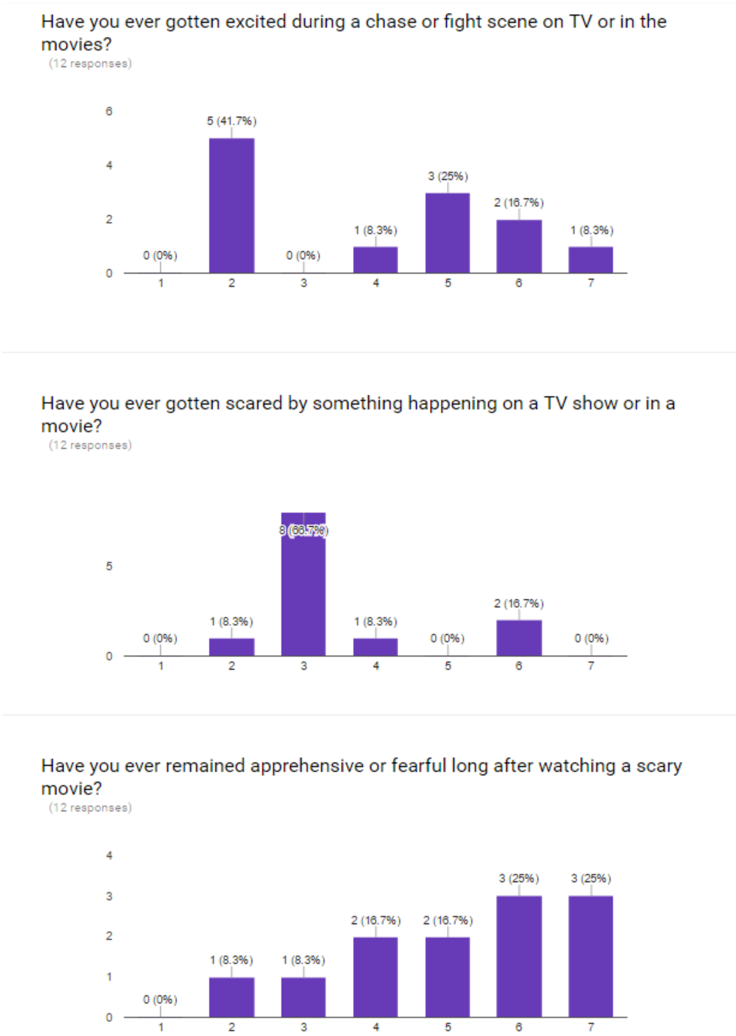


FIGURE A.5: Question 13, 14 and 15.

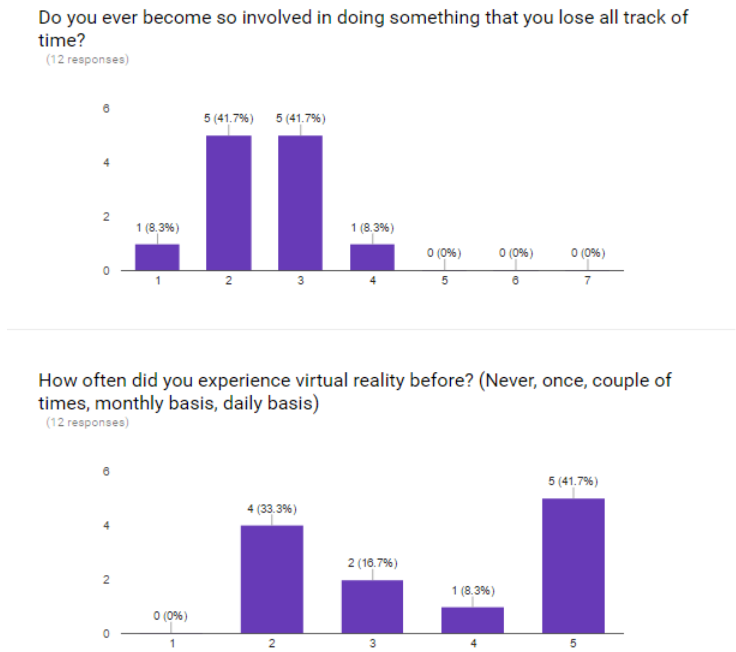


FIGURE A.6: Question 16 and 17.

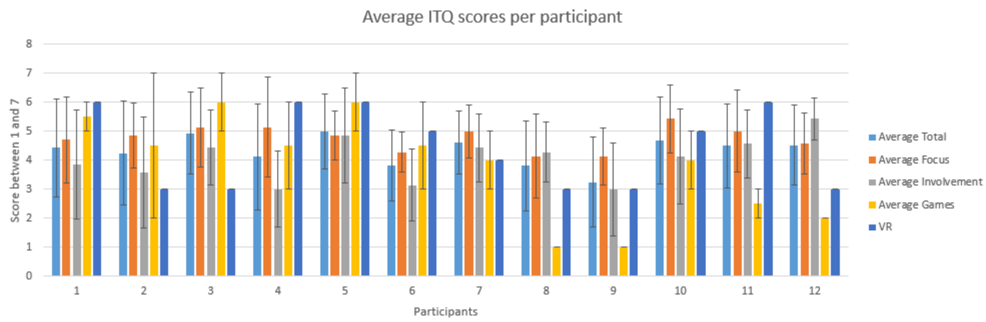


FIGURE A.7: Average ITQ scores per participant for each of the sub-scales.

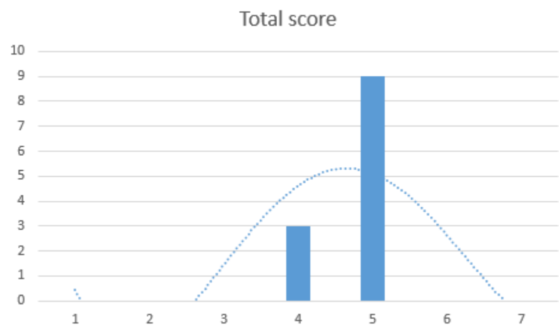


FIGURE A.8: Distribution of the average total ITQ scores.

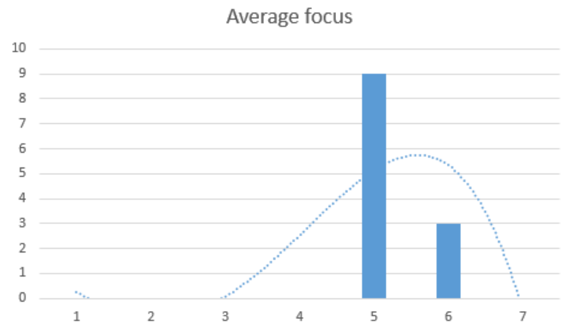


FIGURE A.9: Distribution of the average focus scores.

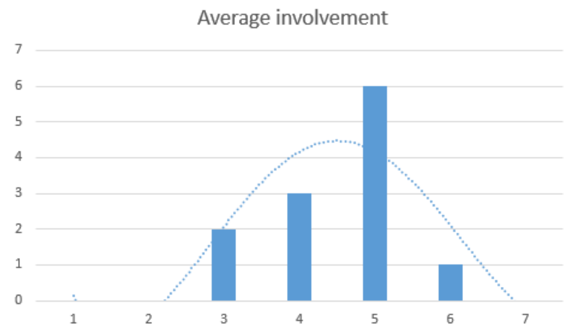


FIGURE A.10: Distribution of the average involvement scores.

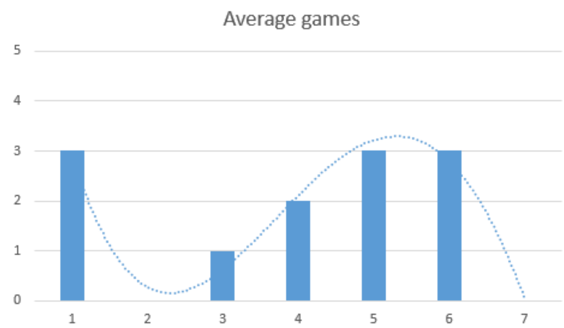


FIGURE A.11: Distribution of the average gaming scores.

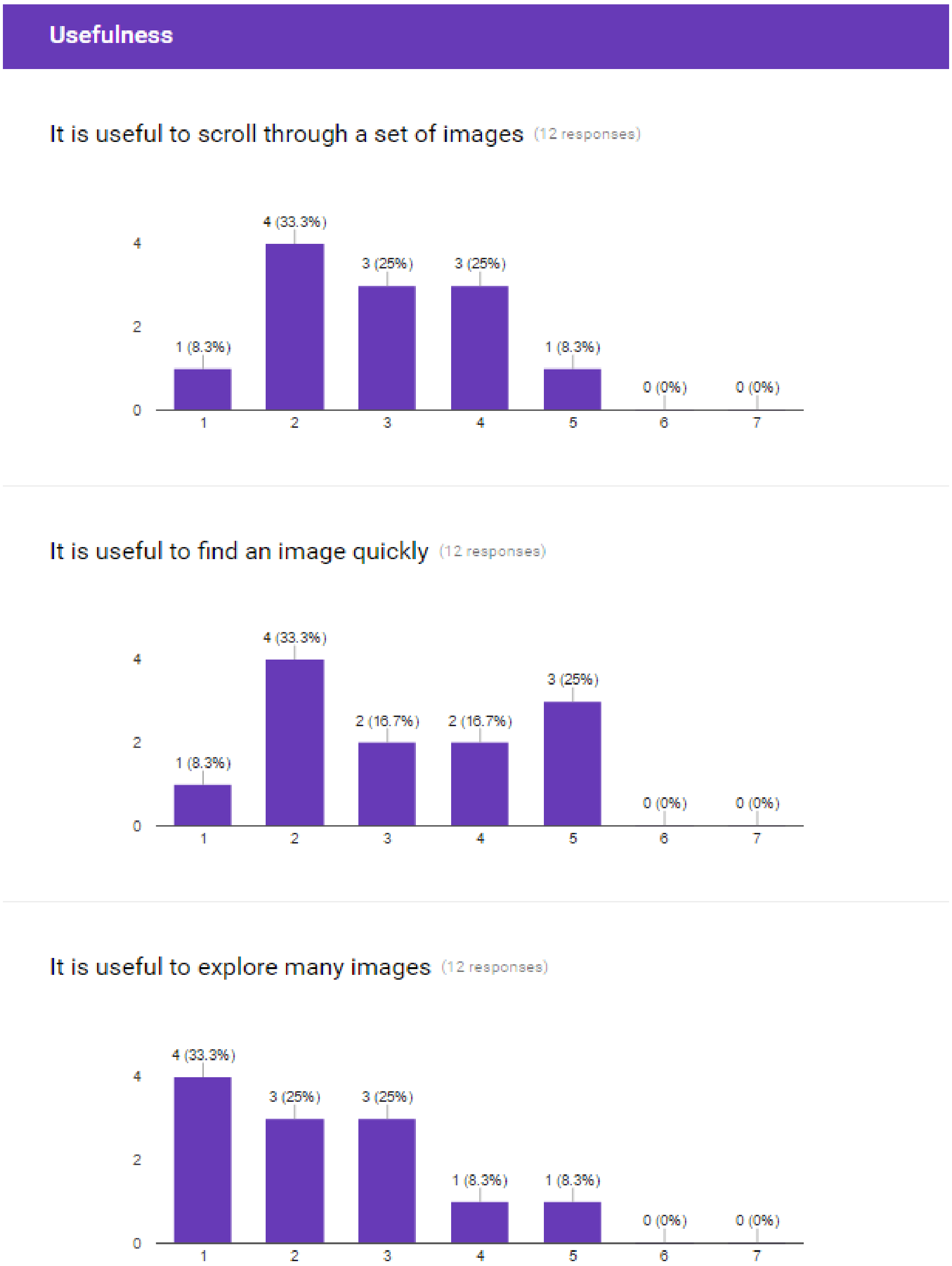


FIGURE A.12: Usefulness: Question 1, 2 and 3.

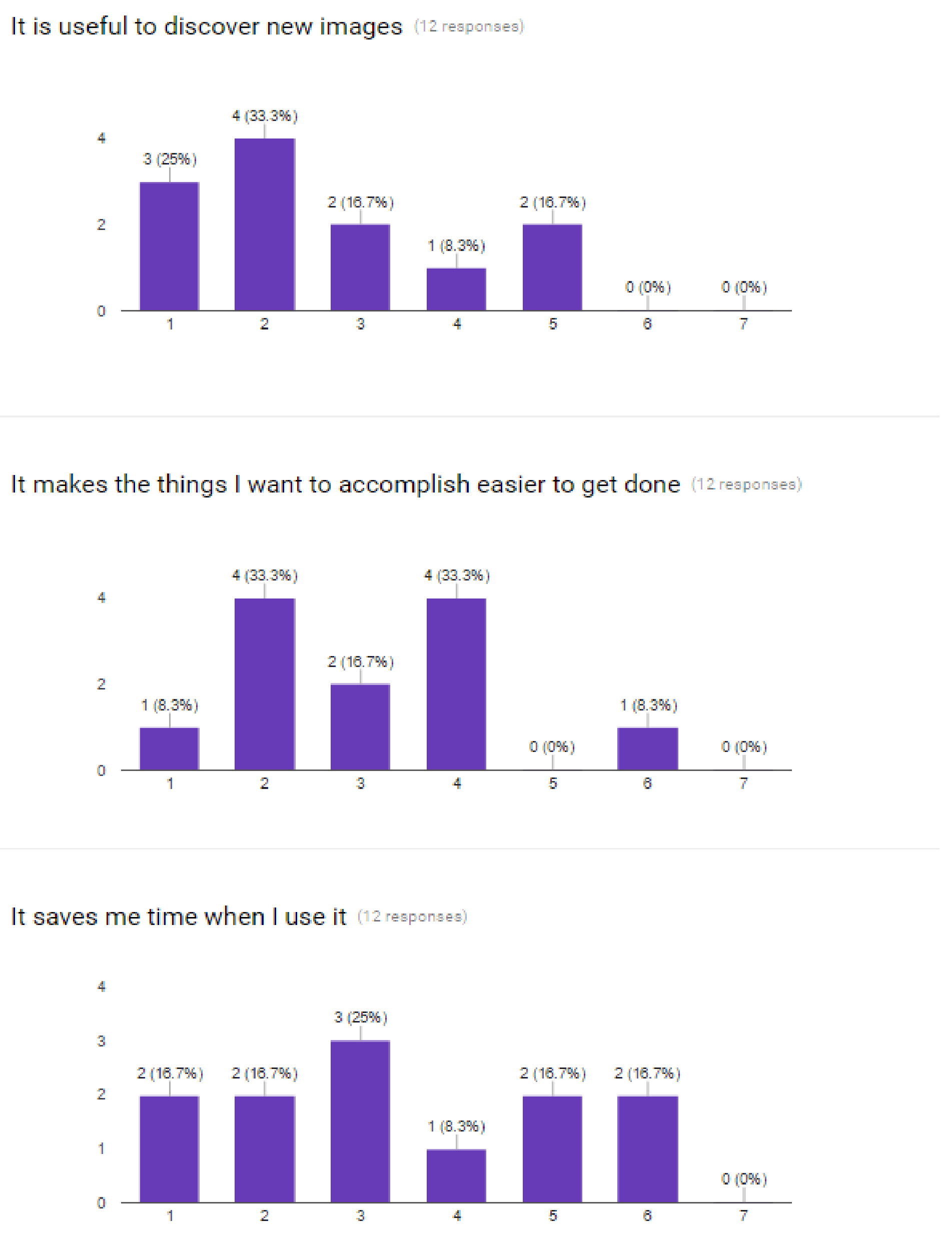


FIGURE A.13: Usefulness: Question 4, 5 and 6.

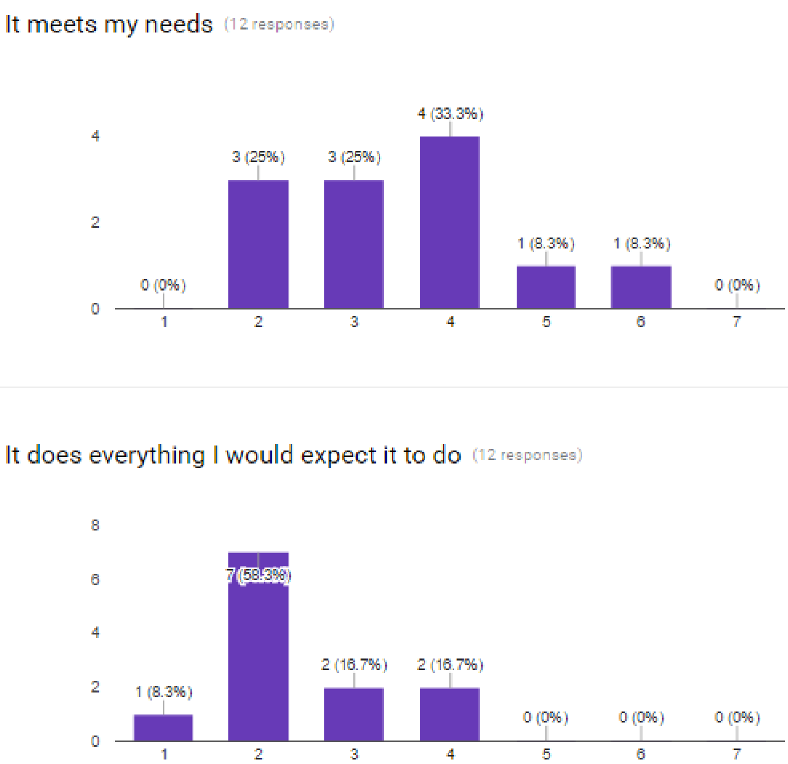


FIGURE A.14: Usefulness: Question 7 and 8.

FIGURE A.15: Ease of use: Question 9, 10 and 11.

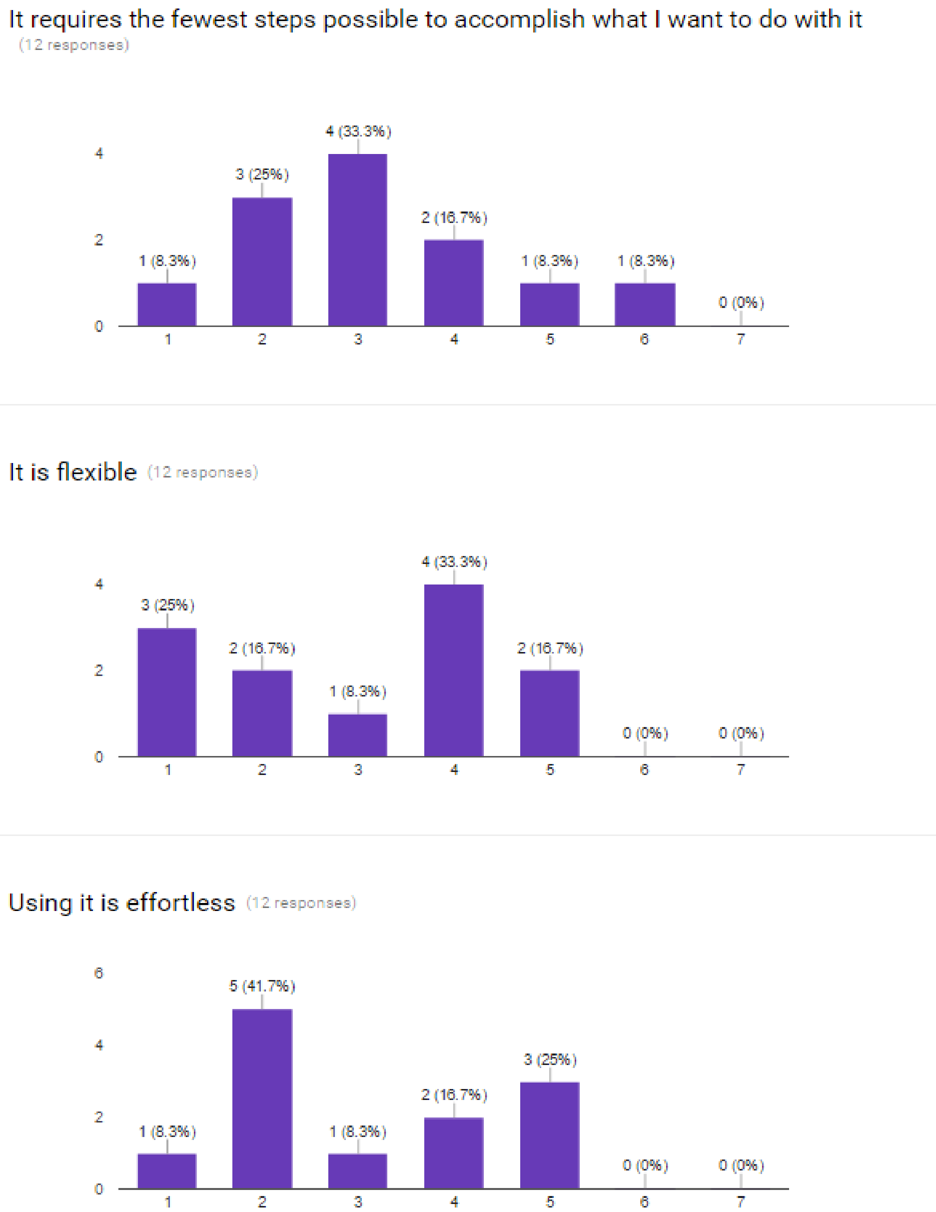


FIGURE A.16: Ease of use: Question 12, 13 and 14.

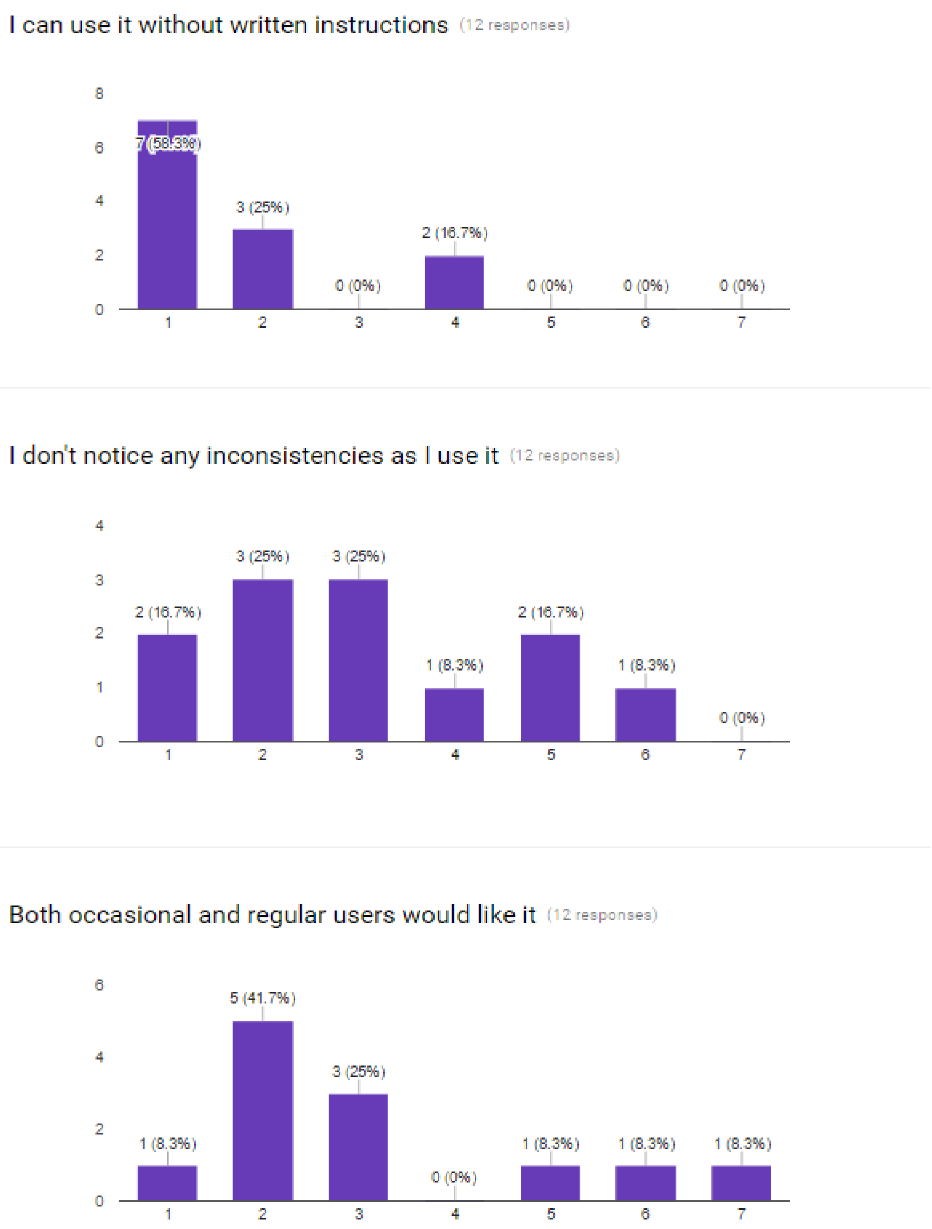
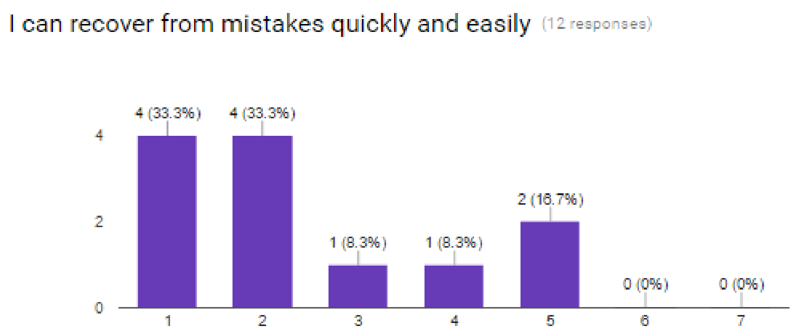


FIGURE A.17: Ease of use: Question 15, 16 and 17.



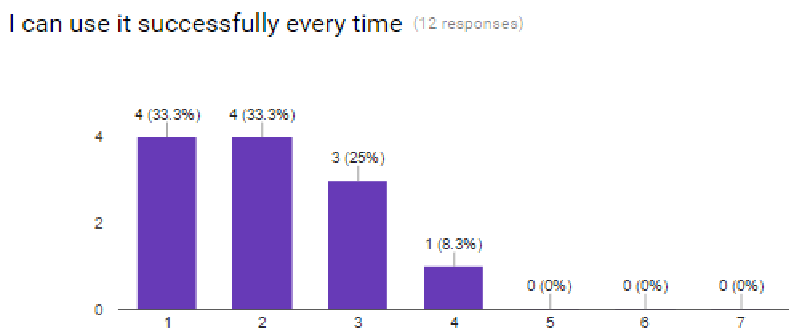


FIGURE A.18: Ease of use: Question 18 and 19.

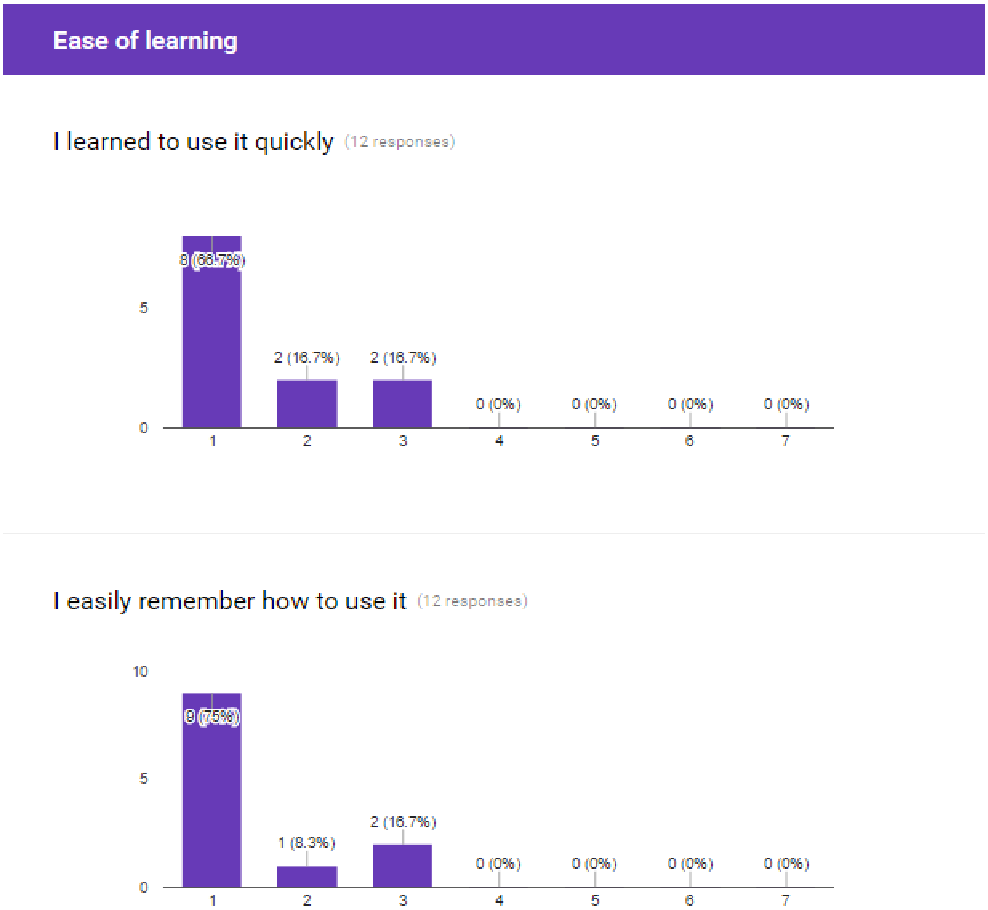


FIGURE A.19: Ease of learning: Question 20 and 21.

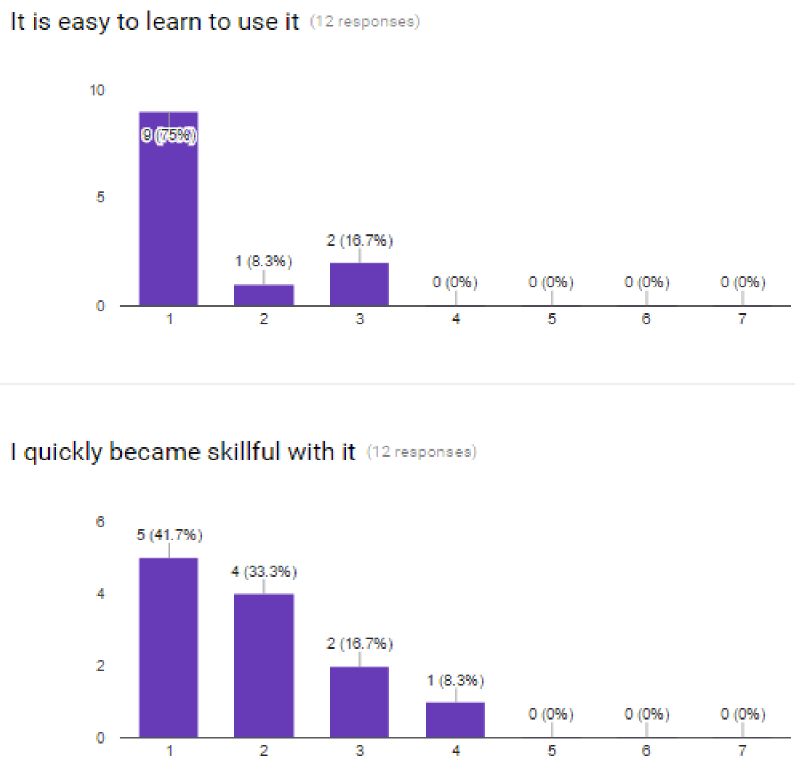


FIGURE A.20: Ease of learning: Question 22 and 23.

FIGURE A.21: Satisfaction: Question 24, 25 and 26.

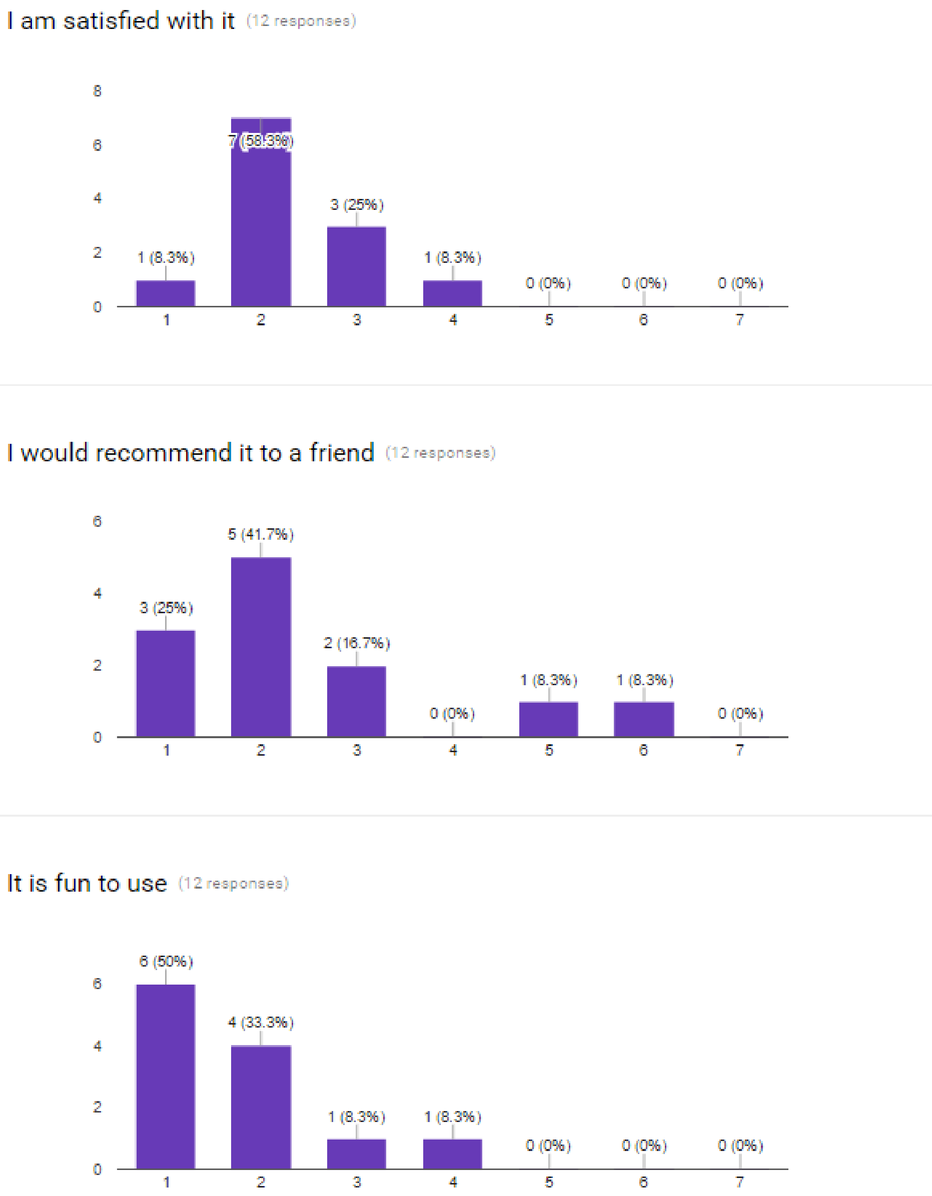
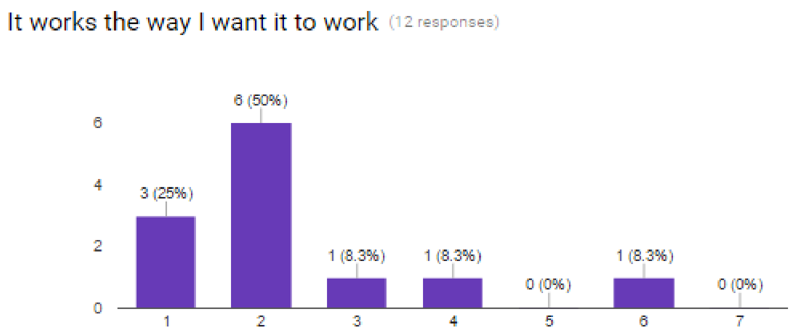


FIGURE A.22: Satisfaction: Question 27, 28 and 29.



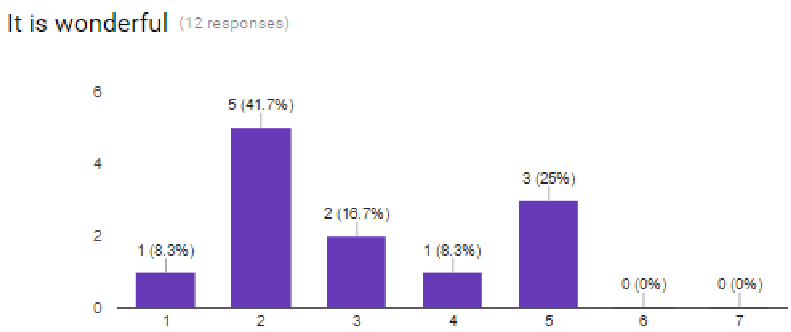


FIGURE A.23: Satisfaction: Question 30 and 31.

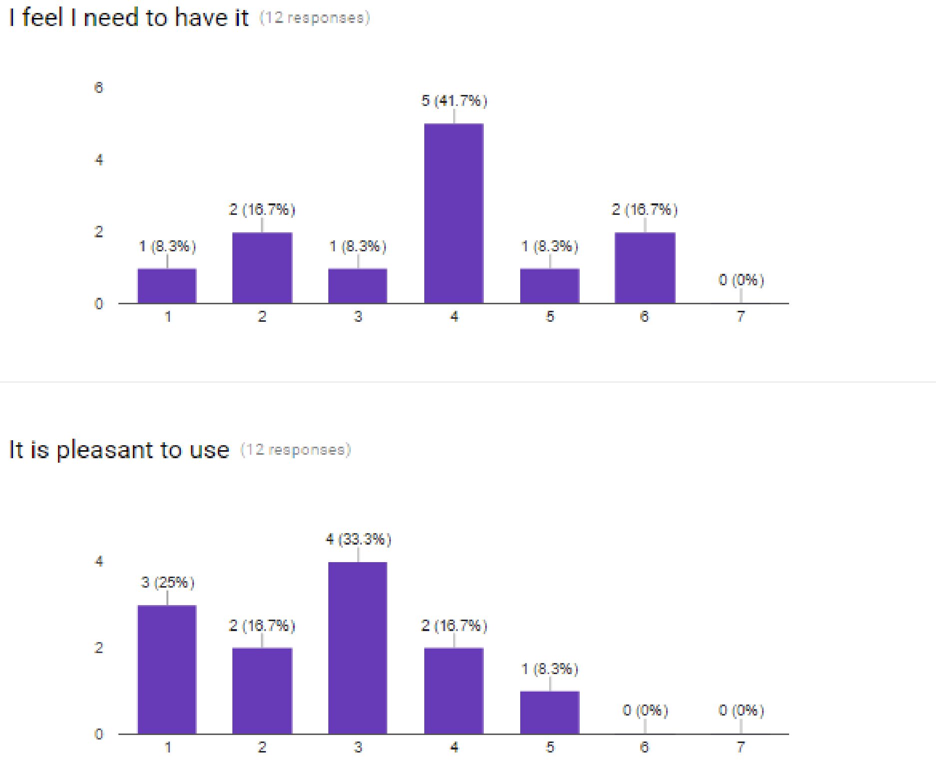


FIGURE A.24: Satisfaction: Question 32 and 33.

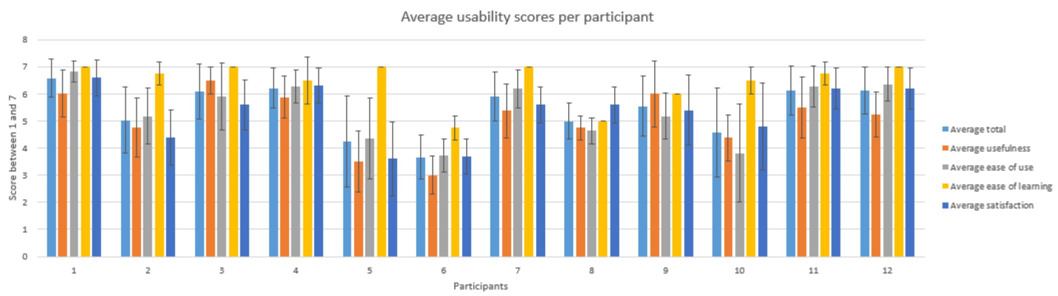


FIGURE A.25: Average scores per participant for each of the usability aspects.

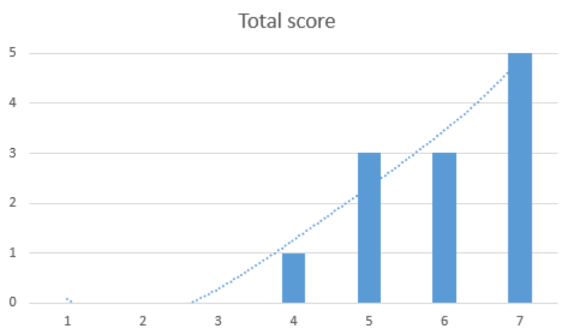


FIGURE A.26: Distribution of the average total usability scores.

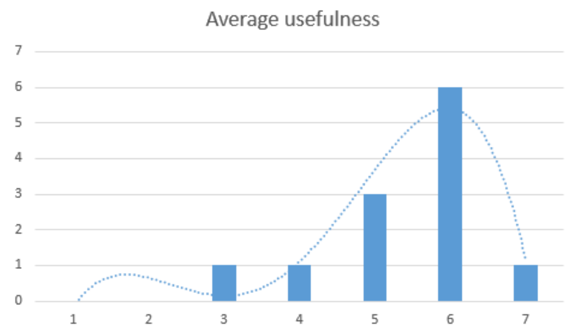


FIGURE A.27: Distribution of the average usefulness scores.

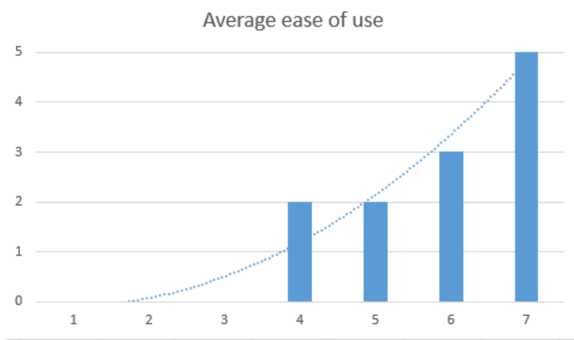


FIGURE A.28: Distribution of the average ease of use scores.

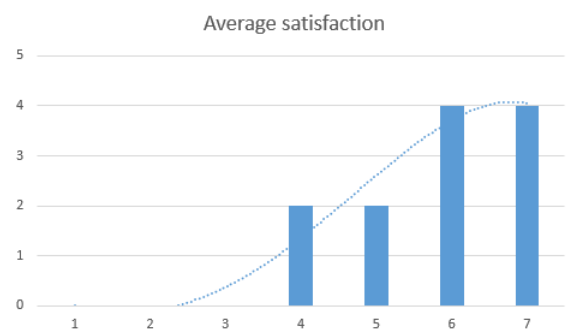


FIGURE A.29: Distribution of the average satisfaction scores.

Bibliography

Alonso-Calvo, Raul et al. (2010). "On distributing load in cloud computing: A real application for very-large image datasets". In: *Procedia Computer Science* 1.1, pp. 2669–2677.

André, Paul et al. (2009). "Designing novel image search interfaces by understanding unique characteristics and usage". In: *Human-Computer Interaction–INTERACT 2009*. Springer, pp. 340–353.

Bailey, Kat (2016). *Opinion: Why I Remain a VR Skeptic*. URL: http://www.usgamer.net/articles/opinion-why-i-remain-a-vr-skeptic.

Bates, Cary Lee and Paul Reuben Day (2004). *Automatic color contrast adjuster*. US Patent 6,809,741.

Bateson, Mary Catherine (1994). *Peripheral visions: Learning along the way*. HarperCollins.

Bernd, Julia et al. (2015). "Kickstarting the Commons: The YFCC100M and the YLI Corpora". In: *Proceedings of the 2015 Workshop on Community-Organized Multimodal Mining: Opportunities for Novel Solutions*. ACM, pp. 1–6.

Bi, Chunyue (2009). "Research and application of SQLite embedded database technology". In: *wseas transactions on computers* 8.1, pp. 83–92.

Biguer, B, M Jeannerod, and C Prablanc (1982). "The coordination of eye, head, and arm movements during reaching at a single visual target". In: *Experimental Brain Research* 46.2, pp. 301–304.

Borup, Mads et al. (2006). "The sociology of expectations in science and technology". In: *Technology analysis & strategic management* 18.3-4, pp. 285–298.

Bowman, Doug A and Ryan P McMahan (2007). "Virtual reality: how much immersion is enough?" In: *Computer* 40.7, pp. 36–43.

Brigner, Willard L and M Brooks Gallagher (1974). "Subjective contour: apparent depth or simultaneous brightness contrast?" In: *Perceptual and Motor Skills* 38.3 suppl, pp. 1047–1053.

Bruggink, Jan-Willem (2013). *Ruim 6 op de 10 mensen dragen een bril of contactlenzen*.

Buering, Thorsten, Jens Gerken, and Harald Reiterer (2006). "User interaction with scatterplots on small screens-a comparative evaluation of geometric-semantic zoom and fisheye distortion". In: *IEEE Transactions on Visualization and Computer Graphics* 12.5, pp. 829–836.

Buffart, Hans and Emanuel Leeuwenberg (1983). "Structural information theory". In: *Advances in psychology* 11, pp. 48–72.

Burdea, Grigore and Philippe Coiffet (2003). "Virtual reality technology". In: *Presence: Teleoperators and virtual environments* 12.6, pp. 663–664.

Byrne, Patrick, Suzanna Becker, and Neil Burgess (2007). "Remembering the past and imagining the future: a neural model of spatial memory and imagery." In: *Psychological review* 114.2, p. 340.

Carlton, Les G (1981). "Processing visual feedback information for movement control." In: *Journal of Experimental Psychology: Human Perception and Performance* 7.5, p. 1019.

Chen, Hsin-liang (2001). "An analysis of image queries in the field of art history". In: *Journal of the American Society for Information Science and Technology* 52.3, pp. 260–273.

Chen, Hsinchun et al. (1998). "Internet browsing and searching: User evaluation of category map and concept space techniques". In: *Journal of the American Society for Information Science, Special Issue on AI Techniques for Emerging Information Systems Applications*.

Chen, Mon Chu, John R Anderson, and Myeong Ho Sohn (2001). "What can a mouse cursor tell us more?: correlation of eye/mouse movements on web browsing". In: *CHI'01 extended abstracts on Human factors in computing systems*. ACM, pp. 281–282.

Cobb, Sue VG et al. (1999). "Virtual reality-induced symptoms and effects (VRISE)". In: *Presence* 8.2, pp. 169–186.

Cockburn, Andy and Joshua Savage (2004). "Comparing speed-dependent automatic zooming with traditional scroll, pan and zoom methods". In: *People and Computers XVII?Designing for Society*. Springer, pp. 87–102.

Cockburn, Andy et al. (2011). "Air pointing: Design and evaluation of spatial target acquisition with and without visual feedback". In: *International Journal of Human-Computer Studies* 69.6, pp. 401–414.

Coren, Stanley (1972). "SUBJECTIVE CONTOURS AND APPARENT DEPTH1". In: *Psychological review* 79.4, 3S9–367.

Cruz-Neira, Carolina, Daniel J Sandin, and Thomas A DeFanti (1993). "Surround-screen projection-based virtual reality: the design and implementation of the CAVE". In: *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*. ACM, pp. 135–142.

Cutting, James E (1997). "How the eye measures reality and virtual reality". In: *Behavior Research Methods, Instruments, & Computers* 29.1, pp. 27–36.

Czerwinski, Mary et al. (1999). "The contribution of thumbnail image, mouse-over text and spatial location memory to web page retrieval in 3D". In: *Proceedings of INTERACT*. Vol. 99, pp. 163–170.

Datta, Ritendra, Jia Li, and James Z Wang (2005). "Content-based image retrieval: approaches and trends of the new age". In: *Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval*. ACM, pp. 253–262.

Davidson, James et al. (2010). "The YouTube video recommendation system". In: *Proceedings of the fourth ACM conference on Recommender systems*. ACM, pp. 293–296.

Dziadosz, Susan and Raman Chandrasekar (2002). "Do thumbnail previews help users make better relevance decisions about web search results?" In: *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, pp. 365–366.

Elliott, Digby and Fran Allard (1985). "The utilization of visual feedback information during rapid pointing movements". In: *The Quarterly Journal of Experimental Psychology* 37.3, pp. 407–425.

Enser, Peter GB (1993). "Query analysis in a visual information retrieval context". In: *Journal of Document and Text Management* 1.1, pp. 25–52.

Enser, Peter GB and Colin G McGregor (1993). *Analysis of visual information retrieval queries*. British Library Board London.

Epstein, Russell A (2008). "Parahippocampal and retrosplenial contributions to human spatial navigation". In: *Trends in cognitive sciences* 12.10, pp. 388–396.

Fisher, Ronald A (1956). "Statistical methods and scientific inference." In:

Furnas, George W (1986). *Generalized fisheye views*. Vol. 17. 4. ACM.

Gallaga, Omar (2015). *Skeptics and obstacles might not stop virtual reality*. URL: http://www.mystatesman.com/news/technology/skeptics-and-obstacles-might-not-stop-virtual-real/npHmm/.

Gigerenzer, Gerd (2004). "Mindless statistics". In: *The Journal of Socio-Economics* 33.5, pp. 587–606.

Goldman, Alvin I (2015). *Theory of human action*. Princeton University Press.

Graphics_Dev (2016). *Unity Answers Statistics*. URL: http://answers.unity3d.com/questions/1233252/unity-answers-statistics.html.

Grossberg, Stephen and Ennio Mingolla (1985). "Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading." In: *Psychological review* 92.2, p. 173.

Gu, Yunhong and Robert L Grossman (2009). "Sector and Sphere: the design and implementation of a high-performance data cloud". In: *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 367.1897, pp. 2429–2445.

Hettinger, Lawrence J and Gary E Riccio (1992). "Visually induced motion sickness in virtual environments". In: *Presence: Teleoperators & Virtual Environments* 1.3, pp. 306–310.

Hollands, JG et al. (1989). “Presenting a graphical network: A comparison of performance using fisheye and scrolling views”. In: *Proceedings of the third international conference on human-computer interaction on Designing and using human-computer interfaces and knowledge based systems (2nd ed.)* Elsevier Science Inc., pp. 313–320.

Igarashi, Takeo and Ken Hinckley (2000). “Speed-dependent automatic zooming for browsing large documents”. In: *Proceedings of the 13th annual ACM symposium on User interface software and technology*. ACM, pp. 139–148.

Jäkel, Frank et al. (2016). “An overview of quantitative approaches in Gestalt perception”. In: *Vision Research* 126, pp. 3–8.

Jansen, Bernard J (2008). “Searching for digital images on the web”. In: *Journal of Documentation* 64.1, pp. 81–101.

Jensen, JF et al. (2002). *Virtual Space: spatiality in virtual inhabited 3D worlds*. Springer Science & Business Media.

Jeon, Seokhee, Hyeongseop Shim, and Gerard Jounghyun Kim (2006). “Viewpoint Usability for Desktop Augmented Reality.” In: *IJVR* 5.3, pp. 33–39.

Kahneman, Daniel (1973). *Attention and effort*. Citeseer.

Kaplan, Ira T and WN Schoenfeld (1966). “Oculomotor patterns during the solution of visually displayed anagrams.” In: *Journal of Experimental Psychology* 72.3, p. 447.

Kaufmann, Hannes and Andreas Dünser (2007). “Summary of usability evaluations of an educational augmented reality application”. In: *International Conference on Virtual Reality*. Springer, pp. 660–669.

Keay, Lisa et al. (2010). “Predictors of early acceptance of free spectacles provided to junior high school students in China”. In: *Archives of ophthalmology* 128.10, pp. 1328–1334.

Keister, Lucinda H (1994). “User types and queries: impact on image access systems”. In: *Challenges in indexing electronic text and images*, pp. 7–22.

Keshavarz, Behrang et al. (2015). “Vection and visually induced motion sickness: how are they related?” In: *Frontiers in psychology* 6, p. 472.

Khanwalkar, Sanket, Shonali Balakrishna, and Ramesh Jain (2016). “Exploration of Large Image Corpuses in Virtual Reality”. In: *Proceedings of the 2016 ACM on Multimedia Conference*. ACM, pp. 596–600.

Kleiman, Yanir et al. (2015). “DynamicMaps: Similarity-based Browsing through a Massive Set of Images”. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, pp. 995–1004.

Kober, Silvia Erika and Christa Neuper (2013). “Personality and presence in virtual reality: does their relationship depend on the used presence measure?” In: *International Journal of Human-Computer Interaction* 29.1, pp. 13–25.

Kofler, Christoph, Martha Larson, and Alan Hanjalic (2016). "User Intent in Multimedia Search: A Survey of the State of the Art and Future Challenges". In: *ACM Computing Surveys (CSUR)* 49.2, p. 36.

Law, Effie Lai-Chong et al. (2009). "Understanding, scoping and defining user experience: a survey approach". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, pp. 719–728.

Lettvin, Jerome Y (1976). "On seeing sidelong". In: *The Sciences* 16.4, pp. 10–20.

Li, Liping et al. (2008). "Spectacle acceptance among secondary school students in rural China: the Xichang Pediatric Refractive Error Study (X-PRES)—report 5". In: *Investigative ophthalmology & visual science* 49.7, pp. 2895–2902.

Liu, Hao et al. (2004). "Effective browsing of web image search results". In: *Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval*. ACM, pp. 84–90.

Lund, Arnold M (2001). "Measuring Usability with the USE Questionnaire12". In:

Marchionini, Gary (2006). "Exploratory search: from finding to understanding". In: *Communications of the ACM* 49.4, pp. 41–46.

Markkula, Marjo and Eero Sormunen (2000). "End-user searching challenges indexing practices in the digital newspaper photo archive". In: *Information retrieval* 1.4, pp. 259–285.

Meeker, Mary (2016). *Internet Trends 2016*. Code Conference. URL: http://www.kpcb.com/internet-trends.

Microsoft, MSDN (2016). *Stopwatch.ElapsedMilliseconds Property*. URL: https://msdn.microsoft.com/en-us/library/system.diagnostics.stopwatch.elapsedmilliseconds.aspx.

Moorrees, Coenraad FA and Martin R Kean (1958). "Natural head position, a basic consideration in the interpretation of cephalometric radiographs". In: *American Journal of Physical Anthropology* 16.2, pp. 213–234.

Nace, Trevor (2016). *4 hurdles that keep virtual reality from being mainstream*. URL: http://thenextweb.com/insider/2016/05/20/hurdles-virtual-reality-mainstream/.

Nurmi, Daniel et al. (2009). "The eucalyptus open-source cloud-computing system". In: *Cluster Computing and the Grid, 2009. CCGRID'09. 9th IEEE/ACM International Symposium on*. IEEE, pp. 124–131.

Oculus VR, LLC. Part of Facebook (2016). *Introduction to Best Practices*. URL: https://developer3.oculus.com/documentation/intro-vr/latest/concepts/bp_intro/.

Ornager, Susanne (1995). "The newspaper image database: empirical supported analysis of users' typology and word association clusters". In: *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, pp. 212–218.

O'BRIEN, VIVIAN (1958). "Contour perception, illusion and reality". In: *JOSA* 48.2, pp. 112–119.

Perlman, G. (2015). *User Interface Usability Evaluation with Web-Based Questionnaires*. URL: http://garyperlman.com/quest/.

Peters, Richard Stanley (2015). *The concept of motivation*. Routledge.

Pierce, Jeffrey S et al. (1997). "Image plane interaction techniques in 3D immersive environments". In: *Proceedings of the 1997 symposium on Interactive 3D graphics*. ACM, 39–ff.

Ramachandran, Vilayanur S, Richard L Gregory, et al. (1991). "Perceptual filling in of artificially induced scotomas in human vision". In: *Nature* 350.6320, pp. 699–702.

Raskin, Jef (1994). "Intuitive equals familiar". In: *Communications of the ACM* 37.9, pp. 17–19.

Rodden, Kerry et al. (2001). "Does organisation by similarity assist image browsing?" In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, pp. 190–197.

Sarkar, Manojit and Marc H Brown (1994). "Graphical fisheye views". In: *Communications of the ACM* 37.12, pp. 73–83.

Schaffer, Doug et al. (1996). "Navigating hierarchically clustered networks through fisheye and full-zoom methods". In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 3.2, pp. 162–188.

Schuemie, Martijn J et al. (2001). "Research on presence in virtual reality: A survey". In: *CyberPsychology & Behavior* 4.2, pp. 183–201.

Slater, Mel (1999). "Measuring presence: A response to the Witmer and Singer presence questionnaire". In: *Presence: Teleoperators and Virtual Environments* 8.5, pp. 560–565.

Slater, Mel and Sylvia Wilbur (1997). "A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments". In: *Presence: Teleoperators and virtual environments* 6.6, pp. 603–616.

Smith, Craig (2016). *By the numbers: 14 interesting Flickr stats*. URL: http://expandedramblings.com/index.php/flickr-stats/.

Statista (2016a). *Amount willing to pay for virtual reality headset in the United States 2015*. URL: https://www.statista.com/statistics/457117/virtual-reality-headset-amount-willing-to-pay-in-the-united-states/.

— (2016b). *Willingness to buy a virtual reality headset in the United States 2015*. URL: https://www.statista.com/statistics/457079/virtual-reality-headset-willingness-to-buy-in-the-united-states/.

Strasburger, Hans, Ingo Rentschler, and Martin Jüttner (2011). "Peripheral vision and pattern recognition: A review". In: *Journal of vision* 11.5, pp. 13–13.

Teather, Robert J and Wolfgang Stuerzlinger (2011). "Pointing at 3D targets in a stereo head-tracked virtual environment". In: *3D User Interfaces (3DUI), 2011 IEEE Symposium on*. IEEE, pp. 87–94.

Thomee, Bart et al. (2016). "YFCC100M: The new data in multimedia research". In: *Communications of the ACM* 59.2, pp. 64–73.

Toates, FM (1972). "Accommodation function of the human eye". In: *Physiological Reviews* 52.4, pp. 828–863.

Unity Technologies (2016). *Company Facts*. URL: https://unity3d.com/public-relations.

Van Tuijl, HFJM (1975). "A new visual illusion: Neonlike color spreading and complementary color induction between subjective contours". In: *Acta psychologica* 39.6, 441–IN1.

Vermeeren, Arnold POS et al. (2010). "User experience evaluation methods: current state and development needs". In: *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*. ACM, pp. 521–530.

Ware, Colin, Kevin Arthur, and Kellogg S Booth (1993). "Fish tank virtual reality". In: *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*. ACM, pp. 37–42.

Witmer, Bob G and Michael J Singer (1998). "Measuring presence in virtual environments: A presence questionnaire". In: *Presence: Teleoperators and virtual environments* 7.3, pp. 225–240.

Wu, Jiyi et al. (2010). "Cloud storage as the infrastructure of cloud computing". In: *Intelligent Computing and Cognitive Informatics (ICICCI), 2010 International Conference on*. IEEE, pp. 380–383.

Yee, Ka-Ping et al. (2003). "Faceted metadata for image search and browsing". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, pp. 401–408.

Zhou, Renjie, Samamon Khemmarat, and Lixin Gao (2010). "The impact of YouTube recommendation system on video views". In: *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*. ACM, pp. 404–410.

Zhou, Renjie et al. (2016). "How YouTube videos are discovered and its impact on video views". In: *Multimedia Tools and Applications*, pp. 1–24.