VR-LIVE: Virtual Reality Lifelog Image Visualization and Exploration

A Comparative Study and Proof-of-Concept System

Yannick Visser - 6585124
Supervisor and first examiner - W.O. Hürst
Second examiner - M. Behrisch

July 3, 2024

Utrecht University

Master Thesis Game and Media Technology
Graduate School of Natural Sciences
Utrecht University
The Netherlands

Abstract

This study introduces VR-LIVE to address challenges in navigating and experiencing Lifelog Image Databases (LIDs) efficiently and immersively. To come to VR-LIVE, three visualizations for lifelog snippets in VR are designed, implemented, and evaluated: GIF, 2D map, and 3D map. Participants rated their confidence in understanding activities and movements in each visualization, with the 3D map perceived as the most immersive. The GIF visualization was found to be the most efficient for quickly looking over the lifelog images to get an idea of what the lifelogger was doing. The advantages and disadvantages of each visualization were identified. A proof-of-concept system, VR-LIVE, was designed and realized with the gained insights about the visualizations. In general, the study provides valuable information on virtual reality life-log image visualization techniques.

Contents

1 Introduction

Large image databases have become more and more common over the years. This is an immediate effect of the increase in the number of mobile devices worldwide [1] and the increasing quality of the camera [2] on board the mobile devices. Large image databases can arise from mobile devices such as smartphones, but can also be the result of devices like wearable cameras (GoPro's, ActionCams, Insta360's) or lifelogging devices. These devices create large image databases in a very short time, because of the high frequency of capturing images.

Regular solutions for image browsing are struggling to keep up with the sheer volume and specific quirks of lifelogging images. While lifelogging has existed in its current form since the Microsoft Sensecam was patented in 2009 [3], research into lifelog imaging has been conducted since the 1980s [4]. Lifelog images are not your typical snapshots; constant automatic capture means that we end up with loads of pictures that are awkwardly framed, blurred, or taken in less than ideal lighting [5]. Managing a (possible) lifetime of data results in sets of Lifelog Image Databases (LIDs) growing to enormous proportions; as these records increase in size, the problem of effectively managing them becomes more difficult [6]. Traditional 2D visualizations just can't handle the scale and unique characteristics of lifelogging images.

Both handheld and non-handheld computing devices commonly employ a Nested Folder Hierarchy to facilitate the organization of diverse data types, such as images, files, or folders [7]. This hierarchical structure serves the dual purpose of allowing users to manually structure their data, involving tasks such as creating folders and categorizing data within them, and also enabling automatic data organization, where folders are generated based on temporal data and other parameters. The Nested Folder Hierarchy model is conceptually akin to a traditional cabinet system, wherein each cabinet may contain distinct folders housing various files. This approach provides a familiar framework for users, aligning digital concepts with conventional storage practices.

However, the Nested Folder Hierarchy encounters challenges when confronted with extensive datasets. The process of subdividing data becomes intricate, resulting in a folder hierarchy with significant depth. This depth poses difficulties for users attempting to navigate the hierarchy to locate specific files, rendering the system less intuitive and more perplexing. Besides the depth-challenges, traditional devices impose screen constraints. The device screen is a limited space, and can only display a limited amount of data at the same time. This poses navigational issues similar to the depth challenge.

The scalability issues of the traditional Nested Folder Hierarchy become apparent, particularly in scenarios involving large amounts of data (like LIDs), necessitating the exploration of alternative organizational paradigms for improved usability and efficiency. Retrieval and indexing of large amount of images has been researched [8].

Many total capture systems inherently address the aspect of recollection or remembering personal experiences [9]. Psychological literature robustly supports the connection between autobiographical memories and visual images [10]. To effectively connect with these memories, the viewing experience of lifelog images should be immersive. Virtual reality (VR) shows great promise in delivering this immersive experience, enabling individuals to re-experience moments captured in their lifelogs.

We introduce three innovative strategies to tackle the challenges posed by traditional 2D gallery visualizations for LIDs. The challenges to be tackled are the scalability and the immersiveness of the lifelog images. We do this by implementing three visualizations in VR: a GIF visualization, a 2D map visualization, and a 3D map visualization. These visualizations are tested and evaluated by a user study. Drawing conclusions from this user study, a proof-of-concept system, VR-LIVE, has been designed and developed. This proof-of-concept system consists of a GIF wall, where GIFs are constructed based on temporal filters applied to the lifelog images. In order to observe events conveniently and immersively, a 3D Street View tour is implemented. This system is designed with insights gained from the preliminary study for the three beforementioned visualizations. The major contributions of this study are:

- The design, implementation, and evaluation of three non-trivial lifelog image visualizations in VR.

  By designing, implementing, and evaluating the three proposed visualizations (GIF, 2D map, 3D map), this study obtains insight into the efficiency and immersiveness of viewing lifelog images in these visualizations.

  These visualizations are evaluated by 19 participants in a user study. The insights obtained during this user study will lay the foundation for the proof-of-concept system realized in this study.

- The proof-of-concept implementation of a system using insights gained during the evaluation of the different visualizations.

  This study implements a proof-of-concept system, in which insights gained during the evaluation of the different visualizations are utilized to come to a system. This system, called VR-LIVE, is built up out of two parts; a GIF wall and a 3D Street View Tour.

  Implementing a GIF wall for lifelogging images in VR is a non-trivial task. The complexity of choosing the right parameters to ensure a solid representation of lifelogging images poses multiple challenges and requires a delicate design and a well-optimized system to be an effective way of viewing lifelog images. A GIF representation has been proven to be an effective way to view lifelog images in VR in the user study conducted in this research.

  A VR 3D Street View Tour has been implemented that enables the user to experience lifelog images in an immersive way, as proven by the conducted evaluation of the three different visualizations. The user will be able to experience lifelogging images within the 3D VR environment, where the user is taken along the locations where the lifelog images were taken.

  Implementing a VR Street View Tour is a non-trivial task. Geo-tagging and spatial alignment of the images is a difficult task, especially when ensuring user experience. With a lot of parameters at hand, tweaking the Street View Tour will be a sensitive topic.

2 Related Work

The dynamic landscape of image database visualization has been the subject of extensive research, reflecting the evolving nature of technology and user preferences. As we embark on a comparative analysis between temporal filtering visualizations and geospatial filtering interfaces, it is imperative to contextualize our study within the broader framework of existing literature. To obtain an optimal LID visualization within VR, we need to understand the following topics:

- Lifelogging
- Virtual Reality
- Current implementations and solutions
  - Geospatial implementations
  - Temporal implementations
  - Combined
- Conceptual interpertations

The current implementations will be divided into the sections temporal filtering visualizations and geospatial filtering visualizations, and a single implementation that attempts to combine the two. Of course, these are not the only VR lifelogging implementations; however, it is important to keep the scope of the research in mind. When we have gained an understanding of these topics, we can uncover open research gaps, and cover these with our new and refined approach.

2.1 Lifelogging

A lifelog represents a personal log of one's daily existence [11]. Individuals engaged in the act of documenting their lives are often called lifeloggers, and the undertaking itself is known as lifelogging. This can be done automatically and digitally, or in simpler forms, such as a diary. Gurrin, Smeaton and Doherty [12] give the following definition for lifelogging: *"the process of passively gathering, processing, and reflecting on life experience data collected by a variety of sensors, and is carried out by an individual, the lifelogger"*. Dodge and Kitchin refined the definition of lifelogging to emphasize more on its large scope. They described it as *"a form of pervasive computing, comprising a consolidated digital account encompassing all aspects of an individual's experiences. This is achieved through various digital sensors, with the resulting data stored permanently as a personal multimedia archive.* [13]

Typically, lifelogging data consists of images automatically captured at regular intervals, such as every 1-3 minutes, using small body-worn cameras. Lifelogging this way is called visual lifelogging and has been the subject of a considerable amount of research [14]. Attributes like GPS location and visual properties of the photos are used as tools to sort within a large dataset. Various applications have been developed to provide users with easy access to lifelog data [15] [16] [17] [18]. These systems generally aim to enhance either the user experience for exploration or the efficiency of queries for retrieval.

Systems focusing on the latter use different techniques, ranging from interface design to back-end systems derived from multimedia/video retrieval. These systems compete annually in the Lifelog Search Challenge (LSC)[19], where each implementation is tasked with efficiently resolving a set of retrieval tasks. The LSC serves as a valuable comparison for state-of-the-art retrieval systems, although it primarily emphasizes search time efficiency rather than user experience and discovery of unknown data.

Lifelogging encompasses more than merely capturing lifelog images with a lifelog camera. In fact, there are numerous types of data that can be recorded during lifelogging, as outlined by Machajdik et al. [20].

- Passive Visual Capture (data in the form of images captured by an ”always on” camera)
- Biometrics (heart rate, skin temperature, body motion, etc)
- Mobile context (GPS data, GSM location data, wireless network data)
- Mobile activity (Call logs, SMSes, email-logs, web activity)
- Desktop computer activity (All computer activity)
- Active capture (written blogs, actively taken pictures, diaries)

To understand what makes a novel lifelogging visualization, an understanding of the psychology behind the need or want to log personal life data. The initial advantage of a lifelogging system on human memory lies in its ability to facilitate the simple act of recalling or reliving specific life experiences. This process involves reflecting in detail on past experiences, commonly known as episodic memory [21]. This form of recollection allows individuals to mentally retrace their steps, proving valuable for practical purposes such as locating lost items or remembering faces and names by revisiting the context of when and where a meeting occurred. According to Sellen et al. [9], many total capture systems inherently address the aspect of recollection or remembering personal experiences. Notably, there exists well-established psychological literature indicating a robust connection between autobiographical memories and visual images [22].

Czerwinski et al. [23] suggest that individuals can benefit from the memory-supporting features of a lifelog, for example for reviewing past events and reflecting on them, or finding lost objects.

Another purpose that can be identified is for people to engage in revisiting their life experiences. However, in this context, the focus is on emotional or sentimental reflection rather than practical considerations.

Other research [24] [25] in regard of the motivation to lifelog indicates several lifelog purposes:

- Reminiscing: wanting to relive the past or review interesting events
- Memory backup: using lifelog as reminders.
- Telling and passing life stories: passing on stories to loved ones
- Re-use: lifelog data being saved so it can be reused

- Evidence: keeping lifelog data to use them as evidence in court
- Collection and archiving: the fun of collecting can be a motivation to capture lifelog images
- Learning about unknown early age: collecting memories from the early age
- Well-being and better organization: capturing lifelog data in order to notice bad habits or patterns.

Although most research focuses on the retrieval of data [26], this study will focus on the exploration of data.

2.2 Virtual Reality

Over the course of documented history, humanity has devised methods to materialize fanciful realms and narratives into our tangible world, through art, literature, and, in the past century, digital media [27]. From the art of storytelling to the advent of modern virtual reality (VR) devices, creators and inventors have sought to captivate their audiences by transporting them into alternate realities, aiming to instill a profound sense of presence and immersion [27]. VR has shown to be beneficial in various fields and applications. Various applications have showcased the versatility of virtual reality. While commonly associated with entertainment, such as in video games or movies [28], it is essential to recognize that its utility extends beyond this realm. Virtual reality has repeatedly proven its value in diverse industries, including simulations, medical fields, and engineering [29]. The most popular and commercialized form of VR are head-mounted displays (HMDs). These are goggles designed to be worn on the head. The integrated displays enables standalone applications to be played directly on the headset. Additionally, the headset features onboard memory capable of storing data, including information acquired during lifelogging. Generally, VR headsets employ a different screen for each eye, enabling a stereoscopic effect. This allows for a higher perception of environmental depth [30].

Virtual reality presents opportunities that extend into the realm of research. The technology has demonstrated a novel approach to exploring data [31], leveraging its additional dimension and increased immersion. VR holds the potential to facilitate a more intuitive visual comprehension of research data, proving particularly advantageous for larger datasets [32].

In addition to general use cases, it is important to study if VR offers a viable solution for lifelog exploration. Previous work has shown that VR offers intuitive ways to browse images [33]. While image browsing is relatively common on VR devices, being able to handle the massive amount of data from lifelogging is a non-trivial task. Early research in this field have found that VR is a viable platform for lifelog exploration [34]. For 'conventional' lifelogging, as for VR, there are frameworks available for lifelogging system designs [35] [36].

Yang et al. [37] explored the visualization of flow maps in virtual reality. They conducted experiments involving the creation of flow maps positioned on or above 2D/3D maps or globes to determine the most effective combination. The findings indicated that the incorporation of the third dimension proved beneficial in minimizing visual clutter on the screen.

2.3 Current implementations and solutions

Currently available LID browsing systems either focus on temporal or geospatial filtering. In this section we will evaluate and digest available solutions for lifelogging visualizations. As mentioned in section 2.1, most of these implementations focus on the performance of search tasks within LIDs. However, since this research is focused on the exploration of lifelog images, the evaluation of existing systems will be conducted on systems that also focus on the exploration of LIDs instead of performances of search tasks.

2.3.1 Geospatial implementations

Within this section, we will examine current geospatial implementations designed to visualize LIDs, with the focus on exploration. Common ways to visualize geospatial data of images is by placing markers on a map that represent an image [38] [39] [40]. In this section, we will dive deeper in the implementation "Geospatial Access To Lifelogging Images in VR" by Ouwehand [16].

2.3.1.1 Geospatial Access To Lifelogging Images in VR

Geospatial Access To Lifelogging Images in VR introduced a new approach to leisure browsing of geo-tagged lifelogging images in VR. Their system, which utilizes a map-based interface, was evaluated through a pilot study involving ten general users and the two lifeloggers who contributed the dataset. The study's findings suggest that the map-based approach offers a viable method for leisure browsing of lifelogging data. Additionally, the system's ability to handle large image datasets ($n > 50,000$) in real-time highlights its practical potential. Furthermore, the high entertainment value of the VR system further reinforces its suitability for leisure purposes.

Ouwehands system employs pins to denote geographical clusters [16], as depicted in figure 1, where each cluster corresponds to data spanning one or more days. As users approach these pins, they transform into image billboards, presenting an image specific to that location. When accessing images associated with a particular location, they are arranged in a row around the user's controller. Additionally, users have the capability to filter through these images by selectively choosing or unselecting tags.

As denoted in the qualitative aspect of Ouwehands, the map view makes perfect sense for lifelogging data [16]. However, the filtering aspect could be improved, suggesting more accurate and specific filters, and better interaction with the filtering menu.

2.3.2 Temporal implementations

Within this section, we will examine current temporal implementations designed to visualize LIDs, focusing on exploration. Here, we will explore vitrivr-VR [18]

2.3.2.1 vitrivr-VR

vitrivr-VR [18] presented a virtual reality (VR)-based multimedia retrieval system called vitrivr-VR. This system combines conventional user interface elements with VR-exclusive interaction methods to facilitate efficient and engaging multimedia retrieval. vitrivr-VR has demonstrated promising results in lifelog search competitions and offers valuable insights into the application of conventional UI methods in VR and the development of VR-specific interaction techniques.

Figure 1: Geospatial representation as seen in [16]

2.3.3 Combined

Within this section, we will examine current implementations that combine temporal and geospatial data and are designed to visualize LIDs. With the chronological aspect of lifelogging data, and the metadata accompanied by it, it makes sense to combine the two in a spatiotemporal visualization. In 2D, this has been proven to be an effective way to organize and visualize "Visits" [41]. However, keeping focus on VR, in this section we will explore:

- Lifelog Access In VR Via Filtering [17]
- Joining Temporal and Geospatial Representations of Lifelogging Data [15]

2.3.3.1 Lifelog Access In VR Via Filtering

Lifelog Access In VR Via Filtering explored the potential of virtual reality (VR) for enhancing lifelog exploration. They proposed a visualization-based approach that uses maps, calendars, and tags to represent location, time, and content type, respectively. Their comparative study evaluated the effectiveness of this approach in both general exploration and targeted search tasks. While the map-based visualization encountered certain limitations, participants found the tag-based interface highly beneficial for task completion. These findings demonstrate the feasibility of a visualization-based method for lifelog access in VR and provide valuable insights for future development.

Veer employs three distinct representations for filtering and viewing life log images, namely geospatial, temporal, and tag-based filtering [17].

In the geospatial representation, images are presented as clusters on a map. Users can manipulate the map, and clusters located outside the map boundaries are excluded from view. Two geospatial representations, namely heatmaps and markers, were utilized in this research.

Temporal filtering is facilitated through a DateTime interface, allowing users to selectively choose or exclude specific time periods, down to the hour level. Interestingly, participants slightly favored the DateTime interface over the geospatial interface. The

Figure 2: VR lifelog retrieval interface as shown by Veer[17]

DateTime interface performed optimally when participants were provided with straightforward time buttons, as opposed to a more open-ended approach where participants could select times on a clock.

For tag filtering, a mechanism is implemented to filter images based on tags. Participants reported difficulties in locating specific tags due to their non-existence. Readability issues were also noted in this regard.

2.3.3.2 Joining Temporal and Geospatial Representations of Lifelogging Data

Van Geel [15] explores large image corpora, recognizing the potential of virtual reality for free exploration of expansive image datasets. The visualization leverages geographic metadata and 3D imaging techniques. Using geographically clustered data, the research expands on this approach by incorporating the temporal component of lifelogging data into filtering and clustering methods, using time information to enhance the visualization of the image set. A novel filtering technique has been deployed for filtering lifelogging data based on the accompanying tags. A user study was conducted to assess the system's usability, revealing that the visualizations contribute to the interactive exploration of the data.

The presented Figure 3 shows the spiral timeline visualization, where images are arranged in chronological order. Users have the option to directly manipulate the timeline or physically navigate within the VR environment to examine the data.

The participants expressed appreciation for the utility of geographical clustering and tag-based filtering. However, there was a divergence of opinions among participants regarding the effectiveness of the timeline visualization. While some found it user-friendly, others encountered challenges in navigation, perceiving it as disorganized. Despite this,

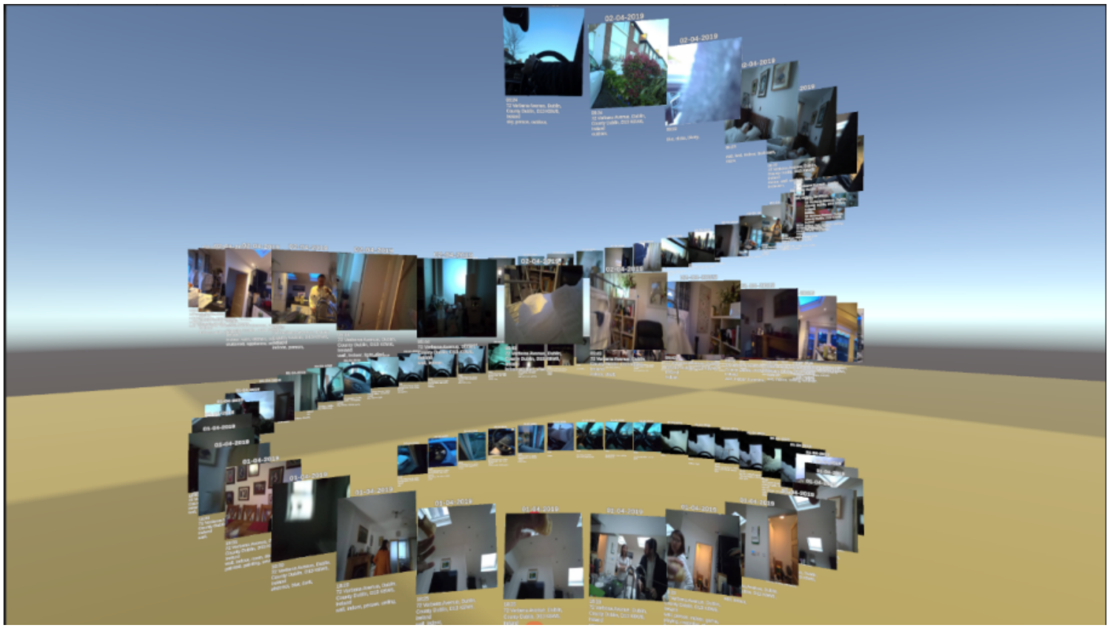

Figure 3: Timeline visualization to incorporate temporal data in the filtering

the integration of temporal visualization with geographical and conceptual data was identified as beneficial for exploring lifelogging data.

2.4 Conceptual interpratations

The emergence of large image databases (exceeding 10,000) has necessitated the development of tools capable of automatically searching and organizing images based on their content [42]. As seen in Section 3.4.1.2, the lifelog image dataset [43] used is accompanied by metadata, including 'tags'. These tags are conceptual interpretations of what is in the image. These tags are widely employed to organize lifelog images within visualization interfaces. Khanwall et al. [33] proposed two methods for linking images in a hierarchical structure based on image concepts. Their VR image exploration visualization was found intuitive and allows image navigation along multiple dimensions. Conceptual filtering or clustering has been either explicitly or implicitly deployed in many VR visualizations for LID's [16], and has been proven an effective way to navigate LID's [33] [44]

2.5 Open Research Gaps

Numerous research initiatives are currently directed towards the exploration of either temporal or geospatial filtering techniques applied to lifelog data. Frankly, almost all researches incorporate conceptual data filtering, which has been proven to be helpful for exploring lifelog data. A noteworthy and open avenue of investigation lies in the exploration of methodologies that adeptly combine both temporal and geospatial filtering mechanisms. The challenge here is not only to merge these filtering modalities but also to devise representations that facilitate the concurrent utilization of temporal and geospatial filters, ensuring a comprehensive and cohesive analytical framework for lifelog data. This research pursuits seeks to enhance our understanding of the intricate interplay between temporal dynamics and spatial contexts within lifelog datasets. Very few papers propose a system usable for undirected browsing, as most of them are about performance or accomplishing some task.

Looking at the purposes of lifelogging (such as reminiscing, memory backup, etc. [24] [25]), and its existing problems (dimensionality issue of LIDs and creating meaningful interactions with the data), it becomes evident that a lifelog browsing system should be efficient, yet immersive. A lifelog browsing system should provide the ability to efficiently browse through lifelog images. Whereas other research [26] is focusing on pure efficiency of data retrieval, lifelog image browsing could greatly benefit from unintended, yet meaningful interactions with the images. Besides being efficient, this process should also be captivating and engaging to promote explorability of unknown data. To fit the purposes of lifelogging, a VR lifelog browsing system should be highly immersive.

2.5.1 GIF's

Similar to how images can be arranged in space, they can also be arranged temporally. Take, for instance, an animated GIF image, which comprises a series of images accompanied by a specific protocol for displaying them sequentially. Helfman et al.[45] has explored several applications that use temporal organizations of images, the most promising ones combine animation with some form of spatial organization, particularly because non-overlapping images are easier to select.

Hopfgartner et al. [35] propose using a timeline as a way for a lifelogger to relive their day. According to them, each event of a day should be depicted on a timeline through a representative lifelog photo. The lifelogger can then select an event and view the photos of that event one by one. LifeSeeker 4.0 [43] deployed a similar technique which they called "Event clustering", in which they clustered consecutive images of a single invent into a sequence, in which the main or middle image is the representative image. This lead to a decrease in occurence of repeated images, and optimising the screen real estate. Considering lifelog images being a Time Cut [46], the "stack" of images can be seen as time slices of a certain event. Time slices are typically performed multiple times, and used in combination with animation or interaction [46]. Instead of keeping these static as seen in LifeSeeker 4.0 [43], we can exploit the temporal properties of these time slices, and create GIFs. Event segmentation based lifelog visualizations have been proven an effective way of visualizing lifelog images [47]

When looking at GIFs, research [48] discovered that animation, absence of sound, immediate consumption, minimal time requirements, storytelling capabilities, and the utility for expressing emotions were pivotal factors that rendered GIFs the most captivating content on Tumblr. Additionally, research indicates that compelling GIFs often feature faces and exhibit higher motion energy and uniformity. These findings are aligned with media theories and hold implications for the design of effective content dashboards, video summarization tools, and ranking algorithms aimed at enhancing engagement. These findings are promising for visualizing lifelog images.

According to Xu et al.[49], GIF thumbnails provide more vivid snapshots, and consequently will stimulate users' stronger interest in clicking and viewing the thumbnail. This can be exploited for VR lifelog image viewing.

A study conducted by Liang [50] suggests that Instagram Reels have a higher engagement score than pictures. Instagram Reels are short, scrollable videos that loop, very similar to GIF's. GIF engagement could thus also be potentially higher for VR lifelog image visualization. More so, GIFs have the capacity to augment and shape our affective performances [51]. This could boost engagement and attention retention within our VR LID visualization.

Research has also shown that humans have an inherent preference for moving objects [52]. When responding to motion pictures or GIFs, individuals undergo physiological [53] changes and are often more aroused [54]. Physiological stimulation frequently translates into emotional assessments [55]. In a parallel vein, the distinctiveness theory posits that animation has the potential to draw attention to specific screen elements in the initial phases of information processing, owing to their visual divergence from other stimuli [56]. Once the viewer's attention is captured, these distinctive objects may contribute to enhancing memory retention of the highlighted content [57]. GIFs have become the new trend in photojournalism [58], and are thus a perfect fit for displaying lifelogging images.

GIFs are promising visualization methods for displaying and exploring LIDs. Combined with geospatial, temporal, or conceptual filtering, it offers a captivating way to navigate LIDs in VR. Even when filtering, the user is left with a huge amount of pictures to explore and browse through. GIFs could provide a solution for the dimensionality issue of LIDs, while still providing a captivating experience.

2.5.2 Street View

Building upon the geospatial representation shown in previous research [16], Street View can be employed to further improve the experience of users. Besides organizing and visualizing the lifelog images, open research gaps can be found in how to experience the lifelog images. Most current implementations (as seen in previous sections) employ a way of navigating the images manually. The user can scroll through the images after a query, by using the joysticks on their controller [15] [16].

Whereas research [59] has shown that Google Streetview can help recollect memories [60], it could greatly help transmit the goal and message of lifelog images to users. Social Street View [61] also emphasizes exploiting Street View in VR to recollect and share memories. MapMash [62] employs a similar approach to model lifelog data on a street view map.

Street View could provide visual spatial context for the lifelogging images that are shown on a geospatial representation. Users could then see where each moment occurred geographically, triggering more vivid recollections of the event. The other way around, location identification and visual recognition from images have been a proven way to effectively guess where images were taken [63].

Relating back to research regarding the motivation of lifelogging [24] [25], a Street View map-based approach could potentially help in invoking a stronger emotional connection to the lifelog images, serving reminiscing. This Street View approach could also help as a storytelling tool for lifelog images, whereas previous research has proven that map-based narratives are an effective way of visual storytelling [64].

Presence refers to the perception of the virtual environment as being genuine, influenced by the vividness of the simulation and user interaction, along with individual user traits [27]. Integrating a Street View tour through lifelog images could enhance this feeling of presence when viewing lifelog images in VR.

Some research projects have shown possible feasability to employ Street View as a way to model lifelog data [65]. However, there is room for research in the way Street View can assist in improving the immersivity of lifelog images.

3 Research Goal & Approach

In this section, the research goal and approach will be discussed. The following sections will elaborate on the topics.

3.1 Research Goal

Concluding from section 2.5, there is are research gaps for browsing LIDs and creating immersive experiences out of lifelog images. VR environments can transform passive lifelogging data into engaging experiences. However, this is only possible if the data is accessible in a manageable manner, making it easy to skim through the huge amount of data. By investigating three different VR visualization methods, this research seeks to identify approaches that maximize user engagement and provide meaningful interactions with personal data. Besides the desired immersion into to personal data, this research aims to uncover ways to efficiently skim lifelogging data. Achieving ways to efficiently skim data and view them immersively, a proof-of-concept system can be designed.

3.2 Research Questions

For a proof-of-concept system for lifelog image browsing that fits the purposes of lifelogging [24] [25], both efficiently skimming lifelog data and immersive interactions with the data are very important, as mentioned in section 3.1. Therefore, this study will try to answer the following research questions:

- Which visualization is the most suitable to relive specific memories?
- Which visualization is the most suitable to browse lifelog images?

These research questions try to gain insight into the user's experience regarding browsing lifelogging images, and immersive ways to view them. With these questions answered, a proof-of-concept system can be designed.

3.3 Experiment System Design

To evaluate these questions, an experiment system has been designed to conduct the experiment described in section 4. This system consists of three main parts, presented in figure 4. This graph visualizes the dynamic between the three main parts of the system: the main experiment hub, the visualization hub, and the lifelog snippet visualization. To get an answer to the research questions, three possible visualizations are proposed. A GIF visualization, a 2D map visualization, and a 3D map visualization are employed to answer the research questions. The GIF visualization is a basic 'viewer' visualization, where the GIF plays on a big canvas in front of the user. The 2D map visualization uses geospatial information to take the user on a 'tour' above a 2D map to points where the pictures were taken. The 3D map does the same thing, but instead of a flat 2D map, it uses colored 3D models. The images in these visualizations are retrieved from a Sqlite3 database [66]. How the visualizations came to be and their requirements are discussed in the following sections.

3.4 Hardware

The hardware used to interact with the VR application is the Meta Quest 2 [67]. Implementing six degrees of freedom, the headset tracks the movement of both your head

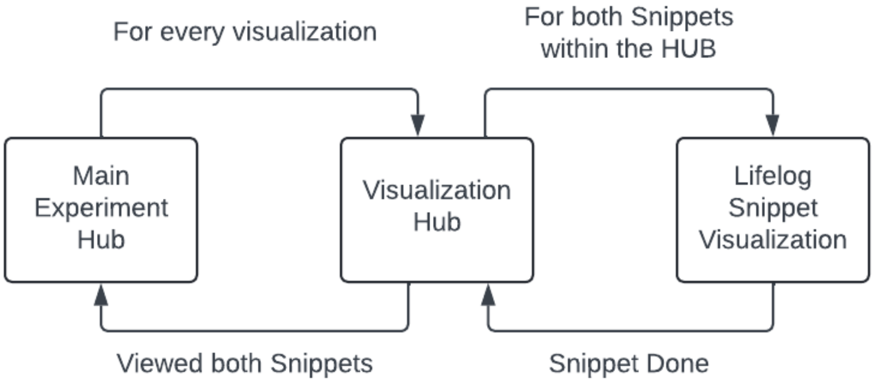

Figure 4: System Flow, containing the three main parts of the system

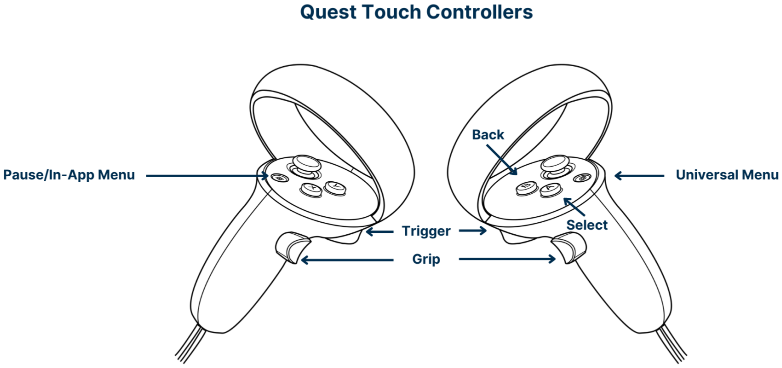

Figure 5: Button layout on the Meta Quest 2 Controller, taken from https://learn.gienc.org/course/meta-quest-2-user-guide/meta-quest-2-user-guide

and body, translating them into VR with realistic precision. The Meta Quest 2 has the advantage of being a standalone head mounted display; no external sensors are required. The screen has a resolution of 1832x1920 per eye, supporting 90Hz refresh rates to ensure smooth content display. The Meta Quest 2 has an attachment that allows users to wear glasses while using the HMD. The controllers used to control the system are shown in figure 5

The system is connected to a desktop, using Virtual Desktop [68]. This is done because the Quest 2 could not run the application standalone. To make it easier to gather more participants for the study, and make it easier to test it remotely without bringing a desktop PC, the application ran in Unity's play-mode 6. Virtual Desktop was used to run the application on a remote Quest 2. To make this possible with as little delay as possible, a 1 Gbps ethernet connection was used on the host side. The eduroam wifi from Utrecht University was sufficiently fast to run the application with a delay of 25ms. The desktop is equipped with 16GB DDR4 RAM, an i7 4790k processor, and a NVIDIA RTX 4060Ti 8GB.

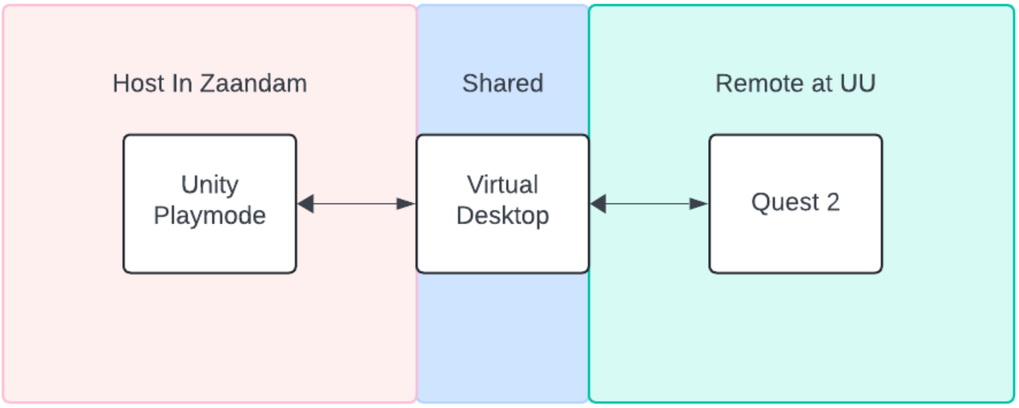

Figure 6: System interaction with the Unity's play-mode host being remotely accessed by the Quest 2

3.4.1 Dataset

In this section the properties and characteristics of the dataset will be evaluated. The LSC'23 dataset is used for this research [43]. The dataset consists of three parts; the actual images, accompanying metadata csv, and detected concepts csv. These parts will be discussed in the following sections. The last subsection of this section explains how the original dataset is converted to a SQLite3 [66] database.

3.4.1.1 Images

The dataset contains 787.681 images, from one lifelogger. The lifelogger collected these images between March 2019 and March 2020. The images have a resolution of 1024x768, and are encoded in JPEG format.

3.4.1.2 Metadata

The metadata CSV file stores more information about the images, and contains the following data:

1. Local time, denoted by a year-month-day hour:minute:second:microsecond format
2. Exact GPS coordinate, denoted by latitude and longitude.
3. Semantic name of the locations the lifelogger visited
4. Time zone the lifelogger was in when the image was taken.
5. The heart rate of the lifelogger in beats per minute
6. Calories burned by the lifelogger.
7. Distance traveled by the lifelogger between pictures.
8. Current music playing, denoted by the artist name, song name, and album name.
9. Sleep data, containing sleep levels, minutes awake, minutes to fall asleep, etc.

Of the metadata, only the Local time and the GPS coordinates will be used in the system. These are used to link images to a location and a time. Images lacking GPS coordinates are omitted from the system; it is impossible to place these on a map. 531.101 images are not accompanied by a GPS location and are therefore omitted from the dataset. The remaining database consists of 256.580 images.

Besides the metadata csv, there is a csv with visual concepts. This csv contains the following data:

1. ImageID. Used to link the tags to the corresponding image
2. Tags, visual concepts detected by computer vision programs
3. OCR (optical character recognition), characters detected by computer vision programs
4. Caption, a sentence describing what is happening in the image
5. CaptionScore, a score describing the accuracy of the caption.

However, only 725.948 images contain visual concepts. This research will not employ any data from the visual concepts csv. Therefore, it is not important whether datapoints lack tags or captions.

ImageID (text) Altitude (real)	MinuteId (text)	Latitude (real)	Longitude (real)
20190329_084752_000.jpg 2.2334551	20190329_0847	53.40310439	-6.17982238

Table 1: Database table with an example entry

3.4.1.3 Database Creation

Scripts were developed in Python3 to read and parse the csv metadata and concepts tag. These scripts were utilized to generate SQL statements, which were then saved to a temporary file to create and populate a SQLite3 database [66]. Subsequently, the SQLite3 program was employed to read these statements from the file and execute them, resulting in the creation of the SQLite3 database. It's important to note that the database includes all relevant information, except for the actual LSC images, which are accessed directly from the desktops' harddisk.

The database is rather simple and consists of a single table; the metadata table. This table and an example entry is shown in table 3.4.1.3. This table allows the system to retrieve images with a single SQL query. The metadata table includes the imageID, minuteID and the latitude, longitude and altitude. The ImageID is the primary key. The database is used as an intermediate metadata representation between the LSC dataformat and the main system in Unity. SQLite3 offers the possibility to function without a running server, unlike other database systems such as PostgreSQL or MySQL. This simplifies the system significantly.

3.4.2 Main experiment hub

The main experiment hub is the place where the participant can navigate to the three different visualizations. The proposed visualizations are a GIF representation, a 2D map, and a 3D map visualization. These visualizations will be discussed in subsections 3.4.4. The main hub is made purposely as plain as possible, to avoid affecting the user experience / the usability of the system, since the purpose of the system is not to evaluate the system, but the visualizations. The main hub is visualized in figure 9.1.5.

The three blue squares with their accompanying text are 'portals' to the visualization hubs 3.4.2. Using the joystick on the left controller, the participant can navigate to a portal. When they are within the blue area, the respective visualization hub is loaded.

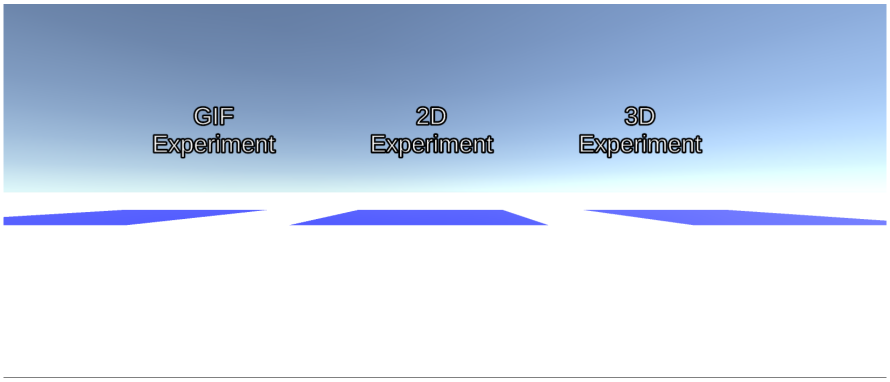

Figure 7: Main hub of the experiment

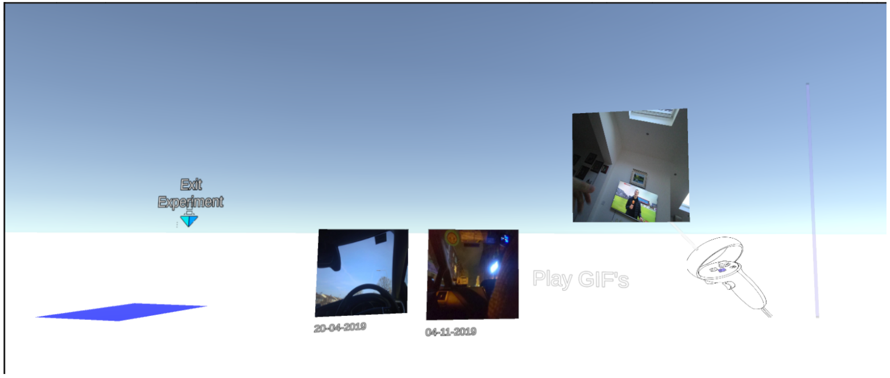

Figure 8: The generic hub of a visualization

3.4.3 Visualization hub

Each visualization hub has an identical layout, independent of visualization 3.4.2. When the participant has entered a visualization hub, two lifelog snippets will be loaded according to a Latin Square, explained in 4.2. The two snippet images have a size of 150x150 pixels, and are presented next to each other. The dates corresponding to each lifelog snippet are displayed below the snippets. The possible interactions of the participants with the snippets are described in 3.4.4.

To guide the user, the possible interaction with the snippet is displayed behind the snippets. This is done in the form of a visualization of both a dummy snippet, and highlighting the correct button the participant should use to interact with the snippet. This is done to let the user interact with the snippets without any interference of the researcher. When they have questions, the researcher can come to aid, described in section 4.

Furthermore, there is an exit portal to allow the user to return to the main experiment hub after viewing both snippets.

3.4.4 Lifelog Snippet Visualization

In the following subsection, the interactions and configuration of each snippet visualization is explained. However, some properties among the different snippet visualizations are the same for each visualization. The generic configuration for the snippets are discussed in this section.

Each snippet is a sequence between 37 and 45 images long, to ensure that snippets do not become too long, limiting the time it takes to run the experiment. Each snippet is from a different date, and each visualization contains two snippets; one fitting the category 'movement', and one fitting the category 'activity'. A 'movement' snippet has to have a minimum 10 frames of consecutive frames where the GPS location significantly (more than 2.5m apart) differs from the previous frame. An 'activity' snippet needs to have minimum of 80% of the frames something different than driving, and not constantly the same view. These categories are chosen to give the user representative snippets, in order to answer the questions the user is presented with during the experiment (Section 4. The questions presented to the participant are chosen to be able to properly answer the research questions posed in section 3.1.

As explained in 11, some tiles don't offer high quality 3D polygons. Because of the Latin Square design 4.2, each snippet could be used in the 3D map visualizations. Therefore, all chosen lifelog snippets have to start and end in a high quality 3D tile, to not detract the 3D visualization.

Whenever the participant is loaded into a visualization hub, for each date in the snippet combination (see 4.2), the following query is committed to the Sqlite3 database:

```
SELECT ImageId, MinuteId, Longitude, Latitude, Altitude FROM
metadata WHERE ImageId LIKE month LIMIT 1;
```

The *LIMIT 1* makes sure file I/O between disk and application is limited, ensuring minimal loading times. When the participant interacts with a snippet, the rest of the images of the snippet are loaded from the database with the following query:

```
SELECT ImageID, minuteId, Longitude, Latitude, Altitude FROM metadata
WHERE ImageID LIKE date;
```

A decision had to be made in terms of play speed of the snippets. The play speed of a snippet is how long images are shown before the next image in the sequence is presented. In the case of variable snippet speed, the snippet speed is determined by the distance traveled by the lifelogger between two consecutive images. The more distance traveled, the faster the snippet plays. This is could be done by using the Haversine formula [69] to calculate the distance between 2 GPS coordinates. These distances are divided by the time between the logged date-times of two consecutive images. This time between two images is not always the same, considering suspended images due to missing GPS data. In case of a constant speed between snippet frames, the snippet frames play at a rate of 1 second between frame. This ensures good visibility of each frame, yet a smooth and fast representation of the lifelog snippet. In the end, the choice falls for constant GIF speed, since this gives the most constant and predictable experience for the user. Due to the research gap for GIF playspeed as mentioned in 8.1.2, opting for a constant GIF speed is the safest option, to not have it affect the user experience too much.

3.4.4.1 GIF visualization

This visualization employs the GIF properties hypothesized to be beneficial for LID exploration as described in section 2.5. Within the GIF visualization, the user is presented with the two snippets as seen in 31. In this visualization, the user can interact with the snippet by pressing the select button on the right controller. When the user selects the snippet, the snippet plays in a big canvas 31 in front of the user, with a size

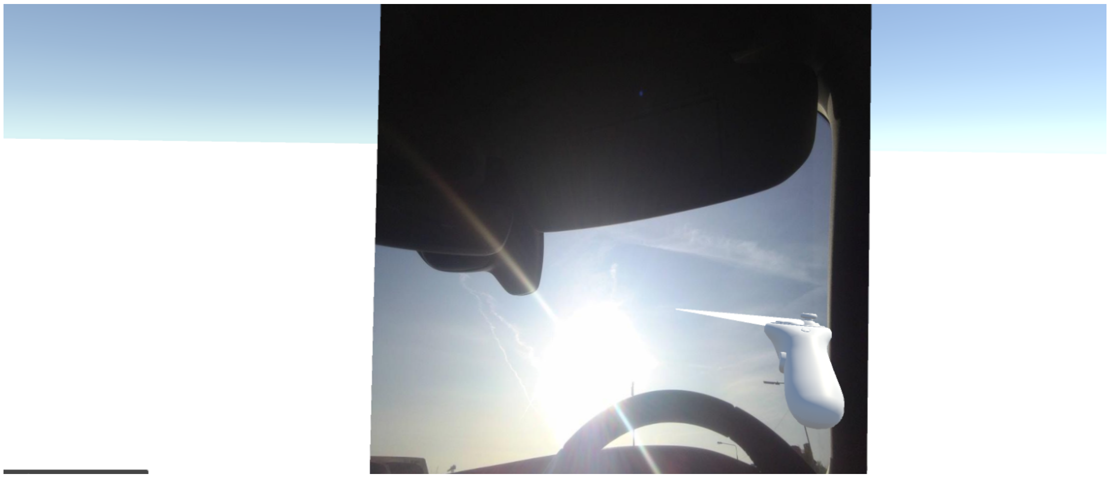

Figure 9: Participant perspective of the GIF visualization.

Figure 10: Participant perspective of the 2D map visualization.

of 500x500 pixels, taking up almost the whole viewing height of the participant. Within this environment, no other objects are displayed. When the snippet has reached its final frame, the user is returned to visualization hub.

3.4.4.2 2D map visualization

This section will describe the decisions made for the 2D map visualization, depicted in figure 10. This visualization implements a map view; Bing Maps API [70]. Bing Maps API offers a 2D map view. The goal of this visualization is to exploit the beneficial properties of a Street View visualization for LID exploration as hypothesized in section 2.5.

Bing Maps are integrated into the Unity system using the Cesium Ion Package [71]. Cesium Ion offers the capability to make direct API calls to various tile-providing services, including Bing Maps API. This allows the Unity system to seamlessly retrieve Bing Maps tiles based on GPS coordinates.

The 2D map visualization starts when the user clicks on the snippet in the visualization using the grab button on the right controller. The 2D map visualization takes place in the following steps:

1. Bing Maps tiles are retrieved by Cesium at the GPS coordinates of the first image of the snippet.
2. The user is placed in the 2D map and sees the first image in front of them.
3. The user can look around, look at the image, and when the select button on the left controller is pressed, proceeds to the next image.
4. The move vector is calculated between the current coordinate and the coordinate where the next image is taken. The user is rotated accordingly.
5. The users position linearly interpolates from the location where the shown image is taken to the next one.
6. Repeat from step 3 until there are no more images in the snippet.
7. The user is returned to the visualization view.

The following considerations have been made while implementing the 2D map view. When the user is in the starting location, the angle is calculated between the current and next GPS location. This is only calculated if the calculated distance between two images is bigger than 2.5 meters. During the development of the 2D map visualization, it was found that turning the camera for smaller movements was very confusing. For example, when the lifelogger was at home, the lifelogger would turn a lot while walking through the house. This resulted in a lot of sudden and confusing movements. The images are placed 3 unit distances from the camera, locked in place in the original direction the camera is initialized in. The image does not turn when the user looks around. This was implemented at first, but was found to be very occluding.

When the user presses the next frame button on the controller, and if the distance is large enough, the camera is rotated accordingly towards the next frame. The user starts to automatically move towards the location of the next frame, while having the next image in front of them. The time it takes for the user to reach the next location is constant, the same as the snippet speed mentioned in 3.4.4.1. This ensures that longer travels don't take up a lot of time, whereas car rides can be quite lengthy, while the pictures all look very similar; images of the steering wheel and dashboard. Next to that, the time between two images is usually roughly 30 seconds. If the lifelogger has travelled a greater distance in that same time frame, the lifelogger has had a greater speed. This translates directly to the constant travel time within the Street View Tour.

Two options were considered, implemented, and pragmatically analyzed by the researcher to proceed to the next image of the lifelogger. Firstly, the 2D map progresses to the next frame automatically without any input from the user, resulting in a fully automatic 'tour' where the user gets 'guided' through the selected snippet. Secondly, an implementation could be considered where the user signals the system the tour can proceed to the next frame. This is done by clicking on the Y-button on the left controller. This implementation enables the user to view the images at their own pace. In the end, the choice has fallen on the manual button implementation. The system can

Figure 11: Participant perspective of the 3D map visualization.

not predict which images or locations are important to the lifelogger or which images the lifelogger wants to inspect more thoroughly. Therefore, the automatic tour can feel rushed and occluded, and the choice has fallen on the 'manual' tour. However, if the button is kept pressed, the snippet plays at a constant speed until it is released.

3.4.4.3 3D map visualization

This section will describe the decisions made for the 3D map visualization, as shown in figure 11. The goal of this visualization is to exploit the beneficial properties of a Street View visualization for LID exploration as hypothesized in section 2.5. This visualization implements a realistic map view; Google Street View Photorealistic 3D Tiles API [72] (from now on, referred to as Google Street View Tiles). Google Street View Tiles offers 3D tiles for next-generation use cases, such as VR. As seen on their website, Google claims: *"These views help your users better understand geographic context, improve how they navigate, and they can showcase a place for storytelling."* [72].

The Google Street View Tiles are integrated into the Unity system using the Cesium Ion Package [71]. As mentioned in 3.4.4.2, Cesium Ion offers the capability to make direct API calls to various tile providing services, including Google Street View Tiles. This allows the Unity system to seamlessly retrieve Google Street View Tiles based on GPS coordinates.

The 3D map visualization starts when the user clicks on the snippet in the visualization hub using the grab button on the right controller. The 3D map visualization takes place in the following steps:

1. Google Street View Tiles are retrieved by Cesium at the GPS coordinates of the first image of the snippet.
2. The user is placed in the 3D map and sees the first image in front of them.
3. The user can look around, look at the image, and when pressing the button the left controller, to advance to the next image.

4. The move vector is calculated between the current coordinate and the coordinate where the next image is taken. The user is rotated accordingly.
5. The users position linearly interpolates from the location the shown image is taken, and the next one.
6. Repeat from step 3 until there are no more images in the snippet.
7. The user is returned to the visualization view.

Multiple challenges had to be overcome while implementing the 3D map visualization, resulting in the following decisions.

As mentioned in 3.4.4.2, the Cesium plugin is used to retrieve the Street View Tiles using API calls. Passing the longitude, latitude, and height, the correct tiles are retrieved. However, the height parameter does not correspond one on one to the altitude attached to the images. The height is not meters above sealevel (where altitude is), but the meters above the average ellipsoid of the earth in Cesium. This results in differences of sometimes tens of meters between the altitude and the height reference in Cesium [73]. This leads to unexpected behaviour when placing the user at the provided altitude; placing the user in the ground or tens of meters above the ground.

Using built-in raycasting in unity, the relative location to the surface can be determined. When the height is determined, the users height gets corrected accordingly. This procedure is done through the whole tour, whereas the lifelogger achieves different altitudes through the tour. A minimum of 25 meters above the surface is maintained; This results in a good overview of the location, and making low polygon areas still look detailed enough. However, if the user achieves higher altitudes (plane flights, etc), the user gets placed at this height.

Travel speed, camera rotation, and manually pressing the next frame button are the same for the 3D map as the 2D map visualization.

The Google Street View Tiles API provides varying levels of detail for its 3D Tiles, which are controlled by the MaximumScreenSpaceError (MSSE) parameter. Setting a lower MSSE yields more detailed tiles, albeit at the expense of performance, while a higher MSSE produces less detailed tiles but enhances performance. In this case, an MSSE of 16 was utilized, resulting in detailed tiles, particularly in terms of user height placement, without any noticeable decline in performance.

There is one downside of using Google Street View Tiles is the various levels of quality of tiles among the world. Metropolises and big cities have high quality 3D tiles, almost resembling the real world one on one, seen in figure 12. Rural areas, and smaller cities however, have lower quality tiles, as seen in figure 13.

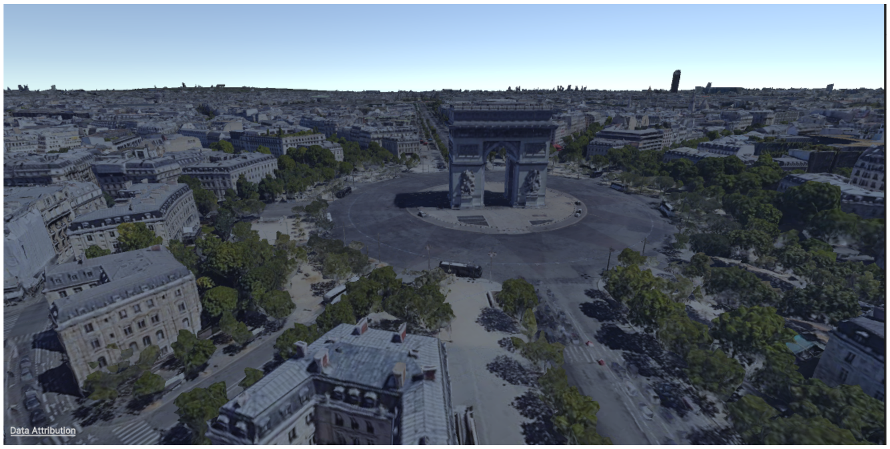

Figure 12: High quality tile for a big city like Paris

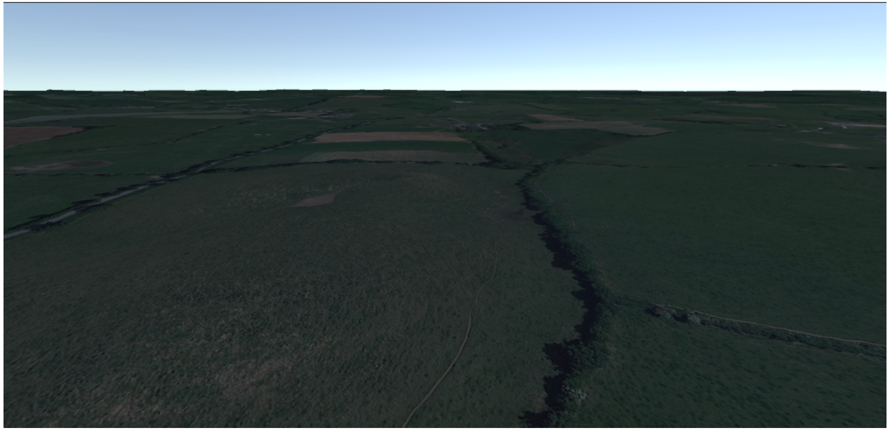

Figure 13: Low quality tile for a rural area in Ireland

4 User Study Design

To assess which visualization is best suited for each purpose (section 3.2), different aspects have to be examined. The primary emphasis is on exploratory search rather than performance-driven tasks. The experiment setup encompasses the following steps:

1. Preparatory steps
2. Qualitative testing
3. Post-experiment procedures

This research uses a qualitative test. The qualitative test is used because the focus of this research is on experience rather than performance-driven tasks. The following sections will provide a thorough explanation of each component of the experiment.

4.1 Preporatory steps

Prior to conducting the experiment, the system and its objectives were explained to the test subject. Moreover, the purpose of the experiment was emphasized, with a focus on testing the system rather than the test subject. It was clarified that taking breaks or discontinuing participation was permissible at any point. Due to the potential for (VR) motion sickness, test subjects were required to sign a consent form (see Appendix A.1). Besides this, it was made clear that the experiment was going to be conducted completely anonymously. Subsequently, the following questions were posed:

1. Age, in years
    (a) Fill in
    (b) Prefer not to say
2. Preferred gender
    (a) Male
    (b) Female
    (c) Non-binary / other
    (d) Prefer not to say
3. How much experience do you have with VR?
    (a) I have never used VR
    (b) I have little experience with VR (a few times)
    (c) I occasionally use VR (e.g. a few hours per month)
    (d) I use VR often (e.g. more then 10 hours per month)
4. How much experience do you have with map-based systems, such as Google Maps?
    (a) I have never used them.
    (b) I have little experience with them (e.g. used them a few times, but do not use it normally)
    (c) I occasionally use them (e.g. a couple times a month)

    (d) I often use them (e.g. more than 10 times a month)

5. How familiar are you with timelapses, reels, or animated GIFs?
    (a) I don't know what they are
    (b) I know what they are, but I rarely see them (e.g. seen them before, but not normally)
    (c) I see them sometimes (e.g. a couple times a month)
    (d) I see them a lot (e.g. more than 10 times a month)

This data of the participants is necessary to draw conclusions about the visualizations. Certain groups might find the system more (e.g. already have experience with lifelogging) or less (e.g. completely new to VR) enjoyable, based on their experience. As stated previously, the experiment aims to evaluate the visualizations rather than the test subjects. Consequently, participants are free to take breaks or discontinue their involvement at any point if they wish. If participants do not complete the experiment in its entirety, their test results will be omitted from the analysis.

After answering the questions, the participant will be shown a piece of text explaining the research, with some accompanying screenshots of the application to make them familiar with the application. This text can be found in the appendix A.2

4.2 Evaluation during experiment

This section will explain the qualitative evaluation during experiment.

When the participant puts on the headset, the participant is presented with the main hub of the experiment, as seen in figure 7. The three blue squares are portals toward each visualization, and the participant is asked to go to one of them. The order in which the participant views the visualizations is determined by a Latin square. A Latin square was used for this experiment since it is possible to measure each subject under every visualization and, in addition, it is necessary to control for changing conditions over the course of the experiment. If every participant views the same visualizations in the same order, and together with that, the same snippets for each visualization, it could possibly introduce carry-over effects. For example, one lifelog snippet may, for whatever reason, suit a certain visualization more than another snippet. This would introduce a bias towards certain visualizations. A Latin square design also allows us to efficiently use participants, whereas if we had a fully randomized design, we would need way more participants to test every configuration. The Latin square used in this research is shown in table 4.2

Participant 1	A1	B2	C3
Participant 2	B1	C2	A3
Participant 3	C1	A2	B3

A −>GIF visualization	1 −>Lifelog snippet combination 1
B −>2D visualization	2 −>Lifelog snippet combination 2
C −>3D visualization	3 −>Lifelog snippet combination 3

Table 2: Latin square used for the experiment

The qualitative evaluation during the experiment testing is done in three parts. The steps 1 and 2 are done for each visualization, so three times in total, making the participant answer each question for each visualization.

1. View two lifelog snippets in visualization X
2. Answer the following questions a Likert scale:
   (a) Please rate the following statements about the GIF visualization on a scale from 1 (strongly disagree) to 5 (strongly agree)
      i. I feel confident that I have a good understanding of the activities done by the person wearing the lifelog camera.
      ii. I feel confident that I have a good understanding of how the person wearing the lifelog camera was moving around.

Considering the nature of the chosen lifelog snippets, these questions directly evaluate if the visualization conveys the 'message' of the lifelog snippet. The answers filled in are on a Likert scale [74]. As mentioned in section 3.4.4, each visualization contains two snippets; one fitting the category 'movement', and one fitting the category 'activity'. However, the participant is not aware of this.

The participant is then asked to navigate to the visualization as determined by the Latin square. The visualization hub is shown to them, as seen in figure 3.4.2, and the snippets are also chosen according to the Latin square. The participant is asked to first view the left snippet, by using the A-button the on the right controller. When viewing a snippet, the user is loaded into the actual visualization environment. In figure 31 the GIF visualization is shown. In figure 10 the 2D map visualization is shown. In figure 11 the 3D visualization is shown. When in the environment, the user can click the left Y button to advance to the next image, or keep it pressed to automatically show the images one after the other. After the snippet reaches its final image, the user is returned to the visualization hub. Then, the participant is asked to view the snippet on the right side. After viewing both snippets in the visualization, the participant is asked to answer the questions mentioned in step 2 for the corresponding visualization. After answering the questions, the participant is asked to leave the visualization hub and go to the main hub by using the blue portal on the left, as seen in figure 3.4.2. This process repeats until all three visualizations have been shown.

4.3 Post-experiment procedures

After viewing the three visualizations, the participant is asked to fill in some last questions that complement the questions asked during the experiment. The following questions were asked:

1. Please pick one of the visualizations for the following statements
   (a) Which of the three interfaces would you consider the most immersive?
   (b) Which of the three interfaces would you consider the most efficient to quickly glance at the lifelog images and get an idea of what the person was doing?
2. For the GIF visualization, mention up to three advantages or positive aspects that you liked about the visualization

3. For the GIF visualization, mention up to three disadvantages or negative aspects that you did not like about the visualization
4. For the 2D map visualization, mention up to three advantages or positive aspects that you liked about the visualization
5. For the 2D map visualization, mention up to three disadvantages or negative aspects that you did not like about the visualization
6. For the 3D map visualization, mention up to three advantages or positive aspects that you liked about the visualization
7. For the 3D map visualization, mention up to three disadvantages or negative aspects that you did not like about the visualization
8. Any general feedback or comments?

Question 1 complements the questions from the during-experiment evaluation. Questions 2 to 7 force the participant to think about each visualization more in-depth, and hopefully get advantages and disadvantages for each visualization. When answering these post-experiment questions, the participant is assured that they can ask questions about the visualization, or view a still image of the visualization to refresh their memory what it looked like. This is done to eliminate any confusion between the names of the visualization and the corresponding visualization.

5 Results

The experiment was conducted with 19 test subjects. In this result section the following data will be presented:

1. Background information on participants
2. Confidence results
3. Post experiment questionnaire results
4. Advantages and disadvantages of different visualizations
5. General comments

5.1 Background information on participants

The results of the pre-experiment questions are shown below. All participants were between the ages of 19 and 28, with an average age of 23.56 ± 2.38. Figure 14 shows the distribution of age of all participants.

12 out of the 19 participants selected male as the preferred gender. 7 participants selected female as the preferred gender. No participants selected 'non-binary / other', and also no participants selected 'prefer not to say'. The preferred gender distribution is shown in figure 15

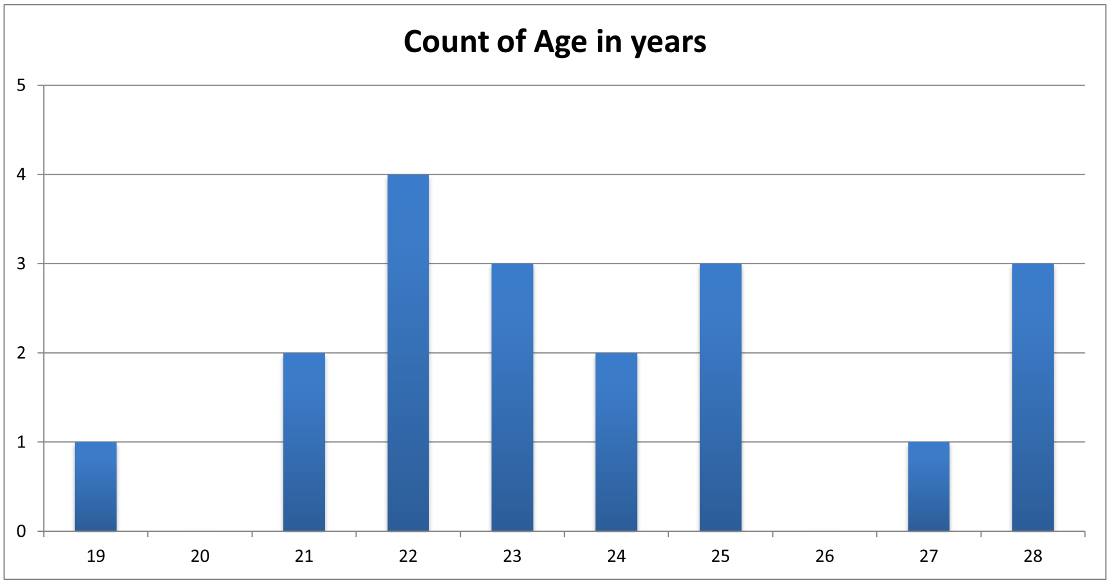

Figure 14: Age distribution of the participants

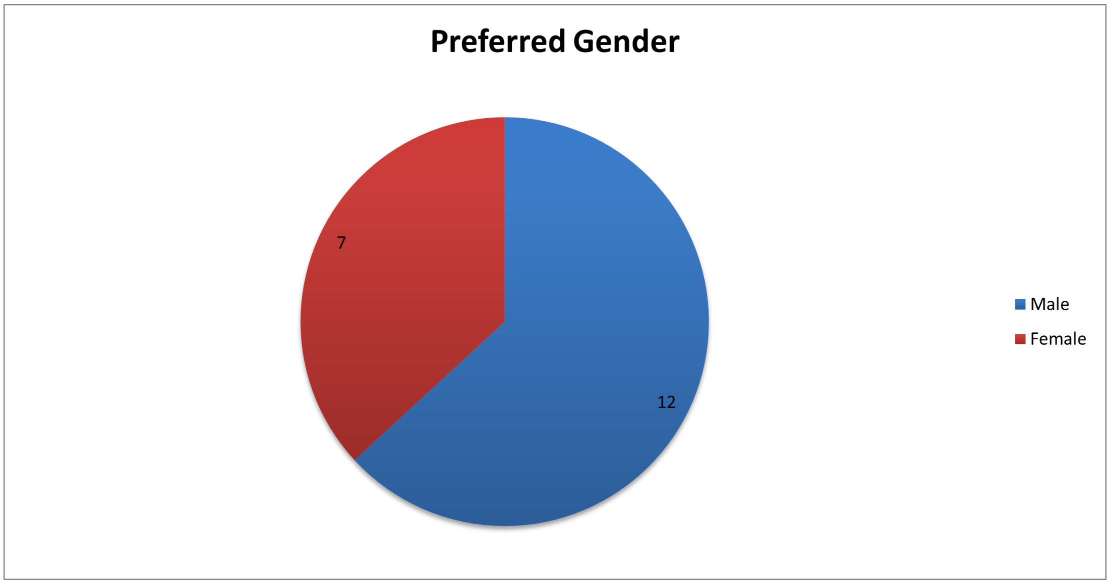

Figure 15: Preferred gender of the participants

When prompted with the question about VR experience, roughly two-thirds of the participants reported that they had used it before (see figure 16), but had little experience with it (e.g. one to five times). One participant reported to be using VR often (e.g. more than 10 hours per month). Two participants had no prior experience with VR. Reported experience with map-based systems is shown in figure 17. Roughly two-thirds of the participants reported using map-based systems more than 10 times a month. The rest of the participants use map-based systems occasionally (e.g. a couple times a month). No participants responded on never using map-based systems, even more so, no participants responded with that they have little experience with them. As for the reported familiarity with timelapses, reels or animated GIFs (see figure 18, almost all (17 out of 19) participants reported that they see them a lot (e.g. more than 10 times a month). The other two participants see them sometimes (a couple times a month). As with the map-based systems, no participant answered not knowing what they are or rarely seeing them.

5.1.1 Confidence results

This section presents the results of the confidence level questions asked during the experiment. As mentioned in section 4.2, the questions were answered on a Likert scale. Because of the low frequency of participants reporting using VR often, or never using VR, the participants can be grouped into two groups; Inexperienced (Participants reporting that they have never used VR, or have little experience with VR) and Experienced (Participants reporting that they occasionally use VR or use VR a lot). Considering all participants responded that they are familiar with GIFs, timelapses or reels, it does not make sense to group participants based on their familiarity with them. Participant distribution for experience with map-based systems is similar to the experience with GIFs, timelapses, or reels. Therefore the participants are not grouped for experience with map-based systems either. Because of this, only the effect of VR experience on the confidence results will be presented.

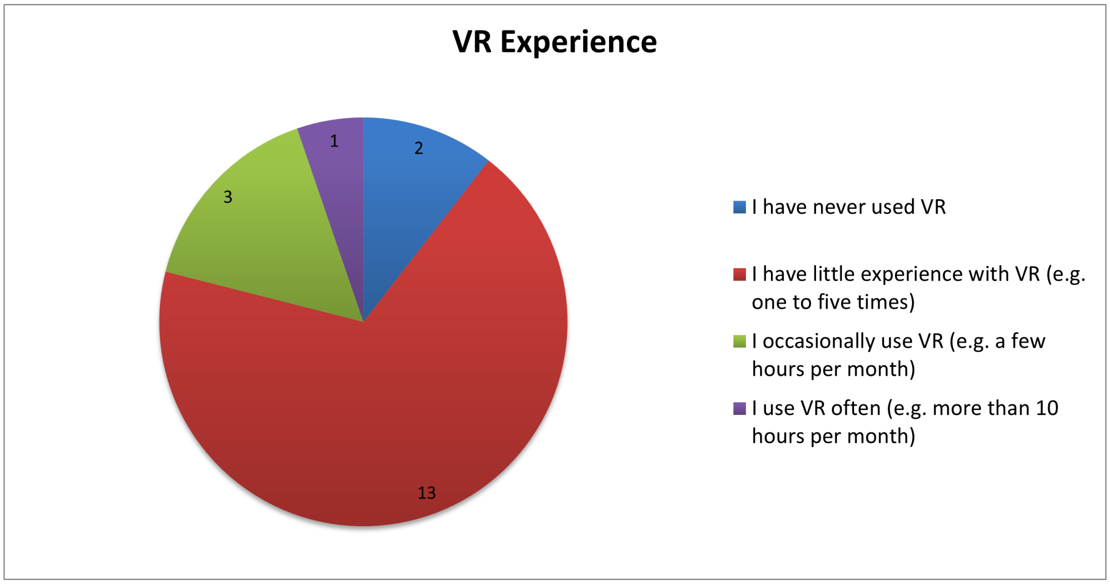

Figure 16: Reported VR experience of the participants

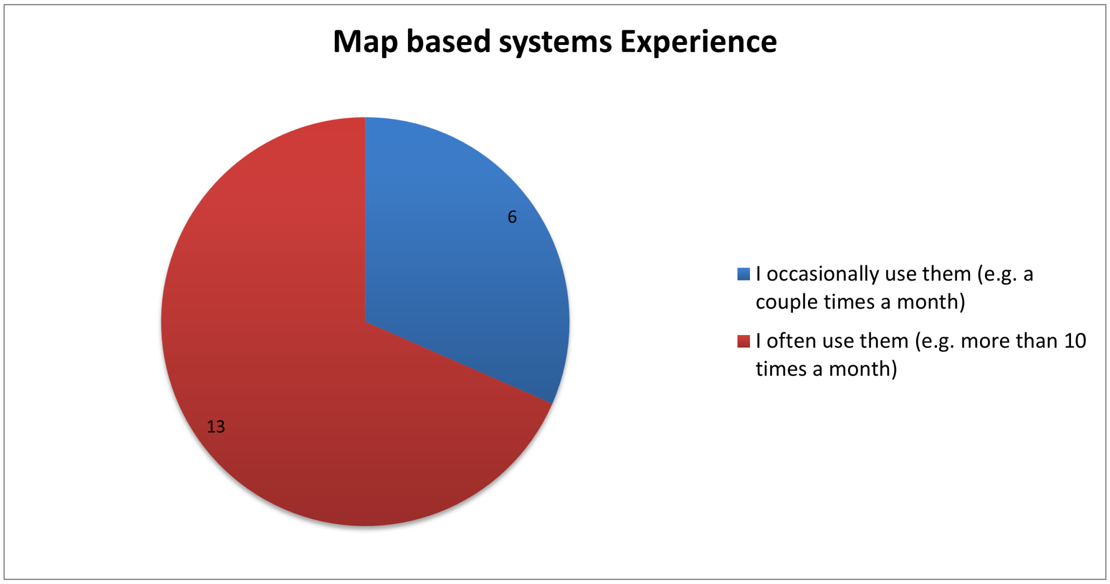

Figure 17: Reported map-based experience of the participants

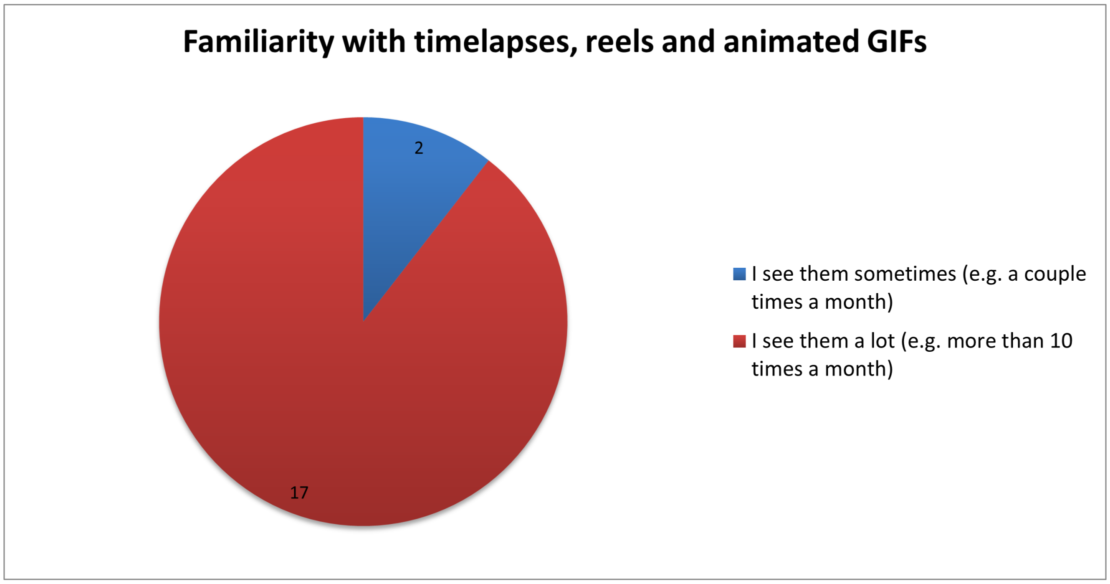

Figure 18: Reported familiarity with timelapses, reels, or animated GIFs of the participants

For the confidence levels about the GIF visualization, the results are presented in figure 25. This graph shows that 9 out of the 19 participants agree with the statement ”I feel confident that I have a good understanding of the activities done by the person wearing the lifelog camera” For the GIF visualization. Out of these 9 participants, 8 strongly agree, and one slightly agrees. Of the other 10 participants, four participants disagree with the statement. One strongly disagreed, while three of them slightly disagreed. The other six participants neither agreed or disagreed. The average response scores a 3.57 ± 1.30 on the Likert scale for this statement.

When looking at the effect of VR experience on the activity confidence for the GIF visualization 19, the confidence level of inexperienced participants was on average a 3.66 ± 1.34. The average activity confidence for experienced users was a 3.50 ± 1.25. Calculating the *p-value* results in a *p-value* of 0.83. This is not a statistically significant result, and thus the VR experience does not have a influence on the activity confidence level for the GIF visualization.

For the statement ”I feel confident that I have a good understanding of how the person wearing the lifelog camera was moving around.”, only 4 out of the 19 participants responded that they agreed with the statement. Just one of them strongly agreed with the statement, and three out of those slightly agreed with the statement. Six of the participants were neutral about the statement, answering ”neither agree or disagree”. The rest of the participants (9) disagreed with the statement. Three participants strongly disagreed with the statement, and six slightly disagreed. The average response scores a 2.63 ± 0.98 on the Likert scale for this statement, slightly below neutral score.

When looking at the effect of VR experience on the movement confidence for the GIF visualization 19, the confidence level of inexperienced participants was on average a 2.60 ± 1.18. The average activity confidence for experienced users was a 2.75 ± 0.96. Calculating the *p-value* results in a *p-value* of 0.82. This is not a statistically significant result, and thus the VR experience does not have a influence on the movement confidence level for the GIF visualization.

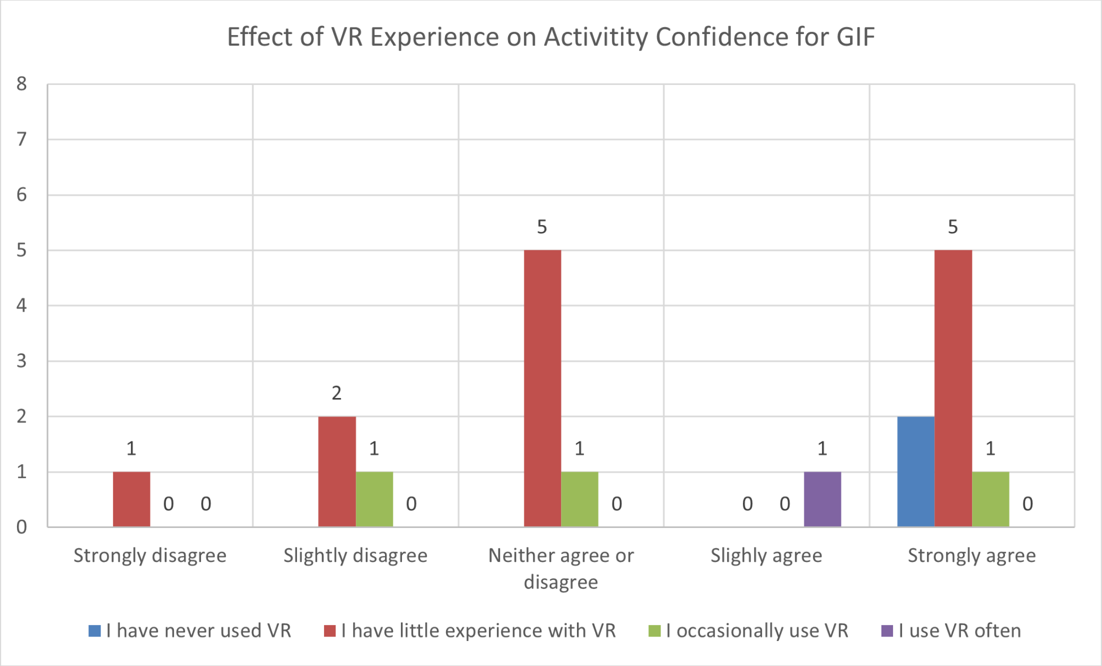

Figure 19: Effect of VR Experience on Activity Confidence for GIFs

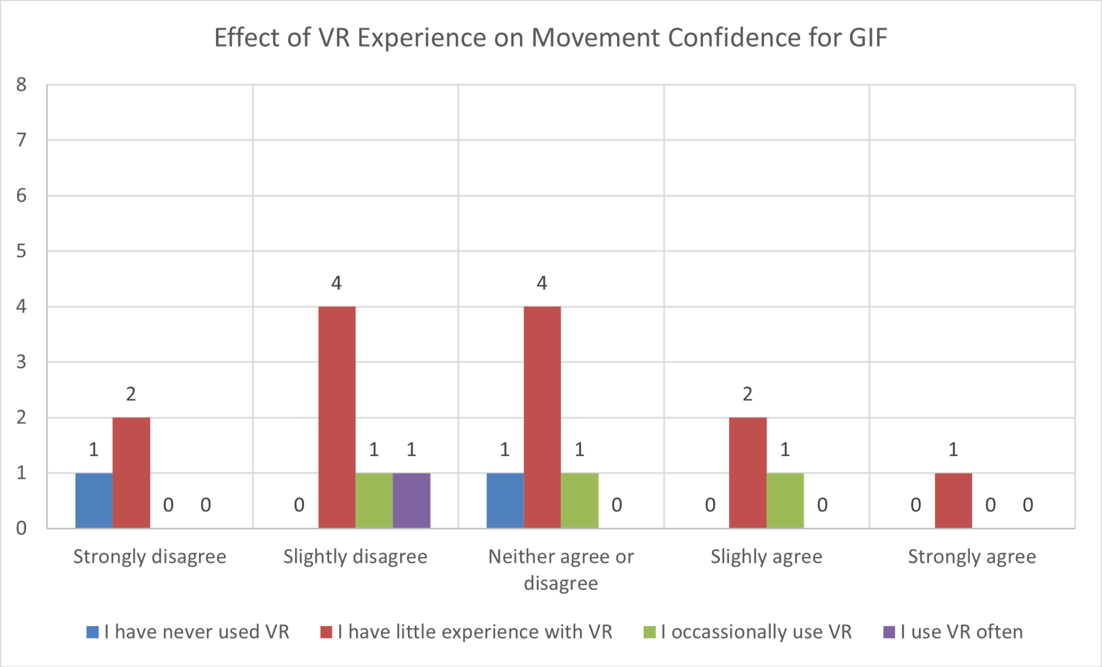

Figure 20: Effect of VR Experience on Movement Confidence for GIFs

For the 2D visualization, the results are presented in figure 26. No participants disagreed with the statement ”I feel confident that I have a good understanding of how the person wearing the lifelog camera was moving around.” for this visualization. Three of the participants neither agreed nor disagreed with the statement. Most of the participants (9) slightly agreed with the statement. Seven of the participants strongly agreed with the statement. This statement scored an average of 4.21 ± 0.69, surpassing

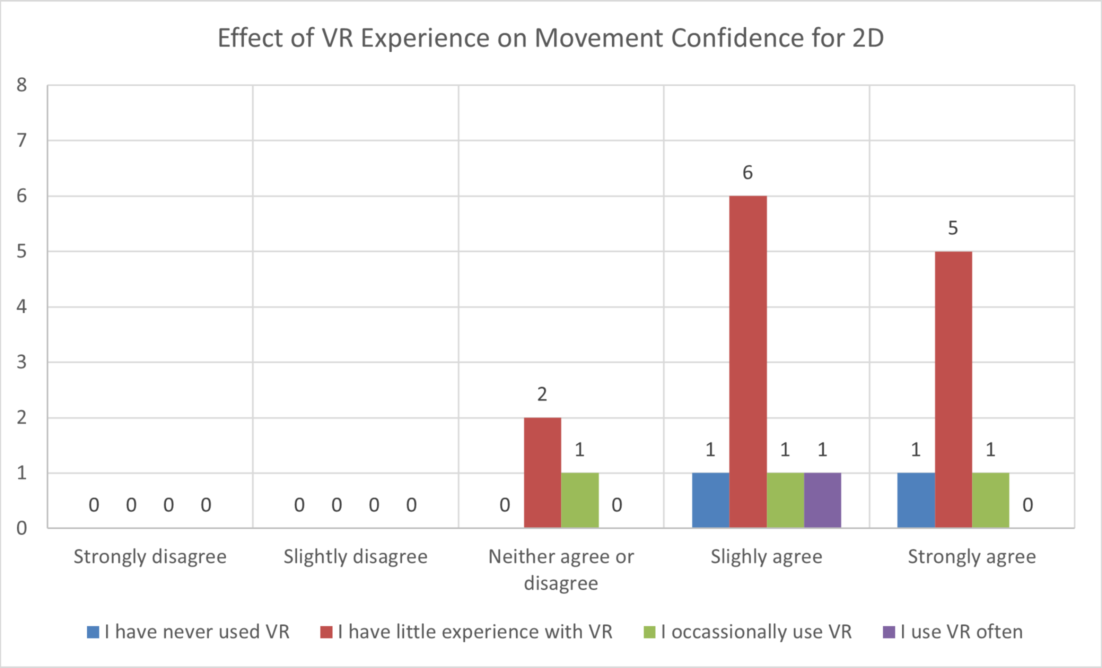

Figure 21: Effect of VR Experience on Movement Confidence for 2D

the ”slightly agrees” response.

When looking at the effect of VR experience on the movement confidence for the 2D visualization 21, the confidence level of inexperienced participants was on average a 4.27 ± 0.70. The average activity confidence for experienced users was a 3.75 ± 1.26. Calculating the *p-value* results in a *p-value* of 0.28. This is not a statistically significant result, and thus the VR experience does not have a influence on the movement confidence level for the 2D visualization.

For the statement ”I feel confident that I have a good understanding of the activities done by the person wearing the lifelog camera”, no participants strongly disagreed. Three participants slightly disagreed with the statement. In total, eleven participants agreed with the statement. Out of these eleven participants, seven of them strongly agreed with the statement, and four of them slightly agreed with it. The average confidence level on the Likert scale of this statement for this visualization is a 3.78 ± 1.10, which nudges towards an average of ”slightly agrees”.

When looking at the effect of VR experience on the activity confidence for the 2D visualization 22, the confidence level of inexperienced participants was on average a 3.80 ± 1.21. The average activity confidence for experienced users was a 3.50 ± 1.00. Calculating the *p-value* results in a *p-value* of 0.66. This is not a statistically significant result, and thus the VR experience does not have a influence on the activity confidence level for the 2D visualization.

For the 3D visualization, the results are presented in figure 27. Interestingly enough, all participants agreed to some extent with the statement ”I feel confident that I have a good understanding of how the person wearing the lifelog camera was moving around.” Twelve of the participants strongly agreed with the statement, while the rest of them slightly agreed with the statement. This statement for this visualization averages a 4.67 ± 0.47, which leans towards a ”strongly agrees” score on the Likert scale.

When looking at the effect of VR experience on the movement confidence for the

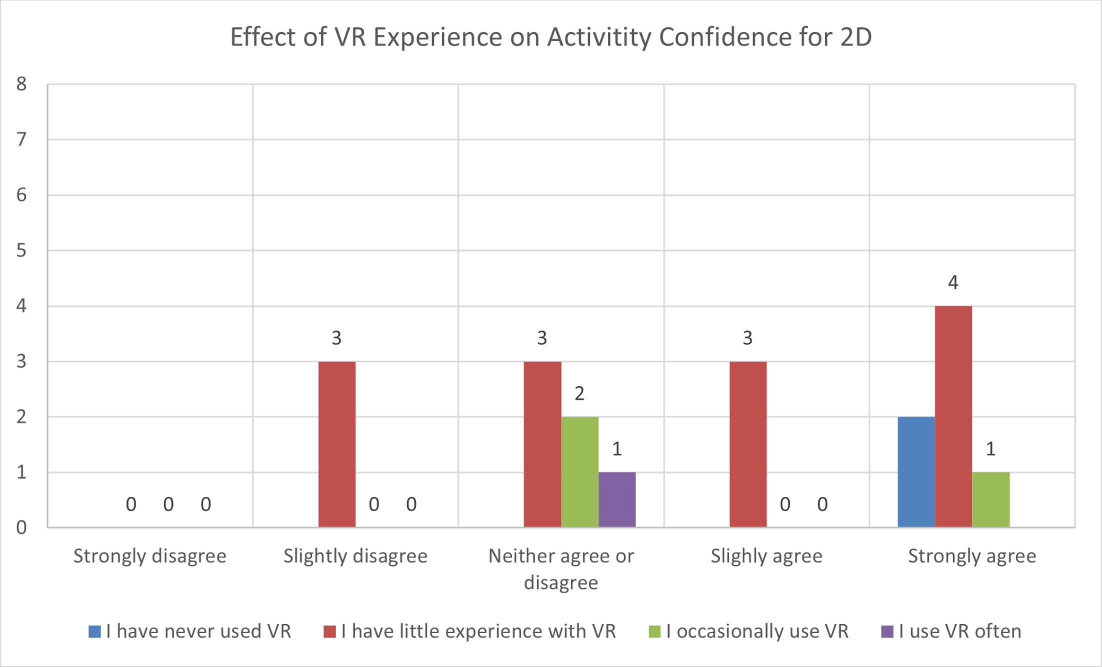

Figure 22: Effect of VR Experience on Activity Confidence for 2D

	Activity Confidence	Movement Confidence
GIF	3.57 ± 1.30	2.63 ± 0.98
2D	3.78±1.10	4.21 ± 0.69
3D	4.12 ± 0.97	4.67 ± 0.47

Table 3: Likert scale confidence levels for the different visualizations.

3D visualization 23, the confidence level of inexperienced participants was on average a 4.67 ± 0.49. The average activity confidence for experienced users was a 4.75 ± 0.50. Calculating the *p-value* results in a *p-value* of 0.76. This is not a statistically significant result, and thus the VR experience does not have a influence on the movement confidence level for the 3D visualization.

The statement ”I feel confident that I have a good understanding of the activities done by the person wearing the lifelog camera.” scores an average of 4.12 ± 0.97 on the Likert scale. Out of all the participants, 9 of them strongly agreed with the statement. Four of the participants slightly agreed with the statement, and five of them neither agreed or disagreed. Only one participant slightly disagreed with the statement.

When looking at the effect of VR experience on the activity confidence for the 3D visualization 24, the confidence level of inexperienced participants was on average a 4.20 ± 1.01. The average activity confidence for experienced users was a 3.50 ± 1.00. Calculating the *p-value* results in a *p-value* of 0.24. This is not a statistically significant result, and thus the VR experience does not have a influence on the activity confidence level for the 3D visualization

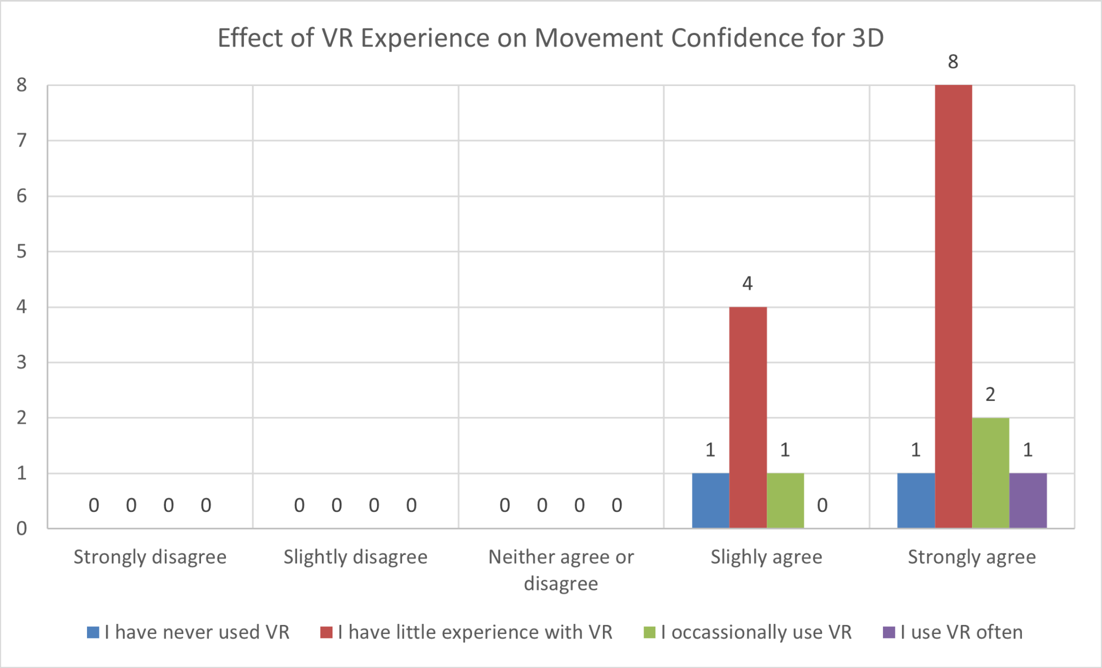

Figure 23: Effect of VR Experience on Movement Confidence for 3D

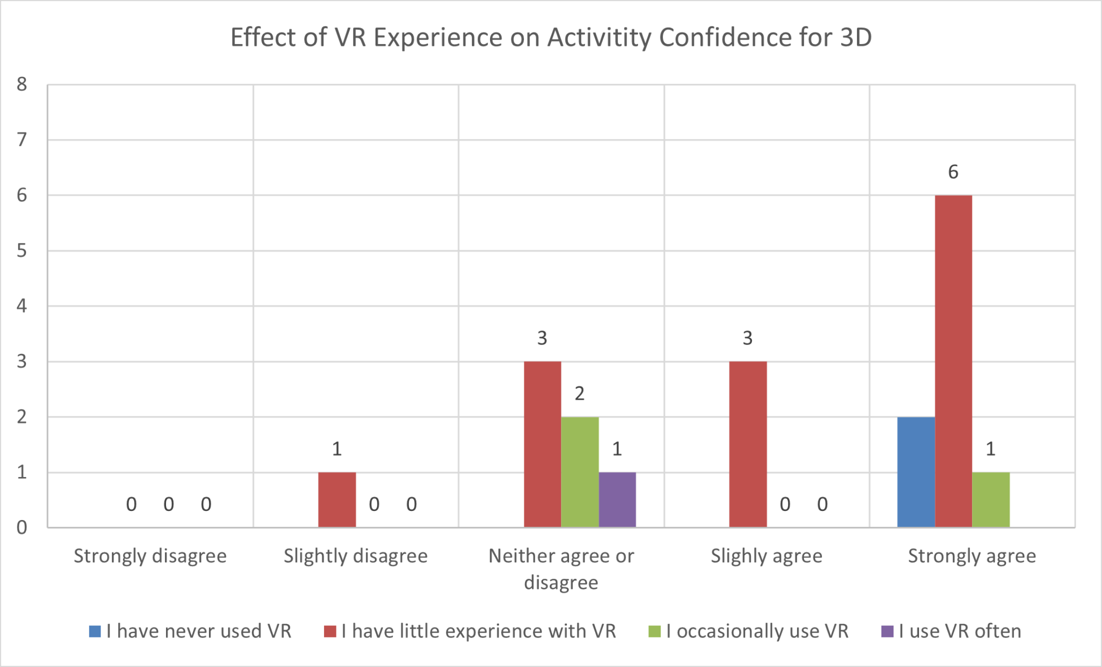

Figure 24: Effect of VR Experience on Activity Confidence for 3D

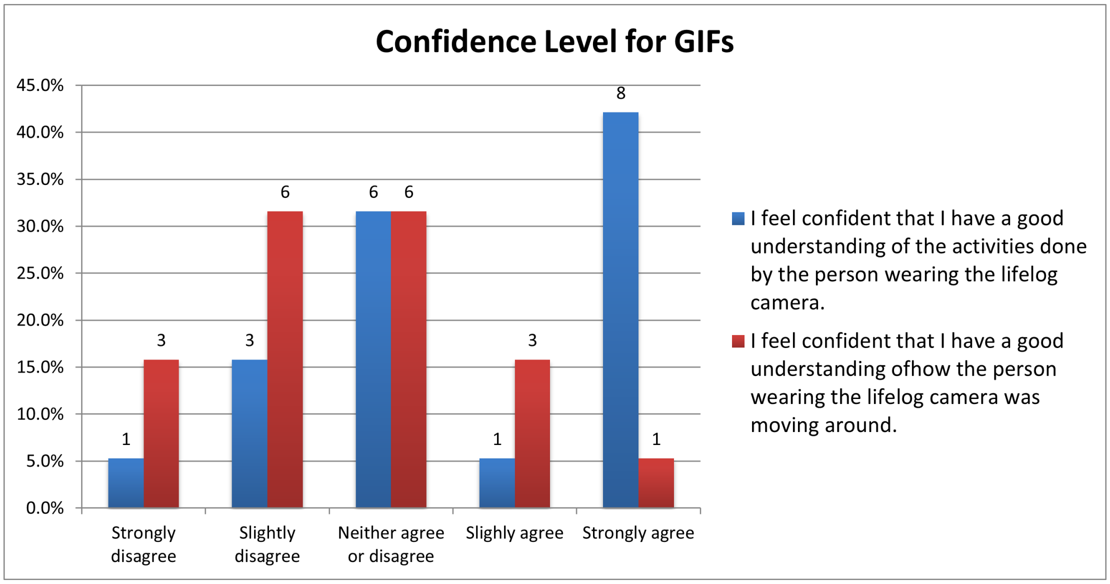

Figure 25: Agreement of perceived confidence level about the statements of the GIF visualization.

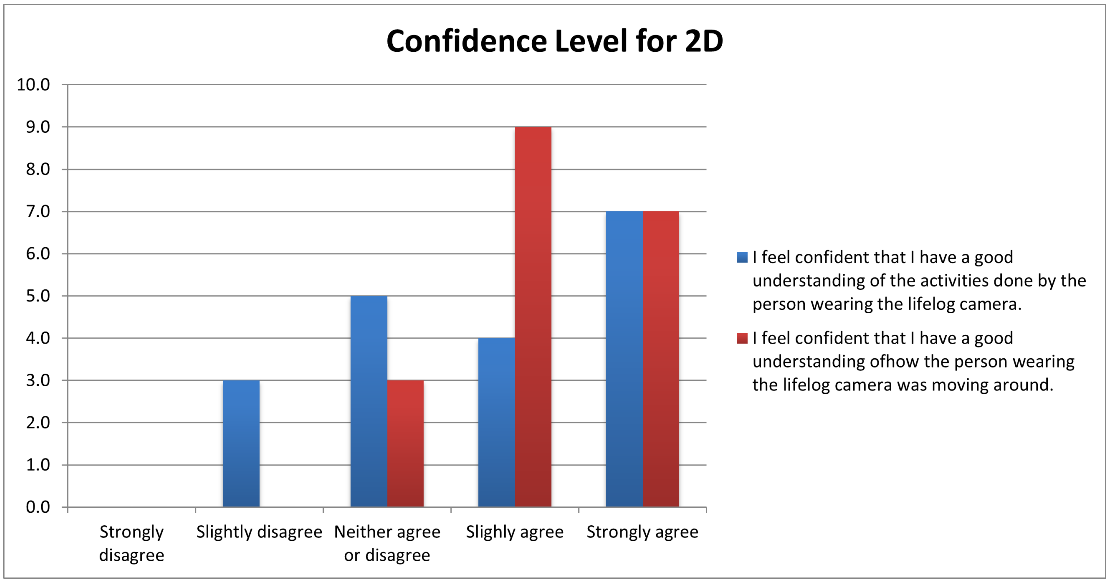

Figure 26: Agreement of perceived confidence level about the statements of the 2D visualization.

		Inexperienced	Experienced	*p-value*	Statistically significant?
GIF	Activity Confidence	3.66 ± 1.34	3.50 ± 1.25	0.83	No
	Movement Confidence	2.60 ± 1.18	2.75 ± 0.96	0.82	No
2D	Activity Confidence	3.80 ± 1.21	3.50 ± 1.00	0.66	No
	Movement Confidence	4.27 ± 0.70	3.75 ± 1.26	0.28	No
3D	Activity Confidence	4.20 ± 1.01	3.50 ± 1.00	0.2	No
	Movement Confidence	4.67 ± 0.49	4.75 ± 0.50	0.76	No

Table 4: Effect of VR Experience on confidence levels for the different visualizations

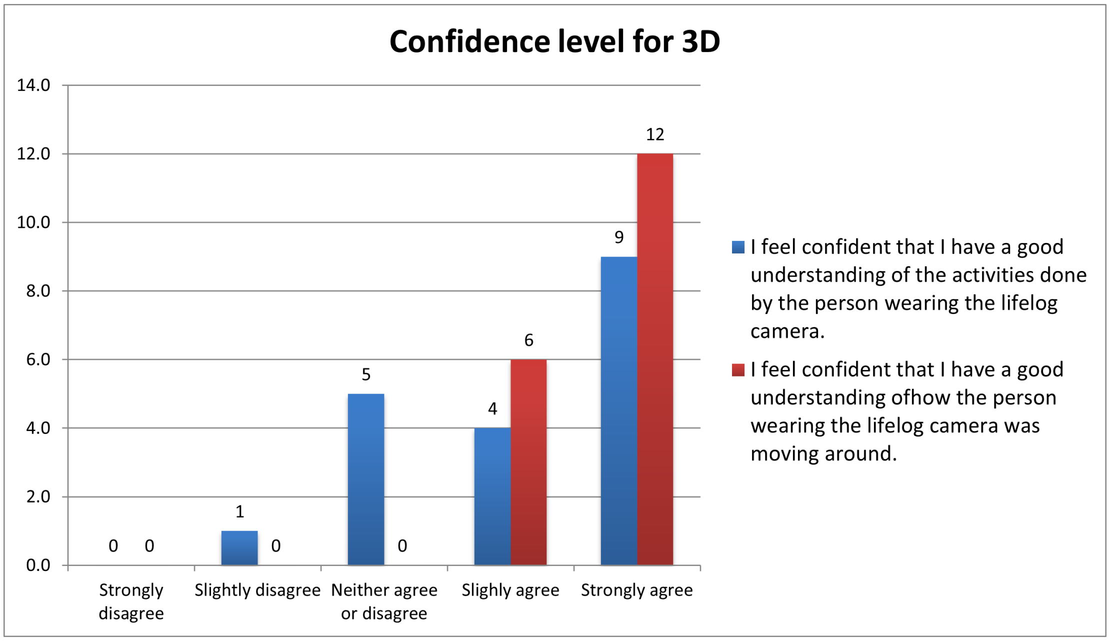

Figure 27: Agreement of perceived confidence level about the statements of the 3D visualization.

5.2 Post experiment questionnaire results

As mentioned in 4, after viewing all visualizations and answering the Confidence Questions, the participants were asked to answer some final questions. This subsection will take a look at the results of the two post experiment questions;

1. Which of the three interfaces would you consider the most immersive?
2. Which of the three interfaces would you consider the most efficient to quickly glance at the lifelog images and get an idea of what the person was doing?

Interestingly, all participants responded with the 3D visualization when asked which of the three interfaces they considered the most immersive. This is an interesting result, considering the distribution of VR experience of the participants. Achieving complete consensus in human responses is rare unless the question has a factual answer (e.g. ”What is 2+2?”). Therefore, this result should be treated with care.

The responses for ”Which of the three interfaces would you consider the most efficient to quickly glance at the lifelog images and get an idea of what the person was doing?”, the responses were more divided. Roughly three quarter (14 participants) of the participants reported that the GIF visualization was the most efficient to quickly glance at the lifelog images. The other quarter of the participants (5) reported that the 3D visualization was the most efficient to glance at the pictures.

5.3 Advantages and disadvantages of different visualizations

This section will report the typed responses for the advantages and disadvantages of each visualization. The results listed in this section will be summarized and categorized. For all the answers, please refer to the Appendix A.2.1.

5.3.1 GIF Visualization

When looking at the typed responses for the advantages of the GIF visualization, we can come to the following categories:

- Lack of Distractions
- Efficiency
- Simplicity and Ease of use:

As for Lack of Distractions, many statements highlight the absence of distractions, which allows for a subjective better focus. Examples include: "There were no distractions. This helped looking at the pictures better", and, "It was a clear interface with no distractions, which is nice if you actually want to focus on the images". As for efficiency, several statements emphasize the efficiency of the experience, noting that it was quick and easy to view the content without waiting. One participant mentioned: "It was nice that there was no 'loading' of the environment in this view which made it quick and smooth". Simplicity and Ease of use were mentioned in various statements, ranging from "This view felt basic, yet it felt trusted" to "Minimalistic design made it easy to focus on the GIF". These statements go hand in hand with the aforementioned categories. Most of the given comments directly relate to the question "Which of the three interfaces would you consider the most efficient to quickly glance at the lifelog images and get an idea of what the person was doing?".

The typed responses for disadvantages of the GIF visualization can be categorized in a similar manner:

- Lack of engagement
- Lack of immersion
- Insufficient information

Within the lack of engagement category, comments can be found like "A bit boring" and "Obviously, it was less 'cool' than the others". This seemed to be a common theme among the responses. Besides the lack of engagement, users also reported a lack of immersion, with comments such as "This one didn't really benefit from the VR, as it was quite static". Participants also noted difficulties in understanding the content due to the lack of context. An example of such a comment is: "You don't get a feeling of time and you don't know where it was.". This is directly related to the results presented in 5.2, where no participant reported the GIF to be the most immersive.

5.3.2 2D Visualization

When looking at the typed responses for the advantages of the 2D visualization, we can come to the following categories:

- Enhanced Engagement and Immersion
- Ease of following movement
- Contextual awareness

The category "Enhanced engagement and immersion" summarizes statements like "More interesting than the GIF one", and, "The map itself was fun". One participant reported "It was fun being taken along the images and seeing where they were taken, like Snapchat maps". Snapchat maps [75] is a function within the Snapchat mobile application, that saves images on a 2D map based on the location where they were taken. This particular comment from this participants indicates that the visualization was familiar for them. The "ease of following movement" category encapsulates comments such as "Clear map with visible street names", and, "For movement, it was more clear where the person went". Quotes such as "It was nice to see where the pictures where taken and that you were taken along the roads" and, "An advantage was seeing where the person was and the exact street names" are categorized within the category "Contextual awareness".

As for the disadvantages of the 2D view, the following categories can be constructed:

- Distraction and Focus Issues
- Lack of Aesthetic and Visual Appeal
- Functional and Usability Concerns

There were quite a few participants reporting distraction and focus issues with the 2D visualization. A few comments fitting this category are: "It was slightly distracting", "I was looking a the names of the roads more than actually paying attention to the images" and "Hovering above a 2D map breaks the immersion a bit". Multiple participants reported that the map was not that aesthetically pleasing, with comments ranging from "The map was not the most pretty" and "The map itself was quite plain" to "The map itself felt a bit abstract, which I don't know if it fits the purpose. The last identified category is the "Functional and usability Concerns". Disadvantages fitting this category include "It felt a bit strange to "fly" above a map", and "Because the map had a similar color all around, it was hard to distinguish the locations from each other".

5.3.3 3D Visualization

Similarly to the 2D and GIF visualizations, we can identify categories for the typed advantages and disadvantages of the 3D visualization.

- Immersive and Realistic Experience
- Visual Appeal and Enjoyment
- Clarity and Sense of Space

The number of positive responses for the 3D visualizations was by far the most. A few of the comments categorized as "Immersive and Realistic Experience" include "Very nice to look around, and would love to visit, for example, my old holiday destinations." and "This one was really cool, and the places I was taken to looked like I was there in a drone". Another interesting comment fitting this category is "Very interesting to see the 3D buildings etc. I have seen Google Earth before, but not in VR, which in combination with the images was very fun to see.". For the category "Visual Appeal and Enjoyment", the following comments are within this category: "This one was even more fun than the 2D map. It looked really good.", "The visuals were impressive." and "Obviously the most pretty.". These comments indicate a solution to a disadvantage of the 2D map. The last category is "Clarity and Sense of Space". Comments that fit this category include: "It was less distracting than the maps.", "Gave a good sense of environment and movement." and "Gave a good sense of environment and movement.".

As for the comments about disadvantages of the 3D visualization, the following categories can be distinguished:

- Loading and Visual Quality Issues
- Navigation and Orientation Challenges
- Distraction and Overwhelming

Some participants reported that the environment sometimes had to be loaded. These comments are categorized in the Loading and Visual Quality Issues. One participant mentioned, "A bit nitpicky: sometimes the road was not followed perfectly.". Another participant reported, "When it turned the view, it made me slightly disoriented.". These comments can be categorized as Navigation and Orientation Challenges. One participant reported, "The looking around has the same drawback as the 2D map; it was a little more distracting.". Together with comments such as "The looking around has the same drawback as the 2D map; it was a little more distracting.", the category "Distraction and Overwhelming" was found suitable.

5.4 General Comments

The general comments made by the participants can be found in Appendix A.2.1. When filtering irrelevant comments, such as "goodluck with your thesis" etc, most of the general remarks mention that the experiment was interesting, and the different views impressive to see in VR. A few participants made some suggestions how the outcome of this research could be used in future applications. One comment theorized that this could be used for promotional works for holidays or sightseeing. A handful of participants mentioned that they would really like to see their own images in such visualizations. Most of these comments were made about the 3D visualization.

6 Discussion

In this section, the results will be evaluated and discussed to answer the research questions proposed in 3.2. With the research questions answered, we propose a proof-of-concept prototype system incorporating the findings of this research, to come to an immersive and accessible VR lifelog experience.

To answer the research question "Which visualization is the most suitable to browse lifelog images?", the following data needs to be evaluated:

- Confidence of Activities done
- Efficiency

In order to answer the research question "Which visualization is the most suitable to relive specific memories?", the following data needs to be evaluated:

- Confidence of Activities done
- Confidence of Movement Behaviour
- Immersiveness

Besides this data, typed advantages, disadvantages, and general comments should be taken into account per research question. Prior experience with VR, map-based systems, and timelapses / reels will not be considered because of their lack of statistical significance.

6.1 Confidence of Activities done

As mentioned in the Result section 5, but repeated here for good measure to put the activity confidence into perspective, the average confidence levels for the visualizations are as follows: the GIF visualization scored the worst with an average of 3.57 ± 1.30 on the Likert scale. The 3D visualization scored the best with an averagge of 4.12 ± 0.97 on the Likert scale. The 2D scored a 3.78 ± 1.10.

A reason why the GIF visualization scored the worst could be the lack of contextual information, as indicated by the written disadvantages. Whereas the 2D and 3D visualization provides spatial information about where the picture is taken, it implicitly also provides information about what the camera wearer was doing at that point. Consider a series of images taken on a walk along the street; viewing the images in the GIF visualization might leave participants guessing based on limited visual information, however, images placed on a map could indicate the lifelogger walking from class to class, or from work to a coffeeplace etc. Some images might also provide ambiguous information when viewed in the GIF visualization. Seeing them on a 2D or 3D map can reduce ambiguity and could clarify what activities might be taking place. The extra contextual information, spatial relationships and environmental cues provided by the 2D and 3D maps could result in the confidence of the participants in these visualizations.

The 3D visualization scored better than the 2D visualization. This could be due to similar reasons as to why the GIF visualization scored worse. The beforementioned spatial understanding, realistic visualization, and even better contextual clues could lead to a higher confidence level of the activities done by the lifelogger.

6.2 Confidence of Movement Behaviour

As mentioned in the Result section 5, but repeated here for good measure to put the movement confidence into perspective, the average confidence levels for the visualizations are as follows: the GIF visualization scored the worst with an average of 2.63 ± 0.98 on the Likert scale. The 3D visualization scored the best with an average of 4.67 ± 0.47 on the Likert scale. The 2D scored 4.21 ± 0.69 on the Likert scale.

A reason why the GIF visualization scored the worst, could be that, same as with the activity confidence, the lack of contextual information. A map provides critical contextual, spatial and temporal information that could enhance the participants' ability to accurately and confidently estimate movement behaviour in a series of images. For example, in the map visualizations, the participant can estimate the distance traveled between two frames, which makes it easier to estimate movement behaviour. The additional layers of information that a map provides help to create a coherent narrative of movement, which is much harder to achieve by viewing images in isolation.

Similar to the evaluation of the confidence for the activity, the difference in results between the 2D and 3D map can be explained by the fact that the 3D map inhibits the same properties as the 2D map, but to a greater extent.

6.3 Efficiency

In order to answer the research questions, it is important to evaluate the post-experiment question "Which of the three interfaces would you consider the most efficient to quickly glance at the lifelog images and get an idea of what the person was doing?". As stated in the result section 5.2, roughly three quarters reported the GIF visualization to be the most efficient for getting an idea of what the person was doing. This is a surprising result, considering the GIF visualization scored the worst for the confidence level of the activities done by the lifelogger. This could possibly be explained by either: The users reporting a low confidence in activities done did not pick the GIF visualization for this question, or, the GIF visualization scored the worst in the confidence level question about activity, yet it was still enough to get an impression of what was happening in the GIF.

6.4 Immersiveness

The post-experiment question "Which of the three interfaces would you consider the most immersive?" is of utmost importance to answer the research questions. As stated in the result section 5, all participants found the 3D map to be the most immersive. This is an unusual statistic, whereas consensus in human responses is rather rare. Therefore, the responses from this answer should be treated with care and be further evaluated. The reason for a consensus within this experiment setup (e.g. excluding potential bias or influence, since no participant could be influenced by other participants) could be that the question is unambiguous, leaving no room for interpretation.

Another reason could be a sampling issue. Participants from a homogeneous group (e.g. same background, education, etc.) might give more similar responses than those from a diverse group. However, achieving identical responses still remains uncommon unless the group is extraordinarily uniform. In this research, the backgrounds of participants had some overlap, but was not identical from subject to subject.

Outside coercion or pressure is not relevant in this experiment, whereas the researcher only gave instructions, or additional information when asked.

Summing all this up, the most likely is that it is probably an unambiguous question, combined with the fact that it is most likely the most immersive visualization.

7 Conclusion

To conclude this research, the proposed research question will be answered. With the answers to these research questions, a proof-of-concept system can be designed and implemented.

7.1 Which visualization is the most suitable to browse lifelog images?

To be able to answer this research question, the results and analysis of the "Confidence of Activity" and "Efficiency" should be taken into account, as mentioned in section 6.

Looking at "Confidence of Activity", the 3D visualization scored the best. However, theorized in 6.3, when this category is combined with the "Efficiency" questions and the typed responses of the participants, the GIF implementation still gave a sufficient idea of what was happening in the GIF, yet proving to be more user friendly and efficient to quickly navigate and glance at snippets. Therefore, the GIF visualization was found to be the most suitable to browse lifelong images.

7.2 Which visualization is the most suitable to relive specific memories?

To be able to answer this research question, the results have to be interpreted in a rather abstract manner. The participants have not seen the lifelog images before, or even more so, not experienced those moments captured by the lifelog wearer. However, the data gathered in the experiment should suffice in giving an answer to which visualization would be the most suitable *if* the observer had experienced those moments.

The most important data to consider when answering this research question, is "Confidence of Movement Behaviour", "Confidence for Activitity" and the "Immersiveness", as mentioned in section 6. Looking at "Confidence of Movement behaviour", the 3D visualization scored by far the best. In the category "Confidence for Activity", the 3D visualization scores the best as well. The 3D visualization was also answered to be the most immersive by the participants. Combining this with the typed out advantages categorized in 5.3.2, the 3D visualization was found to be a realistic experience. It presented clarity and sense of space, with appealing visuals, leading to enjoyment and immersion. Therefore, the 3D map visualization was found to be the most suitable to relive specific memories.

7.3 Final Conclusion

This paper presented three different visualizations to gain insight in ways to make lifelog data accessible and immersive. A GIF visualization, 2D map visualization and 3D map visualization were implemented. A user study among 19 participants was conducted to see which of these visualization was the best for quickly glancing at lifelog snippets, and which visualization was the most immersive. The results and feedback obtained in this study show that the GIF visualization is the best for quickly skimming through and glancing over lifelog snippets, and the 3D visualization was found the most immersive. A few participants mentioned wanting to use the 3D map visualization to be able to view their own images.

As mentioned in the section 3.1, the findings of the study will be used to design and implement a proof-of-concept system. The findings about the three different visualizations will be utilized in the next sections to implement a scientifically proven efficient and immersive way to navigate and re-experience lifelog events.

8 VR-LIVE

Clearly, even when filtering, and looking at previous implementations [16] [17], the LID's are too big to easily navigate, and no novel approach to quickly skim through this huge database has been presented. Besides this, an immersive way to view lifelog images in VR that aligns with the motivations to lifelog [24] [25] is of utmost importance, as mentioned in 2.1. Therefore, to tackle both research gaps (section 2.5), based on the research conducted into the different visualizations and the conclusions drawn from them, we propose the system VR-LIVE: Virtual Reality Lifelog Image Visualization and Exploration. VR-LIVE will provide an efficient way to skim through LIDs, and give the user the possibility to relive experiences through the exploration of lifelog images. VR-LIVE employs a novel approach to efficiently navigate LIDs using GIF walls. These walls consist of looping GIFs, visualizing certain temporal ranges of a users' query. Besides tackling the navigation issue of LIDs, VR-LIVE aims to improve the experience of relieving memories using lifelog images played on a Street View representation, as a small 'movie' of these GIFs in the environment they were captured in.

We propose the creation of a dynamic wall composed of GIFs. These GIFs are organized per temporal range, chosen by the user. This concept of a wall is inspired and proven effective by Duane's memory wall [34]. The GIF visualizations has been proven an effective way to skim through lifelog snippets in the comparative study 7. Combining the wall-concept and the GIF visualizations, a GIF wall should provide a pleasant experience for users to skim through lifelog snippets. Users can interact with the wall by hovering over a GIF, enabling them to play the animation. These GIFs are constructed based on temporal filters; the user can select a time range, and the wall will be constructed based on those constraints. This is explained in more detail in section 8.1.2.

Clicking on a GIF will start a Street View Tour. This tour consists of a small timelapse movie consisting of the images the GIF is composed of. The user will be placed on the Street View map, where the tour will start. This experience moves the user along the geospatial placed lifelog images that are part of the GIFs. As shown in the conclusion of the comparative study 7, the 3D map is proven efficient in providing an immersive experience for the user. The user will relive the experience by viewing the images, while at the same time traveling along the actual path that the lifelogger took when the images where captured in the 3D map visualization. This would be a similar approach to that used in the research by Min Lu et al. [76].

In tandem with the proposal for a dynamic wall constructed from GIFs, the exploration of LID's visualizations delves into the psychological aspects driving the desire to log personal life data. The proposed wall concept aligns with the fundamental advantage of lifelogging systems, particularly in enhancing episodic memory [77] [78]. The ability to recall and relive specific life experiences becomes a central theme, as individuals engage in the intricate process of reflecting on past events using the Street View representation of the GIFs. VR-LIVE aids in mentally retracing steps through the visualization of autobiographical memories. In combination with the fact that lifelogging has been shown to help trace memory, or even in more severe cases, help with recovering from memory loss or amnesia [79], the combination of a lifelog wall made up of GIFs and the 3D Street View representation of these GIFs could potentially greatly increase interactivity and exploration of lifelogging images. Whereas the GIF wall will focus mostly on the exploration aspect of the images, the Street View Tour will focus on re-experiencing the captured moments.

8.1 VR-LIVE Design

As mentioned in previous sections, VR-LIVE introduces an experience-based GIF wall. This wall is constructed based on temporal filtering. The main hub, the construction of the GIFs, the construction of the wall and the construction of the 3D map visualization will be evaluated in the following subsections.

8.1.1 Main Hub

The main hub facilitates the place to construct walls. These walls are based on temporal filters, as explained in 8.1.2. The main hub houses the space to interact with the GIF wall. To make this more convenient for the user, a small, visual tutorial will be available. These images help the user understand the interactions with VR-LIVE. Besides the tutorial, a basic button layout will be presented to the user.

8.1.2 Construction of GIFs

For the wall to be constructed, the GIFs need to be created. GIF construction can be divided in two tasks; splitting the chosen temporal range into smaller subranges to make GIFs from, and defining GIF properties. In this section we will illustrate how both are achieved.

As described, the user will be able to filter the original dataset based on time; the user can select a month through the wrist menu (Figure 32), or multiple months. Such a wrist menu is a proven method for efficiently filtering lifelog images based on temporal properties, as proven by Van Geel [15]. Images falling within the chosen range will be returned by VR-LIVE. These are the images that will be used to construct the GIFs. The GIFs will be separated per day.

It is crucial to take into account the duration of GIFs. Additionally, when separating time between frames from the duration of a GIF, one should consider the number of frames in a GIF. In practice, there is no constraint to the amount of images or duration of a GIF [80]. However, similar to looping GIFs, research and data has shown that the optimal time for a social media reel [81] is between seven to fifteen seconds [82].

Time between frames in a GIF does not have an established FPS (frames per second [83]), but generally ranges between 12-30FPS [84]. Lower FPS can still lead to smooth running GIFs [84]. Considering the time delay between two consecutive images within GIFs (on average taken 30 second apart), the GIFs probably will not result in a smooth running sequence of frames, but rather more of a slideshow images. This will ease the constraint on the high FPS requirement of smooth running GIFs.

Due to the emphasis on exploration in this research, it is advisable to avoid creating GIFs with an excessive number of images or prolonged intervals between frames. This could potentially result in distractions or boredom. The number of image in one GIF is equal to the number of images returned per day for the chosen time range. On the otherhand, GIFs that play too fast or consist of very little frames, it will be difficult to distinguish individual frames. While the number of frames directly corresponds to the filtering process, the preferred time between frames is a characteristic that can vary among users.

The last property of the GIFs that should be considered is instant replay. GIFs are sequences of images that usually loop [85]. Many GIFs are played endlessly whether or not the user is paying attention to them [85]. However, one could argue, with the size of lifelogging image sets, and the immersiveness of VR, this could lead to overstimulation for the user. Therefore, GIFs will only play when the user interacts with them.

8.1.3 Construction of the Wall

The last step to visualizing the GIF wall, is constructing the walls out of the composed GIFs. Similar to [34], we employ a wall visualization. When looking at walls, there are 3 properties to be defined:

- Dimensions
- Interaction with GIFs
- Interaction with walls

Keeping exploration in mind, the dimensions of the display grid is an important property to keep in mind. Given the infinite space of VR, the trade-off between exploration and usability is not a trivial one. Whereas huge dimensions can lead to great explorability, the usability may suffer. In turn, small walls will lead to low explorability. Keeping in mind the scope of this research, being the explorability of the lifelogging images in VR, and the way GIFs are constructed based on previous section, there will be no constraints on the dimensions of the wall. This allows the user to have full control about how they want to explore the dataset.

As mentioned in the section 8, the user will have the ability to interact with the GIFs in order to explore the images within them. To view images within a GIF, the user can hover and select GIFs. When a GIF has been selected, the images that compose the GIFs will be shown one by one in sequence, with the interacted GIF shown in a big canvas, as proven efficient by the comparative study 7.

The interaction with the GIFs needs to be combined with the interaction with the wall. The user will have the possibility to walk around in real life in order to navigate the environment, but, taking into account that sometimes movement space is limited when using a VR, the joysticks of the controllers will also be employed for movement.

8.1.4 3D map Tour

Lastly, the 3D map representation needs to be designed to make the VR-LIVE whole. The main properties that need to be considered for the 3D map representation is the playspeed of the experience, the representation of the map itself, and interaction with individual frames.

The goal of the 3D map is to pull the user into an emerging experience when a GIF is selected. The 3D map visualization has been proven immersive 7, and therefore will be used in the tour. Whenever a GIF is selected, the user will be placed on the map at the GPS location of the first image of the GIF. The user will be placed in a POV manner, where the surroundings of the user will be a 3D map of the location the user is in. In front of the user, the first image of the GIF is displayed. From this moment on, the GIF will start playing, and the surroundings will move according to the geotags accompanying the lifelog images. When the sequence is played, the user will be returned to the GIF wall.

9 Implementation

This section will describe the way the GIFs wall and the Street View are implemented. This will be done by reviewing how the GIFs wall is implemented and how the Street View was integrated.

9.1 VR-LIVE System Overview

In this section, a high-level overview of the designed and implemented system will be given. This consist of a high level overview of VR-LIVEs hierarchy, and used technologies. The hardware used for VR-LIVE is the same as used in the visualization study 3.4. The interaction with unity's play mode with VR-LIVE is done in a similar manner as described in 3.4. However, since VR-LIVE is not going to be verified by participants, for testing and developing, a USB-C to USB-C cable is used to connect the Quest 2 to the desktop.

9.1.1 High Level Technical Overview

As seen in figure 28, VR-LIVE consists of multiple parts. VR-LIVE is composed of the following parts:

- LSC Dataset Metadata
- SQLite Database
- LSC Dataset Images
- Unity System
- Cesium Ion
- Bing Maps API
- Google Street View Tiles API

The LSC Dataset Metadata is discussed in section 3.4.1.2. This green block consists of the CSV files containing the metadata of the images. As described in section 3.4.1.3, the CSV files are converted to a SQLite3 database, in order to simplify and increase the efficiency of image query. The LSC Dataset Images will be read directly from the disk, and is not part of the Sqlite3 database. The database consists of primary keys corresponding to images. This reduces the size of the database significantly.

9.1.2 Unity

The platform was created in Unity (personal edition, version 2022.3.6f1 [86]), accompanied by C# scripts. The platform uses a few plugins to ease development:

- SQLite plugin for interacting with the SQLite3 database [66]
- ListView plugin for creating and managing list views, used for filtering menu [87]
- TaskParallel plugin for managing C# threads, used for loading images in background threads [88]
- CesiumIon plugin [71], for making API calls to both Bing Maps API and Google Street View Tiles API

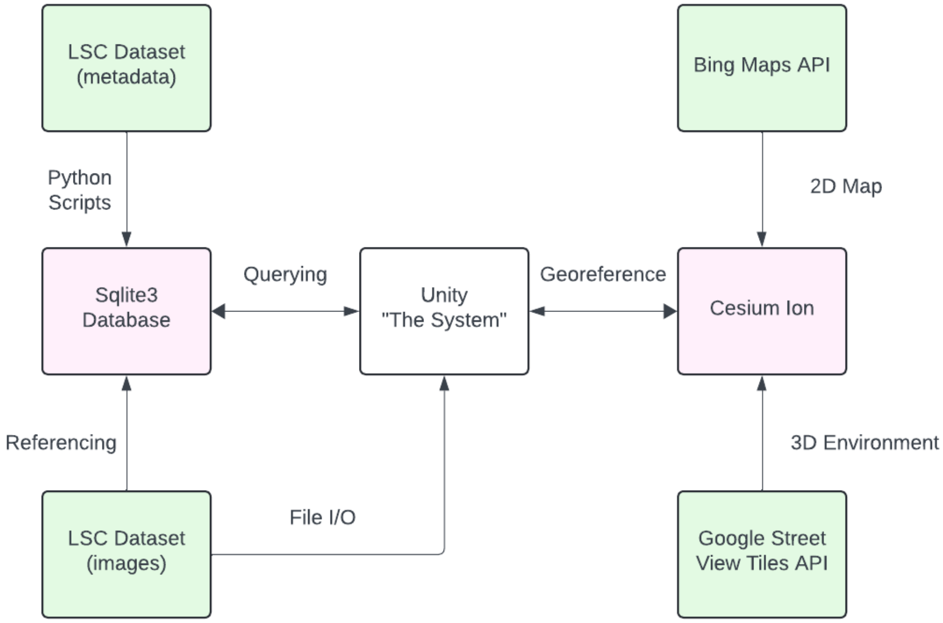

Figure 28: High-level technical overview

9.1.3 Main Hub

As described in section 8.1.1, the main hub facilitates the space to spawn and create walls based on temporal properties. In this same space, a visual tutorial and button layout are present. The visual tutorial consists of four parts:

1. Opening the wrist menu, figure 30
2. Toggling temporal filters, figure 30
3. Interacting with GIFs, figure 31
4. Starting 3D visualization tours, figure 31

The tutorial images switch every second from on, to off, to illustrate what happens when certain buttons get pressed.

The button layout image present in the main hub is shown in figure 29. This button layout helps the user to interact with the VR-LIVE.

As seen on the button layout, the user is able to move around using the joystick on the left controller. The joystick on the right controller can be used to rotate the camera around its Y-axis.

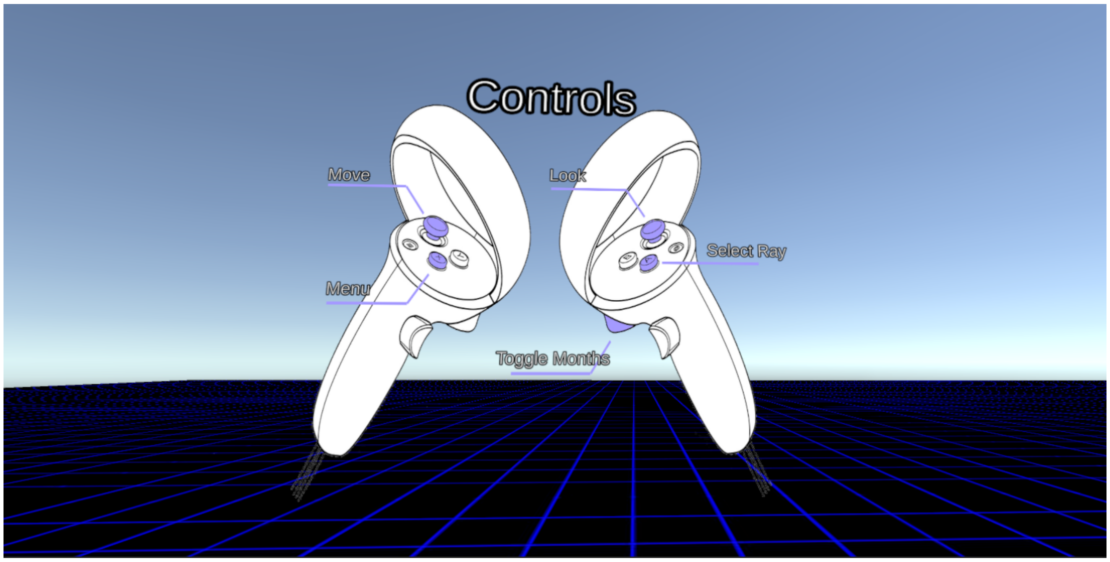

Figure 29: Button layout visualization to help users navigate VR-LIVE

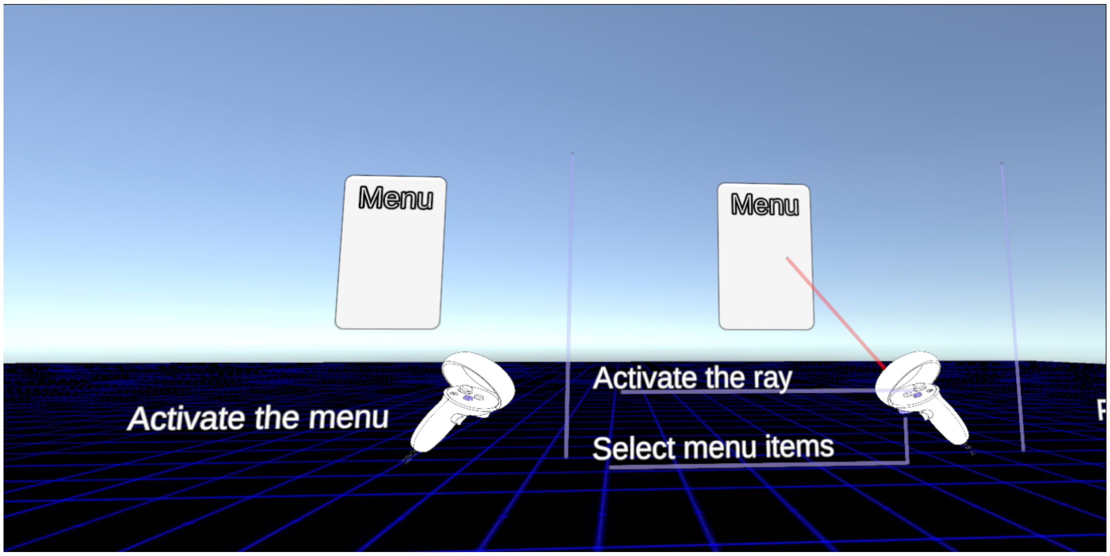

Figure 30: Tutorial images to help the user interact with the menu

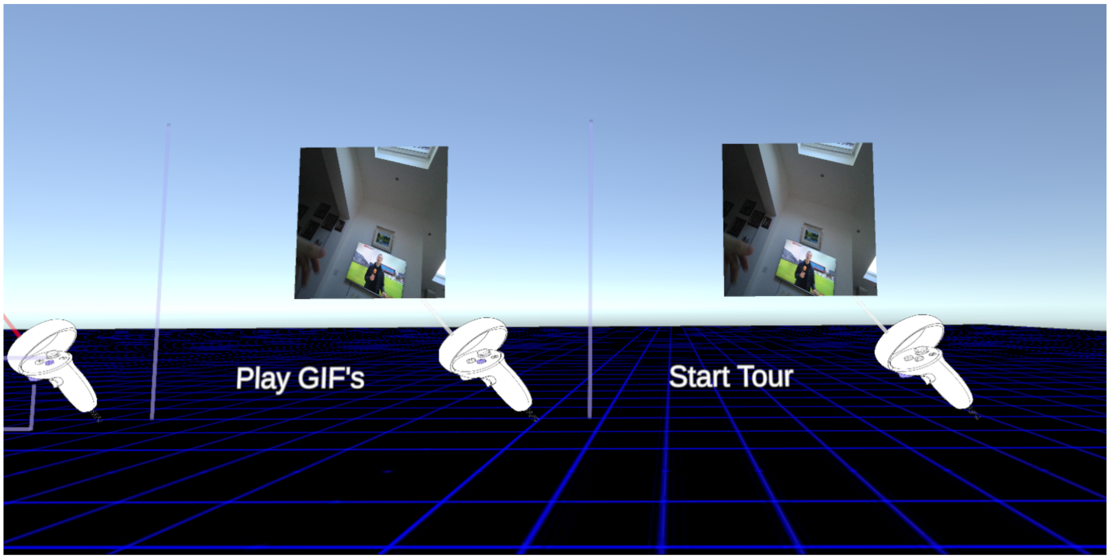

Figure 31: Tutorial images to help the user interact with the GIFs

9.1.4 GIF Generation and Wall Construction

This subsection will explain how the GIF generation is implemented. As described in section 8.1, the GIFs are constructed based on temporal properties. These temporal properties are the months of the lifeloggers activity, selectable from a list view in the wrist menu. Whenever the user clicks on a month (see figure 32), the following query is committed to the Sqlite3 database:

```
SELECT ImageId, minuteId, Longitude, Latitude, Altitude FROM
metadata WHERE ImageId LIKE month LIMIT 1;
```

The LIMIT 1 is to make sure the file I/O between disk and the application is limited, as mentioned before in 3.4.4. This ensures minimal loading times. Without it, for one month, an average of roughly 40.000 images had to be retrieved. This resulted in loading times of roughly a minute. With the LIMIT 1 in the query, loading times are less than a second. The month parameter in the LIKE clause in the query is formatted as YYYYMM.

```
SELECT ImageID, minuteId, Longitude, Latitude, Altitude FROM metadata
WHERE ImageID LIKE date;
```

The date variable in the like clause is formatted as follows: YYYYMMDD

The GIF thumbnails are spaced in a grid-like manner (figure 33), and the thumbnails are each 150x150 pixels big. The size of these thumbnails could be a variable in the interactivity of the user with the wall, but the 150x150 size seems to give a good balance between visibility and of the wall.

The GIFs each correspond to one day of the month, fluctuating between 28 and 31 GIFs per month. GIFs are sorted and created by day. When opting for GIFs split in different time intervals, say, multiple hours, we risk splitting events within the GIFs, making them less representable. When the user interacts with a give by hovering over

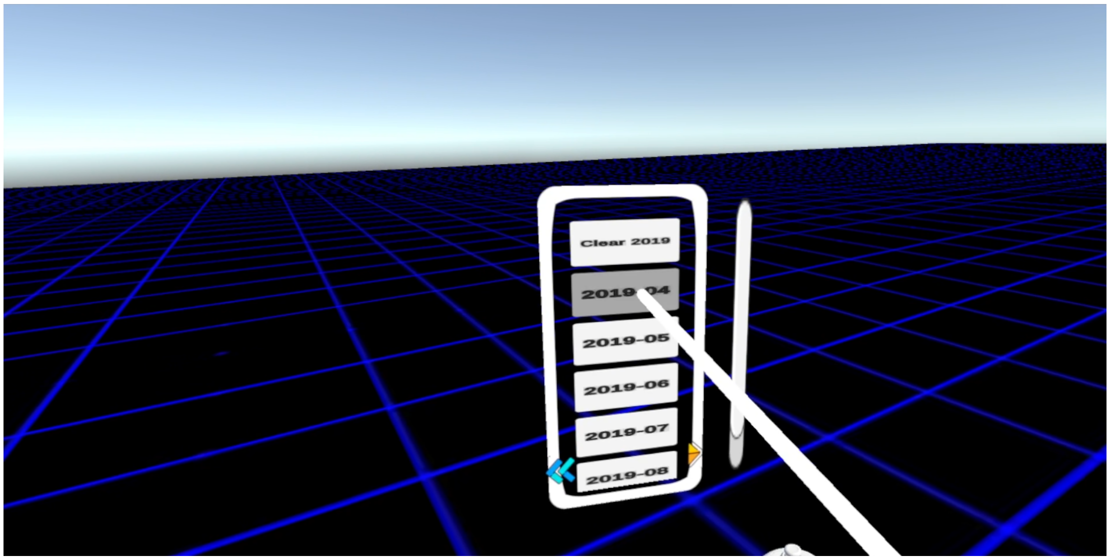

Figure 32: User selecting April 2019

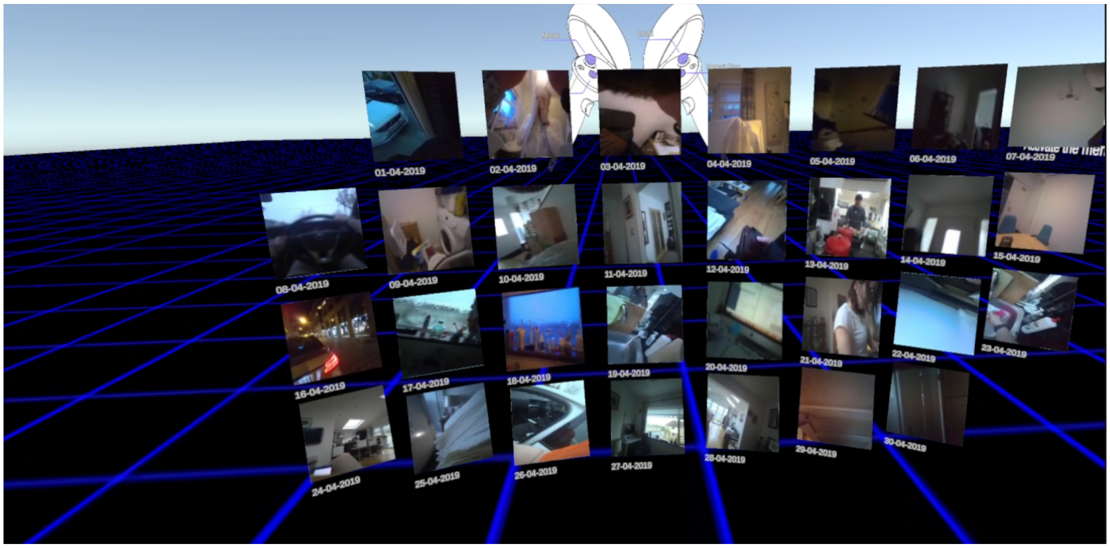

Figure 33: GIF wall of April 2019

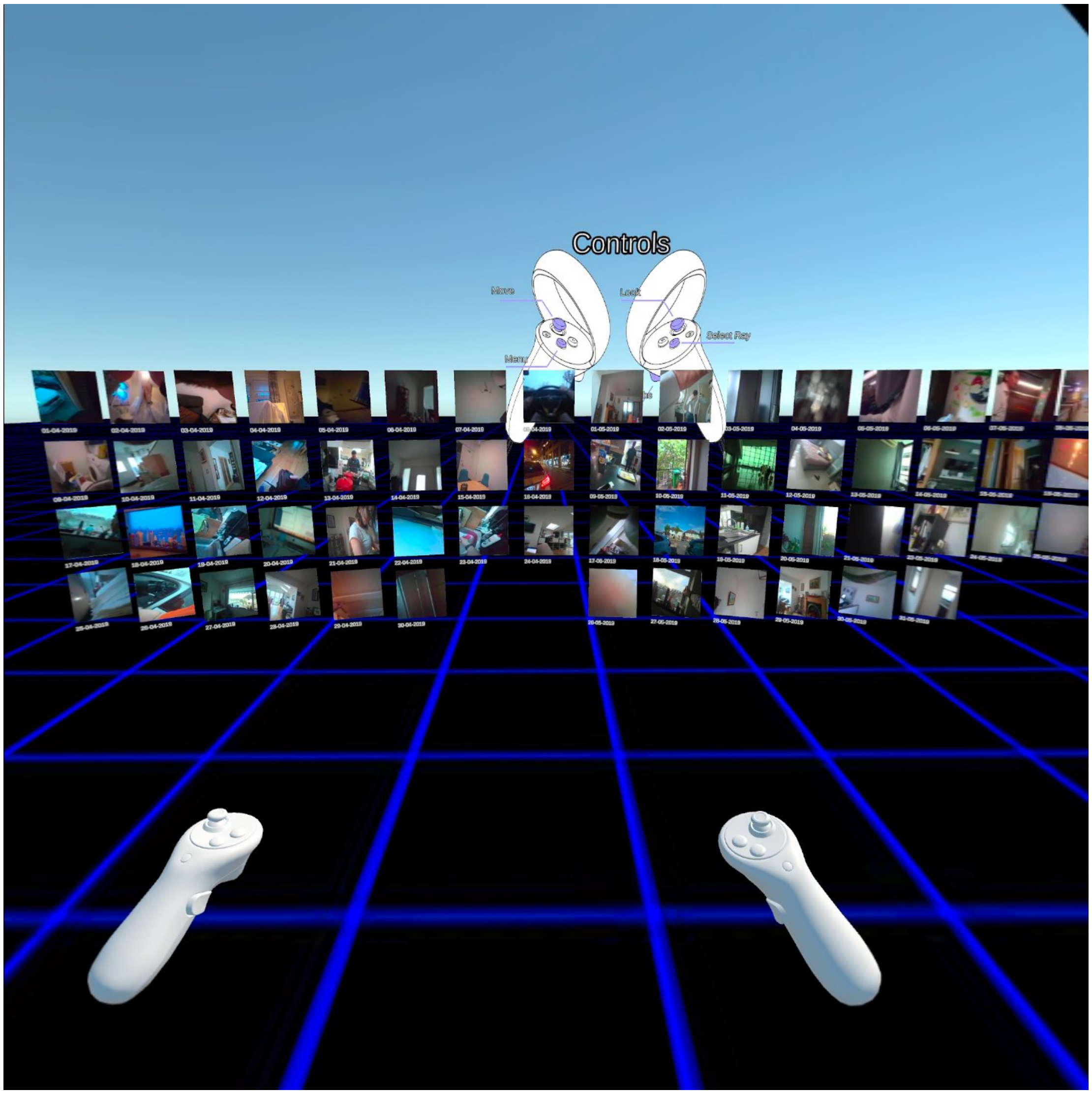

Figure 34: GIF wall of April 2019 and May 2019

it with the right-hand casting ray and pressing the select button, the second query is executed to retrieve the images for this GIF. After analysis, this file I/O is roughly 1 second. Per GIF thumbnail, the date is displayed in a DD-MM-YYYY format. If multiple months are selected, the wall will become longer (figure 34). Since pictures without accompanying GPS coordinates were omitted from the database, some days consist of no pictures.

When hovering over a GIF, the GIF plays the images retrieved by query 3.4.4 in order. This is done in the exact same manner as the GIF visualization, described in 3.4.4. The queries described in 3.4.4 make sure that there is no delay in clicking on the GIF, and playing the GIF. When the user stops clicking on the GIF, the GIF stops playing, and the user can inspect the image. For the GIF properties, such as play speed, the same properties are used as described in 3.4.4. An example of what this looks like is shown in figure 35.

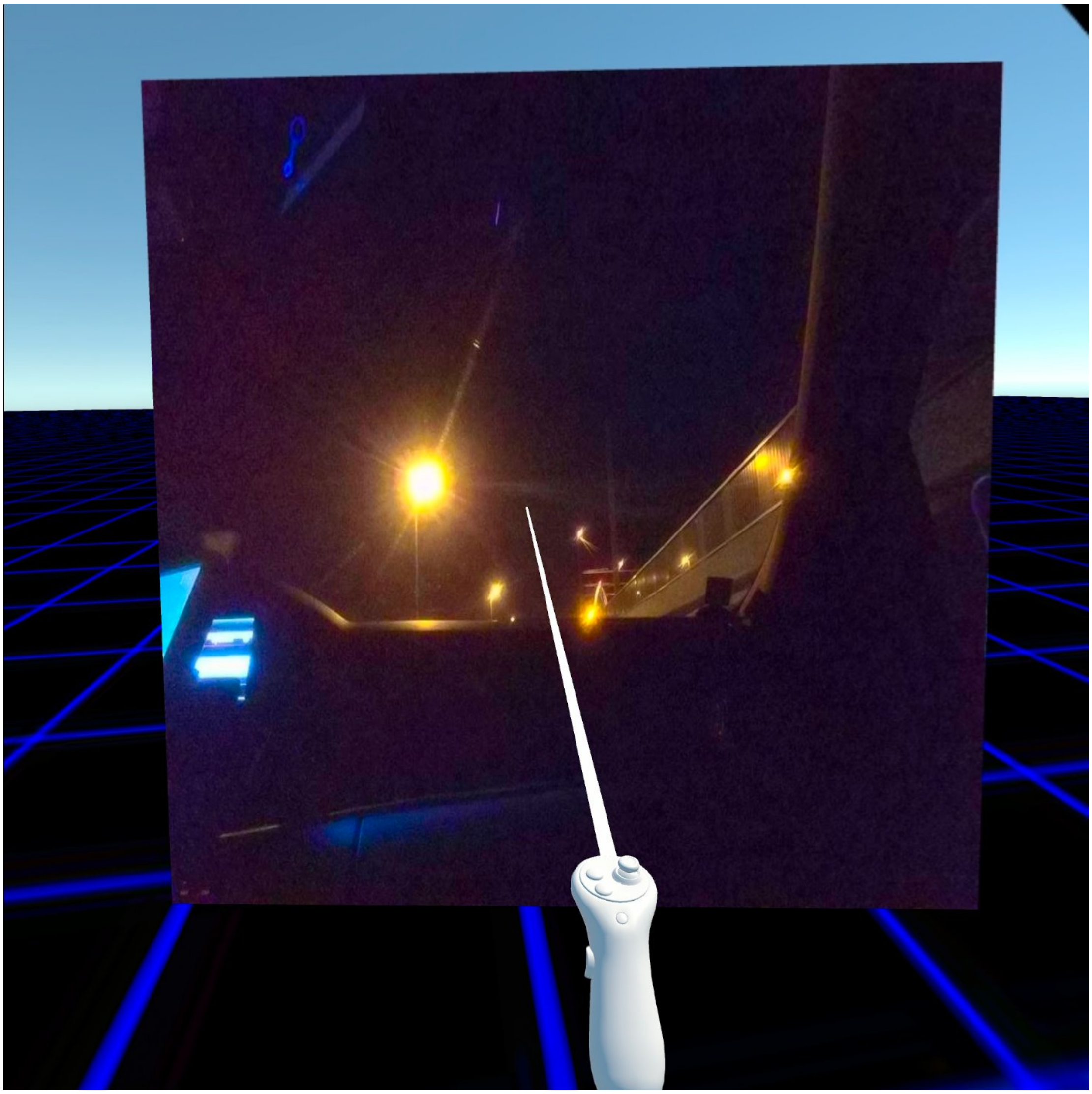

Figure 35: Inspecting and playing a GIF in VR-LIVE

Figure 36: The wrist menu in the 3D map visualization, with an "Exit Tour" button, and a visualization to make clear which button needs to be pressed to go to the next image

9.1.5 Street View Tour Visualization

This section will describe the decisions made in order to facilitate the user with an immersive experience in the Street View Tour. This is achieved by integrating the 3D map visualization described in section 3.4.4.3. As described, when the user selects a GIF in the wall, the Cesium API requests the needed tiles for the geolocation of the first location. Then, the user gets brought to the 3D environment. The user is located at the geographical location where the first image of the GIF was taken. The images viewed in this tour consist of all the images from the selected day. Only the images that have an accompanying GPS location are displayed in the sequence. The user can press the button on the left controller to advance to the next picture and location or keep it pressed to view the images back to back. An additional button on the left controller's wrist menu is implemented, in order to quit the tour before viewing all images. The user is returned to the wall. The wrist menu is shown in figure 36.

A consideration made for the 3D map tour was to include a scrollbar. This scrollbar would offer the functionality of quickly jumping to another image in the GIF during the tour. An average GIF of a day consists of roughly 1000+ images, and therefore it could be desirable to be able to jump to different sections of the day. However, when testing the functionality of the scrollbar, it was found difficult to make a user-friendly scrollbar. This was mainly due to the speed of the Cesium API. When jumping large gaps, the whole environment had to be reloaded, which took a few seconds. This broke the immersiveness proven by 7. Therefore, the scrollbar was omitted from the system.

Another consideration made was to make it possible to select sub-sequences of the GIF to make it more manageable to view a complete sequence in the tour. This would be in line with the efficiency of glancing at lifelog images 7. However, in the visualization research, the subsequences (snippets), where tailored, handpicked, and representative of their purpose. However, when having access to the whole database of 800.000 images, one can not guarantee that constructed sub-sequences (snippets) of lifelog days would be representative. This would then result into an optimization problem of constructing lifelog events, as it would be undesirable to cut certain events/trips into different sub-sequences. Arguing most days start and finish without interrupting an event (except for nightly events), the choice was made to not incorporate sub-sequences.

Whereas the user could not move around during the 3D map visualization in the conducted research, the user can now move around freely while viewing images. This is done in order to make it possible for users to rotate their (possibly own) images in the correct rotation, to match the original direction the picture was taken in. Some different orientations for the same picture is shown in figure 37. The user can move around using the joystick on the right controller, and rotate the camera with the joystick on the right controller. For the current version of VR-LIVE, it is impossible to determine the perfect orientation automatically, since the only data it has are the geolocations, and not the rotation. Whenever the user decides to click the 'next'-button, the user gets teleported to the original location before moving on to the next location.

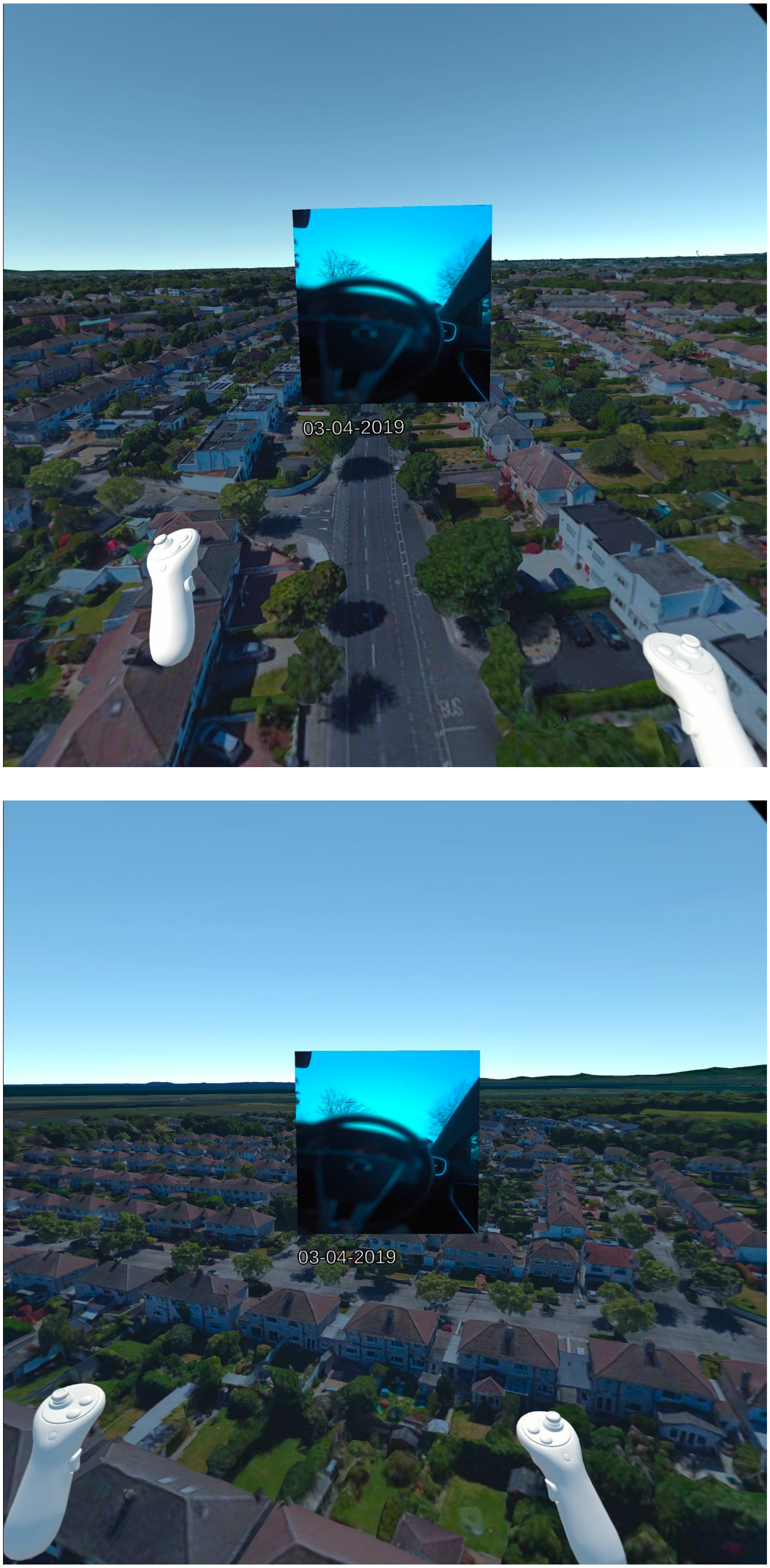

Figure 37: The same image viewed from different angles.

10 Review of System

In the end, a proof-of-concept system, VR-LIVE, has been created that provides an efficient and immersive experience to view lifelog images. Building on the existing literature regarding timelines [35], VR-LIVE builds a wall with efficient GIF visualizations, proven by the conclusions of the comparative study 7. Such a wall with GIF visualizations stimulates the user interests [49], boosting engagement and attention retention [51] within the VR environment. Therefore, this wall approach, with the GIF visualizations proven efficient by the conclusions of the comparative study 7, promises an efficient and captivating experience to browse a huge amount of lifelog images.

The combination of the GIF wall and the 3D map visualization is a unique way to explore lifelog images. Whereas the wall provides the efficiency to quickly look and glance at the lifelog images to get a feel what the lifelogger was doing at that day (see section 7), the 3D map provides an immersive (section 7) experience that can help the user recollect memories [60] [59]. Using the geolocation of the lifelog images in a realistic 3D Street View environment, the merge of real-life images and a digital 3D world, where location identification helps to estimate where images were taken [63]. As research [65] has hypothesized, VR-LIVE proves the feasibility of the combination of 3D Street View and lifelog images using the results from the researched visualizations.

This research has bridged multiple gaps:

- Gained insight into the efficiency of three different visualizations regarding viewing lifelog images.
- Gained insight into the immersiveness of three different visualizations regarding viewing lifelog images.
- Creating a stimulating, captivating, and efficient way to view lifelog images with a GIF wall.
- Creating an immersive lifelog image viewing experience that could help users recollect and relive memories.

With these research gaps filled, the proof-of-concept system provides a fundamental beginning for future research to build upon.

11 Future work

Based on the outcomes of the visualization research, and the implementation of VR-LIVE, the following items are suggested for future work:

- Besides the three visualizations designed, implemented and evaluated in this study, one could image more visualizations that would improve efficiency and / or immersiveness. These could be designed, implemented and compared to the visualizations designed in this study. Future work could provide more insight into the most optimal way to view and re-experience lifelog images.

- Because of time constraints, VR-LIVE was not evaluated by participants in a user study, or by experts in a heuristic evaluation. Different versions of the system could be implemented, adjusting the parameters of the system mentioned in sections 3.3 and 8.1. These parameters could be adjusted and tested, to further optimize the efficiency and immersiveness of the experience.

- Despite the performance optimizations and limitations on the number of images loaded from the disk in one time, there sometimes was still some delay in the file I/O of the system when loading the wall, especially when requesting multiple months of data at one time. Future work could further try to improve the performance of the file I/O, wall, and GIF construction.

- Despite the performance optimizations regarding culling and map detail, the Cesium API was still slow from time to time. Besides this, the Quest 2 headset was not powerful enough to run either the visualization environments, or VR-LIVE. Determined was that loading in the many polygons was the cause of this. Future work could look into improving the performance of the system, or making it compatible with more powerful hardware.

- As mentioned by some participants in the visualization study, there is interest in viewing their own lifelog images in such a system. Future work could provide with a generic porting framework, where people could load in their own geo-tagged images (e.g. from Google Photos), and view them in such a system.

- Future work could also look at different ways to generate and sort the images in the wall. One could think about classifying events in the lifelog data, and presenting an event wall. In such a system, the wall would be generated based on event filters instead of temporal filters.

Looking ahead, several avenues for future exploration and enhancement of the current research and system implementation present themselves. One promising direction is to delve deeper into the implications of the findings, exploring their broader applications and implications across different contexts. Additionally, refining the methodologies used could yield more nuanced insights and bolster the robustness of the conclusions drawn. Furthermore, integrating cutting-edge technologies or methodologies not previously explored could open new horizons for innovation and expand the capabilities of the system. These potential areas of development aim to build upon the foundational findings and capabilities established in this study, offering new opportunities for innovation and advancement in the field, ultimately contributing to the broader body of knowledge for lifelogging.

12 Acknowledgements

I would like to express my deepest gratitude to everyone who has supported me throughout my master's thesis journey. This work would not have been possible without the guidance, support, and encouragement of many individuals.

First and foremost, I would like to thank my primary supervisor, Dr. Wolfgang Hürst, whose expertise, patience, and insightful feedback were invaluable. I am also grateful to my co-supervisor, Dr. Michael Behrisch, for their invaluable advice and guidance.

Besides the support of my supervisors, I would also like to thank everyone who participated in this research.

References

[1] Federica Laricchia. Number of mobile devices worldwide 2020-2025, Mar 2023.

[2] Top Reviewing. The Evolution of Smartphone Cameras — bokasele101. https://medium.com/@bokasele101/the-evolution-of-smartphone-cameras-1713f0da59a3. [Accessed 21-02-2024].

[3] SenseCam - Microsoft Research — microsoft.com. https://www.microsoft.com/en-us/research/project/sensecam/. [Accessed 21-02-2024].

[4] S. Mann. Wearable computing: a first step toward personal imaging. *Computer*, 30(2):25–32, 1997.

[5] Amel Ksibi, Ala Saleh D. Alluhaidan, Amina Salhi, and Sahar A. El-Rahman. Overview of lifelogging: Current challenges and advances. *IEEE Access*, 9:62630–62641, 2021.

[6] Chelsea Dobbins, Madjid Merabti, Paul Fergus, and David Llewellyn-Jones. The big data obstacle of lifelogging. In *2014 28th International Conference on Advanced Information Networking and Applications Workshops*. IEEE, May 2014.

[7] J. Spragins. Unix for the impatient, second edition [new books]. *IEEE Network*, 10(1):6, January 1996.

[8] Michael Vassilakopoulos, Antonio Corral, Boris Rachev, Irena Valova, and Mariana Stoeva. *Image Database Indexing Techniques*, page 20–27. IGI Global, 2009.

[9] Abigail J. Sellen and Steve Whittaker. Beyond total capture: A constructive critique of lifelogging. *Commun. ACM*, 53(5):70–77, may 2010.

[10] Ki Joon Kim and S. Shyam Sundar. Does screen size matter for smartphones? utilitarian and hedonic effects of screen size on smartphone adoption. *Cyberpsychology, Behavior, and Social Networking*, 17(7):466–473, July 2014.

[11] Martin Dodge and Rob Kitchin. 'outlines of a world coming into existence': Pervasive computing and the ethics of forgetting. *Environment and Planning B: Planning and Design*, 34(3):431–445, June 2007.

[12] Cathal Gurrin, Alan F Smeaton, Aiden R Doherty, et al. Lifelogging: Personal big data. *Foundations and Trends® in information retrieval*, 8(1):1–125, 2014.

[13] Martin Dodge and Rob Kitchin. ‘outlines of a world coming into existence’: Pervasive computing and the ethics of forgetting. *Environment and Planning B: Planning and Design*, 34(3):431–445, 2007.

[14] Marc Bolaños, Mariella Dimiccoli, and Petia Radeva. Towards storytelling from visual lifelogging: An overview. 2015.

[15] Robin Van Geel. Joining temporal and geospatial representations of lifelogging data. `https://studenttheses.uu.nl/bitstream/handle/20.500.12932/45669/Thesis_R_K_van_Geel.pdf?sequence=1`. [Accessed 21-02-2024].

[16] Kevin Ouwehand. Geospatial Access To Lifelogging Images in VR GMT Master Thesis. `https://studenttheses.uu.nl/bitstream/handle/20.500.12932/31890/MSc_Thesis___Kevin___Geospatial_Access_To_Lifelogging_Images_In_VR_Igitur.pdf?sequence=2`. [Accessed 21-02-2024].

[17] Hidde Veer. Lifelog access in VR via filtering - universiteit utrecht. `https://studenttheses.uu.nl/bitstream/handle/20.500.12932/223/Lifelog_Access_in_VR_via_Filtering.pdf?sequence=1`. [Accessed 21-02-2024].

[18] Florian Spiess, Ralph Gasser, Silvan Heller, Luca Rossetto, Loris Sauter, Milan van Zanten, and Heiko Schuldt. Exploring intuitive lifelog retrieval and interaction modes in virtual reality with vitrivr-vr. In *Proceedings of the 4th Annual on Lifelog Search Challenge*, LSC ’21, page 17–22, New York, NY, USA, 2021. Association for Computing Machinery.

[19] *ICMR ’23: Proceedings of the 2023 ACM International Conference on Multimedia Retrieval*, New York, NY, USA, 2023. Association for Computing Machinery.

[20] Jana Machajdik, Allan Hanbury, Angelika Garz, and Robert Sablatnig. Affective computing for wearable diary and lifelogging systems: An overview. 01 2011.

[21] E. Wilding. THE HANDBOOK OF MEMORY DISORDERS. *Brain*, 127(7):1681–1682, 07 2004.

[22] Martin Conway and Christopher Pleydell-Pearce. The construction of autobiographical memories in the self-memory system. *Psychological review*, 107:261–88, 05 2000.

[23] Mary Czerwinski, Douglas W. Gage, Jim Gemmell, Catherine C. Marshall, Manuel A. Pérez-Quiñones, Meredith M. Skeels, and Tiziana Catarci. Digital memories in an era of ubiquitous computing and abundant storage. *Communications of the ACM*, 49(1):44–50, January 2006.

[24] Y. Chen and G. J. Jones. What do people want from their lifelogs? 2012.

[25] Niamh Caprani, Paulina Piasek, Cathal Gurrin, Noel E. O’Connor, Kate Irving, and Alan F. Smeaton. *Life-Long Collections: Motivations and the Implications for Lifelogging With Mobile Devices*, page 1153–1176. IGI Global, 2018.

[26] Zenonas Theodosiou and Andreas Lanitis. Visual lifelogs retrieval: State of the art and future challenges. In *2019 14th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)*, pages 1–5, 2019.

[27] Johnathan Bown, Elisa White, and Akshya Boopalan. *Looking for the Ultimate Display*, page 239–259. Elsevier, 2017.

[28] Isabell Wohlgenannt, Alexander Simons, and Stefan Stieglitz. Virtual reality. *Business amp; Information Systems Engineering*, 62(5):455–461, July 2020.

[29] J.M. Zheng, K.W. Chan, and I. Gibson. Virtual reality. *IEEE Potentials*, 17(2):20–23, 1998.

[30] S. Livatino, G. Muscato, and F. Privitera. Stereo viewing and virtual reality technologies in mobile robot teleguide. *IEEE Transactions on Robotics*, 25(6):1343–1355, December 2009.

[31] Mohamed El Beheiry, Sébastien Doutreligne, Clément Caporal, Cécilia Ostertag, Maxime Dahan, and Jean-Baptiste Masson. Virtual reality: Beyond visualization. *Journal of Molecular Biology*, 431(7):1315–1321, March 2019.

[32] Ekaterina Olshannikova, Aleksandr Ometov, Yevgeni Koucheryavy, and Thomas Olsson. Visualizing big data with augmented and virtual reality: challenges and research agenda. *Journal of Big Data*, 2(1), October 2015.

[33] Sanket Khanwalkar, Shonali Balakrishna, and Ramesh Jain. Exploration of large image corpuses in virtual reality. In *Proceedings of the 24th ACM international conference on Multimedia*, MM '16. ACM, October 2016.

[34] Aaron Duane and Cathal Gurrin. Pilot study to investigate feasibility of visual lifelog exploration in virtual reality. In *Proceedings of the 2nd Workshop on Lifelogging Tools and Applications*, LTA '17, page 29–32, New York, NY, USA, 2017. Association for Computing Machinery.

[35] Frank Hopfgartner, Yang Yang, Lijuan Marissa Zhou, and Cathal Gurrin. *User Interaction Templates for the Design of Lifelogging Systems*, page 187–204. Springer London, 2013.

[36] Aaron Duane. Visual access to lifelog data in a virtual environment, ph.d. thesis. 2020.

[37] Yalong Yang, Tim Dwyer, Bernhard Jenny, Kim Marriott, Maxime Cordeil, and Haohui Chen. Origin-destination flow maps in immersive environments. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):693–703, January 2019.

[38] Alexandar Jaffe, Mor Naaman, Tamir Tassa, and Marc Davis. Generating summaries and visualization for large collections of geo-referenced photographs. In *Proceedings of the 8th ACM international workshop on Multimedia information retrieval*, MM06. ACM, October 2006.

[39] A. Pigeau and M. Gelgon. Organizing a personal image collection with statistical model-based icl clustering on spatio-temporal camera phone meta-data. *Journal of Visual Communication and Image Representation*, 15(3):425–445, September 2004.

[40] Kentaro Toyama, Ron Logan, and Asta Roseway. Geographic location tags on digital images. In *Proceedings of the eleventh ACM international conference on Multimedia*, MM03. ACM, November 2003.

[41] Alice Thudt, Dominikus Baur, and Sheelagh Carpendale. Visits: A spatiotemporal visualization of location histories, 2013.

[42] Jau-Yuen Chen, C.A. Bouman, and J.C. Dalton. Hierarchical browsing and search of large image databases. *IEEE Transactions on Image Processing*, 9(3):442–455, March 2000.

[43] Thao-Nhu Nguyen, Tu-Khiem Le, Van-Tu Ninh, Minh-Triet Tran, Thanh Binh Nguyen, Graham Healy, Sinéad Smyth, Annalina Caputo, and Cathal Gurrin. Lifeseeker 4.0: An interactive lifelog search engine for lsc'22. In *Proceedings of the 5th Annual on Lifelog Search Challenge*, ICMR '22. ACM, June 2022.

[44] Ka-Ping Yee, Kirsten Swearingen, Kevin Li, and Marti Hearst. Faceted metadata for image search and browsing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI03. ACM, April 2003.

[45] Jonathan Helfman and James Hollan. Image representations for accessing and organizing web information. *Proceedings of SPIE - The International Society for Optical Engineering*, 4311, 01 2001.

[46] Benjamin Bach, Pierre Dragicevic, Daniel W. Archambault, Christophe Hurter, and M. Sheelagh T. Carpendale. A review of temporal data visualizations based on space-time cube operations. In *Eurographics Conference on Visualization*, 2014.

[47] Aiden R. Doherty and Alan F. Smeaton. Automatically segmenting lifelog data into events. In *2008 Ninth International Workshop on Image Analysis for Multimedia Interactive Services*. IEEE, 2008.

[48] Saeideh Bakhshi, David A. Shamma, Lyndon Kennedy, Yale Song, Paloma de Juan, and Joseph 'Jofish' Kaye. Fast, cheap, and good: Why animated gifs engage us. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, page 575–586, New York, NY, USA, 2016. Association for Computing Machinery.

[49] Yi Xu, Fan Bai, Yingxuan Shi, Qiuyu Chen, Longwen Gao, Kai Tian, Shuigeng Zhou, and Huyang Sun. Gif thumbnails: Attract more clicks to your videos. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(4):3074–3082, May 2021.

[50] Sarah Liang and Janet Wolfe. Getting a feel of instagram reels: The effects of posting format on online engagement. *Journal of Student Research*, 11(4), 2022.

[51] Kate M. Miltner and Tim Highfield. Never gonna gif you up: Analyzing the cultural significance of the animated gif. *Social Media + Society*, 3(3):205630511772522, July 2017.

[52] S. Shyam Sundar and Sriram Kalyanaraman. Arousal, memory, and impression-formation effects of animation speed in web advertising. *Journal of Advertising*, 33(1):7–17, 2004.

[53] B. Reeves and C. Nass. *The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places.* Cambridge University Press, 1996.

[54] Benjamin H. Detenber, Robert F. Simons, and Gary G. Bennett. Roll ‘em!: The effects of picture motion on emotional responses. *Journal of Broadcasting amp; Electronic Media*, 42(1):113–127, January 1998.

[55] Peter J. Lang. A bio-informational theory of emotional imagery. *Psychophysiology*, 16(6):495–512, November 1979.

[56] J.E. Crawley. Simulation studies of interface design. *IFAC Proceedings Volumes*, 4(3):129–137, 1971.

[57] Hairong Li and Janice L. Bukovac. Cognitive impact of banner ad characteristics: An experimental study. *Journalism amp; Mass Communication Quarterly*, 76(2):341–353, June 1999.

[58] Michael Gygli, Yale Song, and Liangliang Cao. Video2gif: Automatic generation of animated gifs from video. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1001–1009, 2016.

[59] Katherine A. Herdman. Investigating spatial memory and navigation in developmental amnesia: Evidence from a google street view paradigm, mental navigation tasks, and route descriptions, 2019.

[60] Dragomir Anguelov, Carole Dulong, Daniel Filip, Christian Frueh, Stéphane Lafon, Richard Lyon, Abhijit Ogale, Luc Vincent, and Josh Weaver. Google street view: Capturing the world at street level. *Computer*, 43(6):32–38, 2010.

[61] Ruofei Du and Amitabh Varshney. Social street view: blending immersive street views with geo-tagged social media. In *Proceedings of the 21st International Conference on Web3D Technology*, Web3D ’16. ACM, July 2016.

[62] Kohei Takahashi, Akira Shimojo, Shinsuke Matsumoto, and Masahide Nakamura. Mashmap: Application framework for map-based visualization of lifelog with location. In *2012 9th Asia-Pacific Symposium on Information and Telecommunication Technologies (APSITT)*, pages 1–6, 2012.

[63] Yuji Yoshimura, Shan He, Gary Hack, Takehiko Nagakura, and Carlo Ratti. Quantifying memories: Mapping urban perception. *Mobile Networks and Applications*, 25(4):1275–1286, May 2020.

[64] Robert E. Roth. Cartographic design as visual storytelling: Synthesis and review of map-based narratives, genres, and tropes. *The Cartographic Journal*, 58(1):83–114, September 2020.

[65] Cheryl Gilge. Google street view and the image as experience. *GeoHumanities*, 2(2):469–484, July 2016.

[66] SQLite Home Page — sqlite.org. `https://www.sqlite.org/`. [Accessed 04-03-2024].

[67] Meta Quest 2 Virtual Reality Headset. `https://www.meta.com/au/quest/compare/`. [Accessed 21-02-2024].

[68] Home — Virtual Desktop — vrdesktop.net. `https://www.vrdesktop.net/`. [Accessed 24-06-2024].

[69] Haversine formula - Wikipedia — en.wikipedia.org. https://en.wikipedia.org/wiki/Haversine_formula. [Accessed 17-02-2024].

[70] dev admin. Bing Maps API — Free Trial, Explore Features, Talk to Specialists — microsoft.com. https://www.microsoft.com/en-us/maps/bing-maps/choose-your-bing-maps-api. [Accessed 17-06-2024].

[71] Cesium for Unity Quickstart — cesium.com. https://cesium.com/learn/unity/unity-quickstart/. [Accessed 20-03-2024].

[72] Photorealistic 3D Tiles overview — Google Maps Tile API Google for Developers — developers.google.com. https://developers.google.com/maps/documentation/tile/3d-tiles-overview. [Accessed 20-03-2024].

[73] Georeference Origin Height not at 0,0,0 — community.cesium.com. https://community.cesium.com/t/georeference-origin-height-not-at-0-0-0/15189/4. [Accessed 14-04-2024].

[74] Likert scale - Wikipedia — en.wikipedia.org. https://en.wikipedia.org/wiki/Likert_scale. [Accessed 02-07-2024].

[75] Share the moment — Snapchat — snapchat.com. https://www.snapchat.com/. [Accessed 15-06-2024].

[76] Min Lu and Masatoshi Arikawa. *Map-Based Storytelling Tool for Real-World Walking Tour*, page 435–451. Springer Berlin Heidelberg, 2013.

[77] Matthew L. Lee and Anind K. Dey. Lifelogging memory appliance for people with episodic memory impairment. In *Proceedings of the 10th international conference on Ubiquitous computing*, UbiComp08. ACM, September 2008.

[78] Endel Tulving. Episodic memory: From mind to brain. *Annual Review of Psychology*, 53(1):1–25, February 2002.

[79] Tijmen van Teijlingen, Erik Oudman, and Albert Postma. Lifelogging as a rehabilitation tool in patients with amnesia: A narrative literature review on the effect of lifelogging on memory loss. *Neuropsychological Rehabilitation*, page 1–27, September 2021.

[80] How to Make a GIF - The Ultimate Guide to GIF — flexclip.com. https://www.flexclip.com/learn/how-to-make-a-gif.html#:~:text=How%20long%20can%20an%20animated,a%2015%20seconds%20plus%20GIF.). [Accessed 21-02-2024].

[81] Instagram Reels — sproutsocial.com. https://sproutsocial.com/glossary/instagram-reels/. [Accessed 14-02-2024].

[82] The Best Instagram Reels Length - ShareThis — sharethis.com. https://sharethis.com/social-media/2022/11/best-instagram-reels-length-to-boost-engagement/. [Accessed 14-02-2024].

[83] Frame rate - Wikipedia — en.wikipedia.org. https://en.wikipedia.org/wiki/Frame_rate. [Accessed 21-02-2024].

[84] SVGator. Animated gif best practices for gif optimization, Jan 2024.

[85] James Ash. Sensation, networks, and the gif: Toward an allotropic account of affect. *Networked affect*, pages 119–133, 2015.

[86] Unity Real-Time Development Platform — 3D, 2D, VR & AR Engine — unity.com. https://unity.com/. [Accessed 05-03-2024].

[87] ListView — GridView in Unity GUI System — GUI Tools — Unity Asset Store — assetstore.unity.com. https://assetstore.unity.com/packages/tools/gui/listview-gridview-in-unity-gui-system-109577. [Accessed 05-03-2024].

[88] Task Parallel — Integration — Unity Asset Store — assetstore.unity.com. https://assetstore.unity.com/packages/tools/integration/task-parallel-82257. [Accessed 05-03-2024].

A Appendix

A.1 Consent Form

Consent Form
Research VR Lifelogging visualization
Utrecht University
Yannick Visser (supervisor Wolfgang Hürst)

The focus of this research is to study three different visualisations to view lifelogging images in VR. Virtual reality is a fabricated reality view through a headmounted display, like the device in front of you. A lifelog represents a personal log of one's daily existence. For this research, we will be looking at passively gathered lifelog images.

Risks, Discomforts and benefits Participation in this study should be an interesting and enjoyable experience, and the results obtained are expected to make a scientific contribution to the field of computer science.

Yet, when using virtual reality (VR) systems, some people may experience different levels of the following: nausea, vomiting, sweating, pallor, headaches, disorientness, vertigo and/or dizziness. Participants experiencing any of these or other discomforts must stop with the test immediately and inform the testing researcher.

Using VR apps and games have the possibility of inducing epileptic episodes. Therefore, people who are known to suffer from epilepsy are not allowed to participate in this experiment.

Confidentiality Any information that is shared during the study will be treated strictly confidential and handled anonymously. Throughout the study only the aforementioned researchers will have access to the information. The results of this study could be published, but the results will not lead to the possibility of identifying individuals.

Request for Further Information You are encouraged to discuss any concerns regarding the study with the testing researcher at any time, and to ask any questions that you might have.

Refusal or Withdrawal You may refuse to participate in the study and if you do consent to participate then you will be free to withdraw from the study at any time without consequence, fear or prejudice. If you wish to withdraw from the event please contact the researcher and all data pertaining to you will be destroyed.

I have read the consent sheet	YES / NO
I have had the opportunity to ask questions about the procedure	YES / NO
All my questions were answered to my satisfaction	YES / NO
I have received sufficient information about the study	YES / NO
I Understand and accept the risks associated with the use of virtual reality	YES / NO
I certify to have no knowledge of past epileptic episodes	YES / NO
Name	
Date	
Signature	

A.2 Introduction to experiment

Your task is to explore lifelog snippets, that is, temporarily sorted sequences of images recorded throughout a day by someone wearing a lifelog camera. You will do this with three visualizations and watch two snippets per visualization. After pressing a button, you will watch the snippets with the respective interface. Afterward each interface, I will ask you the following questions:

Please rate the following statements on a scale from 1 (strongly disagree) to 10 (strongly agree):

- I feel confident that I have a good understanding of the activities done by the person wearing the lifelog camera.
- I feel confident that I have a good understanding ofhow the person wearing the lifelog camera was moving around.

These are images of the visualizations to give you a feel of what you are about to see. In order from to right, the generic visualization to select snippets, the GIF visualization, the 2D visualization, the 3D visualization

A.2.1 Written Responses Post Experiment Questionnaire

The written responses for the post experiment questionnaire are filtered. Filtered out responses include responses that are empty, or a non-logical response (for example, ”-”, ”None”, ”Nothing”, etc.). Responses consisting of multiple answer are broken up into seperate points.

A.2.1.1 Advantages of GIF Visualization

- This experience felt basic, yet it felt trusted
- this was nice and easy to view. no complications
- There were no distractions. This helped looking at the pictures better.
- no distractions, so I could focus on the video.
- Minimalistic design made it easy to focus on the GIF
- There was no lag
- It was convenient to view one GIF after the other.
- It was nice that there was no ’loading’ of the environments in this representation which made it quick and smooth
- There were no distractions.
- It is was very efficient in regards to time
- you get a quick summary of the activities
- I could focus solely on the images.
- It was a decent way to quickly see the images.

- It was a familiar way to view timelapses (apart from the VR aspect), but very similar to things like youtube where you can view the thumbnails
- it was a clear interface with no distractions which is nice if you actually want to focus on the images
- nice big view, quick and fast.
- You get to see the images more like a video, so you quickly get an overview and get a better sense of the motions made by the wearer. Also, no waiting for loading the scene or images is required.

A.2.1.2 Disadvantages of GIF Visualization

- this experience was a bit boring, and it was difficult to see where the person went etc.
- a bit boring
- More difficult to think formyself to see what the person was doing. It costs more effort.
- Compared to the other ones, this one was slightly 'lame'
- The images went by a bit fast to see what the person was doing
- I didnt like that there wasnt anything else to see besides the pictures
- Obviously, it was less 'cool' than the others
- This one didnt really benefit from the VR, as it was quite static.
- A disadvantages was that the images didnt convey a lot of information about what was happening and where it was happening. Another disadvantage was that it wasnt really as immersive as the other ones.
- it wasnt very immserive.
- If I keep immersiveness in mind, this one was the least immersive. It lacked a sense of environment
- It did not show where he was going, for example, on a map. It was just pictures and not a true visualization of his route.
- you don't get a feeling of time and you don't know where it was
- I did not feel immersived in the experience.
- The images went by in a abrupt fashion, which made it difficult to understand the sequence of events.
- It was not as interesting as the other two. I feel like it could have been done in a normal device and not in VR
- When comparing it to youtube, it would be nice to have more options to hover over, even while one timelapse / thumbnail is playing

- the images were poor so not much information could be deducted from just viewing the images (especially like previous questions mentioned, about movement etc.)
- it would be good for navigation if the other gif didnt dissapear (kind of like social media)
- No geolocation reference.

A.2.1.3 Advantages of 2D Visualization

- much easier to see movement throughout the snippets.
- for sure more interesting than the gif one
- The map itself was fun. The images followed the road nicely.
- At some point in one of the GIFs there was a restaurant / coffeeshop 'scene', next to apparently a lake. My imagination filled in the scenery for me!
- I thought it was nice to see where the pictures were taken, and that you were taken along the roads.
- The map was fun to see
- Interesting point of view.
- Good viewing pace which seemed to resemble actual movement
- The map is clear and not very distracting.
- The streetnames were nicely visible.
- An advantage was that it was nice to see where the person was / went.
- It was also nice to see the exact streetnames
- it was fun being taken along the images and seeing where they were taken, like snapchat maps kind of
- Not a lot of distractions, but still more immersive than the GIF one. Besides that, more environment sense than the GIF one
- I liked how I was taken on his journey on a map, whilst also seeing the pictures in front of me.
- It was effiecient and very clear.
- you know where he was, the streets, parks etc.
- The visusalisation was simple and easy to follow.
- The movement was implemented well, where you could really see the car ride in the correct directions etc.
- there were not a lot of distractions.

- for movement, it was more clear where the person went to. the map itself looked good.
- for travel behaviour definitely better than the gif one
- Geolocation reference and the fact that you could go through the images more quickly when the images where taken at the same location.

A.2.1.4 Disadvantages of 2D Visualization

- it was slightly distracting
- the map was a bit distracting, as it had to be loaded sometimes
- The map was not the most pretty.
- I was looking at the names of the roads more than actually paying attention to the images.
- The google maps was distracting, I was looking more at the maps than at the pictures
- I didnt like how the map looked. The white was very bright and It was not pleasant to look at
- The map itself was quite plain
- The map was less impressive and immersive than the 3D view
- A disadvantage was that it didnt look as good as the 3D visualization
- the map was not that good looking
- It was not that inviting to look down, just to see the names of the streets. It was not visibly appealing
- Even though it was efficient in terms of clarity, it was more time consuming.
- it felt a bit strange to ”fly” above the map.
- Because the map had similar color all around, it was hard to distinguish the locations from each other.
- The map itself felt a bit abstract, which I dont know if it fits the purpose.
- sometimes the camera(?) moved a little bit when the images didnt indicate any movement.
- it was a bit weird to be on a map, instead of having it infront of you.
- the map itself didnt add much apart from giving a sense of distance traveled.
- Hovering above a 2D map breaks the immersion a bit

A.2.1.5 Advantages of 3D Visualization

- impressive view, very nice to look around, and would love to visit for example my old holiday destinations e.g.
- very cool, looks really good and fun to be taken a long a small section of this random person i dont know
- Obviously the most pretty. It was like I was flying with a helicopter and seeing the images where they were taken
- Very nice to see more of the environment and being able to look around and seeing the setting
- I got a realistic view of where the person was driving.
- It was less distracting than the maps.
- Very nice to see the colors of the building and trees
- This one was really cool, and the places I was taken to looked like i was there in a drone
- Really looked like I was there
- Same as the 2D, the movement seemed to resemble the ’movement’ in the pictures.
- The ’car ride’ was really fun
- I have never seen such a thing in VR so it was really cool.
- The environment looked pretty.
- An advantage was that it looked really good (or immersive as mentioned above).
- Other advantages where the same as for the 2D visualization
- this one was even more fun than the 2d map.
- It looked really good.
- This one was very immersive.
- Very interesting to see the 3D buildings etc. I have seen Google earth before, but not in VR, which incombination with the images was very fun to see.
- I liked how Google Earth was incorporated and it gave a very realistic view of his journey.
- The 3D visualization made me feel like I was there and going on his journey with him. you get a better feeling about the place, the buildings.
- you really see where he was and what that building looked like; was it big small, which color etc. you can also look around and get a feeling of the whole environment
- I felt immediately immersed in the world.

- It was enjoyable to look around.
- I had a sense of space and how much distance the was between the locations.
- The visuals were impressive. Gave a good sense of environment and movement.
- The environment was very cool, and really interesting to see it in VR.
- never seen something like that
- it looked really good, almost like i was there.
- the 'car rides' where really fun
- very cool, kind of like you are there. immersive.
- Same as 2D: Geolocation reference. But also that it is on a 3D map, providing a better sense of geolocation reference and immersion (distinguish buildings and other places better).

A.2.1.6 Disadvantages of 3D Visualization

- maybe there could be a minimap or something to see where you are as with the 2d
- it was a bit weird to see the environment being loaded while watching images.
- A bit nitpicky: sometimes the road was not followed perfectly
- The looking around has the same drawback as the 2D map
- it was a little more distracting.
- Sometimes I noticed that the 3D objects had to be loaded. It then didnt look as good.
- Sometimes the 3D objects looked a little weird.
- When it turned the view, it made me slightly disoriented.
- for the first image it took a while for the scene to be loaded.
- While it was very fun to look around, it was also a bit distracting since it was quite overwhelming, especially since I havent seen anything like this before
- also a bit strange to "fly" above the map.
- Some pictures were taken inside a building, while the camera was on top of the building. This felt a bit weird
- sometimes the buildings still had to load which made it look a little ugly for a few seconds when loaded into the visualization
- When you move a significant distance compared to the previous image, it takes time to load the world, which impacts the immersion a bit

A.2.1.7 General Comments

These General Comments are slightly filtered; comments that were not relevant to the experiment, such as "goodluck with your thesis", were omitted.

- maybe there could be a minimap or something to see where you are as with the 2d, it was a bit weird to see the environment being loaded while watching images.
- A bit nitpicky: sometimes the road was not followed perfectly
- The looking around has the same drawback as the 2D map; it was a little more distracting.
- Sometimes I noticed that the 3D objects had to be loaded. It then didnt look as good.
- Sometimes the 3D objects looked a little weird.
- When it turned the view, it made me slightly disoriented.
- for the first image it took a while for the scene to be loaded.
- While it was very fun to look around, it was also a bit distracting since it was quite overwhelming, especially since I havent seen anything like this before
- also a bit strange to "fly" above the map.
- Some pictures were taken inside a building, while the camera was on top of the building. This felt a bit weird
- sometimes the buildings still had to load which made it look a little ugly for a few seconds when loaded into the visualization
- When you move a significant distance compared to the previous image, it takes time to load the world, which impacts the immersion a bit