Visual Access to Lifelog Data in a Virtual Environment

Aaron Duane

Supervised by Dr. Cathal Gurrin

A thesis presented for the degree of Doctor of Philosophy (Ph.D.)

School of Computing

Dublin City University

July, 2019

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed:

ID No.: 15213058

Date: July 2019

Acknowledgements

This thesis would not have been possible without the support of my friends and especially my family. To my parents, Deirdre and Brendan, thank you for always being there for me in my time of need; your constant encouragement and dedication to our family can never be understated. To my brother, Ashley, thank you for always having my back and pushing me when you know I need pushing. To my girlfriend, Joanne, thank you for never ceasing to be a pillar of support throughout this entire process; both of us started our PhDs together and we are going to finish them together. I would also like to thank my supervisor, Cathal, who always made time to meet with me when I was having doubts about my research and knew exactly what to say to steer me back on track. And finally, to my remaining colleagues at Insight and DCU, thank you for the camaraderie over the past few years as this thesis finally came together.

Contents

List of Figures

List of Tables

Abstract

Continuous image capture via a wearable camera is currently one of the most popular methods to establish a comprehensive record of the entirety of an individual's life experience, referred to in the research community as a lifelog. These vast image corpora are further enriched by content analysis and combined with additional data such as biometrics to generate as extensive a record of a person's life as possible. However, interfacing with such datasets remains an active area of research, and despite the advent of new technology and a plethora of competing mediums for processing digital information, there has been little focus on newly emerging platforms such as virtual reality. We hypothesise that the increase in immersion, accessible spatial dimensions, and more, could provide significant benefits in the lifelogging domain over more conventional media. In this work, we motivate virtual reality as a viable method of lifelog exploration by performing an in-depth analysis using a novel application prototype built for the HTC Vive. This research also includes the development of a governing design framework for lifelog applications which supported the development of our prototype but is also intended to support the development of future such lifelog systems.

Chapter 1

Introduction

1.1 Overview

There is no universally accepted definition of lifelogging. In its broadest sense, it can be observed as the comprehensive recording of any data pertaining to a person's life or life experience. However, this level of ambiguity is not very helpful to researchers who often have opposing classifications and approaches. Dodge and Kitchin (2007), for example, define lifelogging as "a form of pervasive computing, consisting of a unified digital record of the totality of an individual's experiences, captured multimodally through digital sensors and stored permanently as a personal multimedia archive". This suggests a much more rigorously defined form of lifelogging that is aimed at nothing less than recording the totality of human experience, yet in practice this type of lifelogging is never actually achieved, as we are still wholly incapable of recording the totality of human experience. As a result, this description more accurately conveys a set of principles, than a specific definition.

In contrast to this, there are efforts like the Quantified Self movement (Barrett et al., 2013; Meyer et al., 2014; Hoy, 2016), which deliberately avoid trying to capture the totality of a person's experiences, instead seeking to capture data on targeted aspects of a person's life such as diet, mood or physical performance in a domain-focused effort to address specific use cases or research questions. We also have more commercial forms of lifelogging in the rapidly growing digital health sector. Applications released by companies like Google[1] and Apple[2] seek to record data related to a user's health and lifestyle. This extends the already well-established industry of recording users' social media life experiences in the form of photos, videos and other online documents like emails or tweets. Regardless of how one defines lifelogging, it is clear that it is a rapidly growing area of focus from both a research and commercial perspective.

However, one consistent aspect of collecting life experiences is the generation of tremendous amounts of data (Gurrin, Smeaton, and Doherty, 2014). Visually accessing very large datasets to garner actionable insight is not an especially new phenomenon and there has already been research in the design community regarding different visualisation practices and techniques (Wang, Guanghui, and Alexander, 2015; Doherty, Caprani, et al., 2011). However, visually accessing large datasets devoted specifically to lifelog data remains an open area of research, especially when it comes to datasets focused on total capture as opposed

[1]Google Fit: https://www.google.com/fit/
[2]Apple Health: https://www.apple.com/lae/ios/health/

to selective or situation-specific capture. One of the most common methods of attempting to record the totality of an individual's life experiences is through the continuous capture of images via a wearable camera, often worn around a person's neck. These cameras can autonomously capture several thousand images per day and will very quickly generate enormous datasets that become far too unwieldy for anyone to explore manually.

To address this issue, researchers have relied on content analysis such as computer vision to organise, segment and annotate this type of lifelog data to better support summary and exploration (Doherty and Smeaton, 2008; Lee et al., 2008). However, without an appropriate mechanism for users to efficiently interact with this metadata, the lifelog dataset only becomes more difficult to navigate or explore. This reinforces how important human-computer interaction is when it comes to interfacing with large lifelog datasets enriched by machine learning techniques. This philosophy not only extends to good user experience and user-centred design, but also to the choice of hardware being used. For example, researchers have already explored the potential of various hardware platforms such as laptops, tablets and phones (Yang, Lee, and Gurrin, 2013; Qiu, Gurrin, and Smeaton, 2016), but there has been very little research into other less conventional platforms such as virtual reality. In this work, we intend to motivate virtual reality as an effective new candidate for visual access to lifelog data and introduce a framework for developing effective user-centred lifelog interfaces for both conventional and unconventional hardware platforms. The subsequent sections of this chapter introduce the various components of this research and set forth the hypothesis and research questions which will be investigated.

1.1.1 Lifelogging

The concept of lifelogging can be traced back as early as 1945 with Vannevar Bush's vision of a 'Memex' (Bush, 1945), a portmanteau of memory and index. The Memex was envisioned as a type of desk with pulleys and levers which would enable the rapid archival of personal information related to documents interacted with at the time. However, early research into the modern concept of lifelogging didn't begin until the 1980s when pioneers such as Steve Mann, often referred to as the father of wearable computing, started to develop increasingly smaller and more sophisticated wearable sensors. Mann has developed many different wearable camera technologies and addressed many early personal imaging challenges which are outlined in Mann (1997), as well as the eye-tap system described in Mann (2004). In more recent years, Mann has developed solutions to some of the common challenges of gathering lifelog data from wearable computers such as supporting different lighting levels to avoid capture blackout or whiteout, outlined in Mann et al. (2012). This technique is now widely deployed as the HDR feature on modern digital cameras.

Another pioneer worth noting in the area of lifelogging is Gordon Bell, known famously for formulating Bell's law of computing systems (Bell, Chen, and Rege, 1972). Bell became the experimental subject of a project which came to be known as 'MyLifeBits', described in Gemmell, Bell, and Lueder (2006). The MyLifeBits project intended to collect a lifetime of storage on and about Gordon Bell and included full-text search, audio and text annotations, as well as hyperlinks. This experiment in lifelogging was an attempt to fulfil Vannevar Bush's original vision of an automated store of the things an individual has experienced in his or her lifetime, to be accessed with speed and ease. To achieve this, Bell and his assistants digitised all documents he had ever read or produced; gathering web pages browsed as well as phone and instant messaging conversations. The book 'Total Recall' (Bell and Gemmell, 2009) describes the vision and implications for a personal, lifetime e-memory archive for recall, work, health, education and immortality, thereby offering a first view of the potential of a holistic lifelogging application.

As previously stated, the concept of lifelogging is not strictly defined, but researchers have noted two subcategories of lifelogging which have become most prevalent: total capture and situation-specific capture (Sellen and Whittaker, 2010). As the name suggests, the focus of total capture lifelogs is a complete record of an individual's life, capturing as many kinds of data as possible, as continuously as possible. These datasets most often contain biometric data, such as heart rate or blood sugar levels and human activity data, such as location and diet, but the most common approach to total capture lifelogs is the recording of a continuous stream of images from a wearable camera. Situation-specific capture on the other hand is more focused in scope and typically aims to capture rich data in specific domains. Some researchers define it as a specialised form of lifelogging, where the goals are often clearer or better defined. The Quantified Self [3] movement is an example of a situation-specific lifelog movement as it aims

[3]Quantified Self: http://quantifiedself.com/

to capture targeted datasets to assist with human health and wellness. However, even the simple recording of audio or video at a workplace meeting or whiteboard session can be defined as a form of situation-specific lifelog capture (Bahrainian and Crestani, 2017a; Bahrainian and Crestani, 2017b; Bahrainian and Crestani, 2018), where the goal is to preserve any ideas or decisions which could be lost afterwards.

The scope of the research outlined in this work focuses on total capture lifelog datasets, specifically those which utilise continuous image capture as the primary method of capturing the entirety of life experience. Critics of this form of lifelogging (Sellen and Whittaker, 2010), have suggested that many such systems ”lack an explicit description of potential value for users”, instead focusing on overcoming technical challenges. They state that memory refers to a complex, multi-faceted set of concepts, and therefore there are different types of memory, which all need to be catered in different ways. Furthermore, they stress that while digital data collections can serve as cues to trigger autobiographical memory about past events, these are not memories themselves. This rhetoric emphasises the importance of understanding the precise relationship between the cues we are able to capture and the memory experiences they trigger. This will enable lifelogging systems to strategically target the weaknesses of human memory and evaluate a system's effectiveness based on the type of memory it is trying to cater for. These principles are employed throughout this study and are expanded upon in later chapters.

1.1.2 Virtual Reality

The exact origins of virtual reality are disputed, but there have been many systems developed over the years that have exhibited elements of virtual reality. For example the first stereoscope, a device which positions two photographs of the same object at slightly different angles to create an impression of depth, was invented as far back as 1832 by Sir Charles Wheatstone (Gregory, 1997). One of the earliest systems to utilise moving pictures was the Sensorama (Heilig, 1960), created in 1961 by Morton Heilig, which sought to become a new and immersive way to experience film. It provided a wide field of view, stereo sounds, seat tilting, vibrations, smell, and even wind. However, the first examples of commercially viable virtual reality systems didn't appear until the 1980s and 1990s when attempts were made by companies like Virtuality[4] and Nintendo, see the Virtual Boy[5], but the limitations of these systems at the time failed to resonate with consumers and they were ultimately discontinued. It has only been in the last few years, largely due to advances in mobile hardware for smartphones, that virtual reality has seen a sudden resurgence, with large tech companies such as Google, Facebook and Valve developing modern head-mounted displays such as the HTC Vive and the Oculus Rift.

However, the relative infancy of these new virtual reality platforms means that a ubiquitous hardware standard has not been established and user interaction and design is still an ongoing area of research. Unlike conventional platforms,

[4]Virtuality: https://en.wikipedia.org/wiki/Virtuality_(gaming)
[5]Virtual Boy https://en.wikipedia.org/wiki/Virtual_Boy

such as laptops or phones, there are not well defined best practices for many VR interaction tasks. Also, the slower adoption rate of virtual reality, when compared to something like the smartphone industry, means the full potential and application of virtual reality remains unexplored. To date, the primary focus of platforms like the HTC Vive and Oculus Rift have been on entertainment, with a particular emphasis on video games. Other more casual VR platforms like the Google Daydream or the Samsung Gear have attempted to market more affordable hardware with emphasis on not just video games, but on accessing the world wide web through virtual reality as well. As the platform continues to mature, and the hardware becomes increasingly lightweight and affordable, it is feasible to assume its applications will become more varied as well. This is compounded by the existence of advancing mixed reality platforms, like the Microsoft HoloLens[6], which could enable us to experience a virtual reality mapped entirely on top of our own reality. The convergence of virtual and augmented reality into a completely mixed reality could signal the end of traditional screen-based platforms as we know it, where digital information is rendered to any surface, at any place and at any time (Garon et al., 2017; Kalantari and Rauschnabel, 2018).

Yet even before this future is realised, there is still much potential for virtual reality and how it might enhance information retrieval in its current form. Some researchers believe the most valuable aspect of the platform is its highly immersive quality and the degree to which it projects stimuli onto the sensory receptors of users in a way that is extensive, matching, surrounding, vivid, interactive and

[6]Microsoft HoloLens: https://www.microsoft.com/en-us/hololens

plot informing (Slater and Wilbur, 1997). There has also been well-established research indicating that actively using more of the human sensory capability and motor skills has been known to increase understanding and learning (Dale, 1969) and more recent research has suggested that immersion can greatly improve user recall (Krokos, Plaisant, and Varshney, 2018). Many virtual reality interactions are also highly intuitive, likely because engaging with digital elements directly in a three-dimensional medium more closely simulates our natural environment than the screen-based two-dimensional analogue users have become accustomed to in the digital era. Mine, J. Brooks, and Sequin (1997) notes that the "underlying belief motivating most virtual reality research is that it will lead to more natural and effective human-computer interfaces and that there have been promising results in several key application domains" (see Table 1.1). It is important to note that many of these results are from pre-2000. This is because the first decade of the 21st century is often referred to as the "VR Winter" (Jerald, 2015), where very little mainstream research occurred in the VR domain until its revival in the early 2010s. For this reason much of the research carried out in this period still remains relevant today. Yet despite all these benefits using virtual reality applications, there has been very little investigation into virtual reality in the context of lifelogging, perhaps due to the relatively niche research area of lifelogging.

When we consider the varied and multifaceted nature of lifelog datasets, containing text, images, audio, video and metadata, the concept of an immersive virtual world to store and explore this wealth of digital information seems quite enticing. This would be especially true when hardware advancements enable the

Domain	Example Applications
Experience for the sake of experience	Phobia therapy: Rothbaum et al. (2000), Aesthetics: Davies and Harrison (1996), Entertainment: Pausch, Proffitt, and Williams (1997)
Training and practice of different skills	Surgery: Hunter et al. (1993), Military: Macedonia et al. (1995), Maintenance: Wilson (1997), Wayfinding: Witmer et al. (1996)
Visualisation of unrealised or unseeable objects	Architecture: F. P. Brooks and P. (1987), Fluid Flow: Bryson and Levit (1992), Nano-surfaces: Taylor et al. (1993)
Design	3D Models: Butterworth et al. (1992), Cityscapes: Mapes and Moshell (1995)

Table 1.1: Successful Virtual-World Application Domains from Mine, J. Brooks, and Sequin (1997)

convergence of virtual and augmented reality so that lifelog information could be exposed to us contextually within our environment at any time. While it is tempting to immediately try and create some early version of these advanced interface concepts and start mapping lifelog data to multiple spatial dimensions to try and create an entirely novel method of digesting information, we must first establish a baseline for interacting with lifelog data inside a virtual environment. For the purposes of this research, the HTC Vive was chosen as our target virtual reality platform as it was deemed the most advanced VR headset at the time of purchase and included a set of wireless controllers as standard. However, it is important to note that the design framework which will be introduced in this research is intended to be hardware agnostic and therefore not strictly limited to the HTC

Vive.

1.2 Hypothesis and Research Questions

The research addressed in this work focuses on the confluence of interactive information retrieval within virtual reality and visual access to lifelog datasets based on total capture. We define interactive information retrieval as described by Kelly (2009) where it is characterised as information retrieval in the context of its relationship to humans or human behaviour. To provide further specificity in our hypothesis, we must also define the concept of interactive *lifelog retrieval*, a specialised form of information retrieval which corresponds to the ability of a system or application to retrieve specific digital information to generate cues which promote autobiographical memory. This is one of the five lifelog application benefits defined by Sellen and Whittaker (2010) which are expanded upon in the following chapter of this work, however, it is necessary to explicitly define it in this section to better establish the scope of this research. Finally, throughout this work we refer to conventional systems as distinct from virtual reality systems, where conventional is interpreted as any hardware platform that has become ubiquitous among modern technology users, such as desktops/laptops, tablets or smartphones. However, since the lifelogging community has primarily developed applications for desktops/laptops, this will be the primary conventional analogue we utilise for our evaluative comparison. With all this in mind, we propose the investigation of the following hypothesis:

Hypothesis

> Interactive lifelog retrieval in virtual reality can be as effective as a conventional system in terms of speed and usability while being more immersive, intuitive, and providing increased user satisfaction.

To effectively investigate this hypothesis we have established three research questions which are as follows:

Research Question 1

> What is an appropriate design framework that integrates lifelog principles, personal data, and the wide variety of available technology, to support the development and evaluation of interactive lifelog systems?

This research question was proposed because, to date, many lifelog applications have directed their attention to overcoming technical challenges rather than providing an explicit description of value (Sellen and Whittaker, 2010). While overcoming these technical issues is an important aspect of advancing the lifelogging research area as a whole, to properly develop and evaluate a lifelog interaction system, it is necessary to devise a model or set of criteria which establishes the scope and use case of such a system. This framework would need to account for the huge variance in data that is characteristic of lifelogging, yet also be hardware agnostic in order to be most applicable to the broad selection of competing platforms available to users today. The intent would be to use such a framework as the foundation for the design and evaluation of a lifelog interaction system.

Research Question 2

> Can a virtual reality system support the generation of lifelog retrieval queries as effectively as a conventional system?

For this research question we restrict our scope to the interactions necessary to generate queries for the purposes of lifelog retrieval. The effectiveness of the system in this context corresponds to the speed and ease of use employed in the generation of these queries via the user interface. The interactions necessary to generate lifelog retrieval queries can be quite simple, ranging from selecting options in a list or navigating between specific interfaces, but the methodology behind these interactions can vary hugely in the transition from a conventional system to a virtual reality system. In addition, the impact this may have on lifelog retrieval and the user experience has not yet been properly evaluated.

Research Question 3

> Can a virtual reality system visualise lifelog data in a manner that supports lifelog retrieval as effectively as a conventional system?

For this research question we evaluate the effectiveness of visualising lifelog data in a virtual environment in comparison to a conventional environment. In this context the effectiveness of the visualisation corresponds to its impact on the speed of lifelog retrieval, its ease of use and its immersive quality. It is important to acknowledge that accurately comparing a three-dimensional and two-dimensional analogue is an inherently difficult task due to the fundamental differences between the two mediums and there are obvious benefits to both approaches. However, the aim of this study is to translate the most common visualisation techniques used to support lifelog retrieval from a conventional environment to a virtual environment to determine a baseline for their effectiveness. It is likely that further study could isolate and expand on the most effective visualisation techniques to accomplish this, but that is outside the scope of the baseline being established in this work.

1.3 Research Contribution

Virtual reality is an immersive new medium of digital consumption which is evolving rapidly. Despite its relative infancy compared to other prevalent hardware platforms, its highly immersive quality and wide variety of application, make it a compelling candidate for lifelog exploration. This is even before one considers the eventual convergence of virtual and augmented reality into a mixed reality platform where contextual information could be exposed to us about our lives as we are living them. However, to date, the full potential of virtual reality as a medium to support lifelogging systems remains unexplored. Through this work, we will contribute to this area of research by providing a novel governing design framework to support the development of interactive lifelog applications. This framework is intended to be hardware agnostic and adaptable to the large variety of data which is characteristic of lifelogging. The development of this will also support us in providing our main contribution, namely the first lifelog retrieval system designed for virtual reality. Finally, upon evaluation of this novel lifelog retrieval prototype, we expect to provide a baseline to serve as guiding principles for future lifelog interaction systems designed for virtual reality

We believe these contributions are noteworthy because the research focus of total capture lifelogging has historically leaned toward the overcoming of technical challenges and the enhancement of non-interactive information retrieval methods such as automatic content analysis (Gurrin, Smeaton, and Doherty, 2014). In that respect this has created an excellent test-bed for researchers to collaborate on multimedia exploration prototypes and experiment with rich new data analytics, but the explicit value of total capture lifelogging systems to human users remains poorly defined, and as a result, the value of the systems being developed in this area cannot be fully evaluated. It is our goal to address this discrepancy via the aforementioned contributions and advance the state of modern interactive lifelog applications as a whole.

Chapter 2

Background and Related Work

In this chapter we provide an overview of existing work which is relevant to the research we are conducting in this area. We begin by outlining a brief history of the two main forms of lifelogging which are prevalent in academia today. We follow this with a review of a variety of related systems and the methodology utilised to evaluate them throughout this work. Finally, we explore virtual reality in the context of human-computer interaction and design principles, and how they can be effectively managed to support a modern lifelog interaction system.

2.1 Evaluating a Lifelog System

As was touched upon in our first chapter, despite the lack of an explicit definition of lifelogging to suit everyone in the research community, two distinct approaches to lifelogging have become most prevalent: situation-specific capture and total capture. These two capture philosophies overlap in many fundamental areas and even require much of the same tools, sensors and data processing techniques. It is for this reason that much of the early research into lifelogging would often neglect to specify or even address which capture philosophy they were engaging with. This was most notable in technical papers where the focus of the research was on developing a new type of sensor, storing a greater quality of data, or some other hardware concern. However, it has become increasingly evident that the nature of capture being implemented should infer the nature of the lifelog research being carried out, particularly as we move away from overcoming technical challenges and toward generating actionable insights for end users.

Of the two capture methodologies, probably the easier to summarise is situation-specific capture. This type of capture seeks to record data as completely and automatically as possible but in a very specific domain, often where the research goal is clearly defined. For example, improving an athlete's performance by collecting specific biometric data or recording their diet and calorie intake. The important distinction between situation-specific lifelog capture and any other type of generalised data capture is the emphasis on complete, automatic and continuous collection of the human sensor data being targeted. One of the oldest and most notable examples of situation-specific lifelogging within the research community is the Quantified Self movement (Meyer et al., 2014), founded by Gary Wolf and Kevin Kelly, who have described it as "a collaboration of users and tool makers who share an interest in self knowledge through self-tracking." Despite much of the research utilising situation-specific capture being focused on health and well-being, it is important to remember that any domain focused collection effort designed to be comprehensive, automatic and continuous falls into the situation-specific category of lifelogging. One example could be a system designed to record the audio, documentation or other relevant data during a workplace meeting to preserve ideas or decisions which could be lost afterwards (Bahrainian and Crestani, 2017a; Bahrainian and Crestani, 2017b; Bahrainian and Crestani, 2018).

Total capture based lifelogging, on the other hand, is a more difficult form of lifelogging to address, where the ultimate goal is to record the totality of human experience, collecting as much data as possible as continuously as possible. In this respect, it could be argued that total capture lifelogging is effectively impossible, or at least currently impossible, because the tools necessary to capture the entirety of human experience do not currently exist. Yet despite this, the pursuit of capturing the totality of human experience has had a notably positive impact on research to develop novel techniques for collecting and analysing personal data (Czerwinski et al., 2006; Gurrin, Smeaton, Qiu, et al., 2013; Gurrin, Smeaton, and Doherty, 2014) especially visual data which has become one of the most popular methodologies to support the total capture of life experience. This passive and continuous collection of first-person images captured via wearable cameras has also supported the development of new storage and access mechanisms as well as providing a medium to evaluate new processing techniques for data analysis and enrichment (Gurrin, Schoeffmann, et al., 2019).

However critics such as Sellen and Whittaker (2010) have argued that total capture based lifelogging systems have not proven their value to end users because they are only in use by a small number of people with direct investment in the technology, such as Gorden Bell and his 'MyLifeBits' project (Gemmell, Bell, and Lueder, 2006) and that even users with large collections of data don't actually access the majority of it (Whittaker, Bergman, and Clough, 2010). To address these issues, the critics have suggested that lifelogging systems should strategically target the weaknesses of human memory and that memory itself refers to a complex, multi-faceted set of concepts; thus it is not simply a question of what the lifelog system captures but determining how the lifelog system is going to be used. This includes recognising that no lifelog system can enable a person to explicitly relive a past experience but rather can enable the retrieval of cues which in turn can promote an autobiographical memory. Through exploring the relationship between human psychology and these cues, Sellen and Whittaker (2010) determined five core aspects of human memory which could benefit from a lifelogging system. We can utilise these considerations to develop a set of design principles that would more explicitly convey the value of total capture lifelog systems to end users and also improve the effectiveness of such systems being created within the lifelogging community.

2.1.1 Recollection

The first benefit a lifelog system could have on human memory is to support the simple act of recollecting, or reliving a specific life experience. This involves thinking back in detail on a past experience, sometimes referred to as an episodic memory (Wilding, 2004; Sellen, Fogg, et al., 2007). This type of remembrance enables us to mentally retrace our steps, useful for a host of practical purposes such as the locating of lost property or remembering faces and names by recalling when and where we met someone. Sellen and Whittaker (2010) state that many total capture systems implicitly address recollection, or remembering personal experiences, and that there is well known psychological literature that there is a strong connection between autobiographical memories and visual images (Conway, 1990). This implies user interfaces for such lifelog systems should focus primarily on images in their core design.

2.1.2 Reminiscence

This can be observed as a specialised form of recollection as it is still a form of reliving one's life experiences but for the purpose of emotional or sentimental reasoning as opposed to practicality, for example watching a home movie or flipping through photo albums with friends and family. If system designers want to support reminiscence, visual images are still very important but other factors also become more relevant, such as optimising the sharing of data with others.

2.1.3 Reflection

This potential lifelogging benefit is based on supporting more abstract representations of personal data to facilitate reflection on, and reviewing of, past experiences. For example, examining patterns in one's behaviour over time which could provide useful information about our physical activity or emotional states in different situations. This could then be related to other data about our health or well-being. Reflection might also involve looking at one's past experiences from different angles or broader perspectives where the value is not in reliving past experiences, like with recollection, but in seeing things anew and framing the past differently. Systems seeking to support reflection would need to focus on abstraction, offering flexible and novel methods for viewing personal data in ways that might surprise, provoke or educate its users.

2.1.4 Remembering Intentions

This type of remembrance is different in that it does not relate to retrospective memories, but to prospective memories, as in remembering things you intend to do. For example, remembering to run errands, take medication or show up for an appointment. To support this within a lifelog system, designers need to focus on delivering timely cues in appropriate contexts if they are to provide effective reminders.

2.1.5 Retrieval

As was defined in the previous chapter of this work, this potential benefit of lifelogging corresponds to a system's ability to retrieve specific digital information the user has previously encountered such as images, documents, etc. In this sense, retrieval can include elements of some of the previous lifelogging benefits to human memory. For example, the retrieval of a specific email might support recollection, or the retrieval of a specific image might support reminiscence. Retrieval can often depend on inferential reasoning, such as trying to deduce keywords in a document or thinking about the document's other likely properties. The consideration of information properties need not involve recollection of past experiences at all as long as other ways are available for finding the desired information. Lifelog systems seeking to support retrieval should focus on efficient ways of searching through large heterogeneous collections of data to provide access to metadata that might support more effective filtering, ranking and search. This is particularly relevant for modern lifelog applications which often contain large collections of continuous image streams. Within the academic community, retrieval is the most commonly evaluated aspect of a lifelog system and is also the primary focus of this work.

2.2 Review of Lifelog Systems

In the previous section we established the design principles which form the foundation of our lifelog evaluation methodology. Here we apply those principles to a selection of seminal lifelog applications published within the academic community to evaluate their effectiveness as total capture lifelogging systems. These systems were selected due to their notoriety within the academic community, their technical novelty or their coverage of common lifelogging trends and conventions. It should be acknowledged that there is also a wealth of multimedia and information retrieval systems which have been developed that share many similarities with the prototypes reviewed in this section but are unrelated to the lifelog domain. However, to maintain the focus of this research, we are restricting the scope of application described in this section to lifelog-specific entries.

2.2.1 Microsoft's SenseCam Photo Viewer (Hodges et al., 2006)

The SenseCam is a small wearable camera developed by Microsoft which can automatically capture several thousand images per day. In addition, each camera contains a sensor to detect light levels, an accelerometer to detect motion, a thermometer to detect ambient temperature and a passive infrared sensor to detect the presence of people. To complement the SenseCam's system architecture, Microsoft developed a PC-based application called the SenseCam Photo Viewer (see Figure 2.1) to manage and replay the image sequences and related sensor data captured by the device. This viewer contributed to many of the early experiments into the application of lifelogging as a memory supplement . The interface itself is quite simple, presenting the user with an image slideshow and a familiar video playback control scheme. The images captured by the SenseCam play at a user-defined speed and can be paused, rewound or deleted at any time. When paused, the user has the option of bookmarking the currently observed image frame and annotating it with a label. These labels typically convey the context or activity present in the image (e.g. getting ready, going to work, eating dinner, etc).

The simplicity of the interface is likely a reflection of the early stage of research into the much more visual form of lifelogging that the SenseCam facilitated. There were no established user needs or requirements at the time of development so the captured data is presented linearly and with minimal interaction mechanisms. For example, the data captured by the SenseCam's various other sensors can be viewed

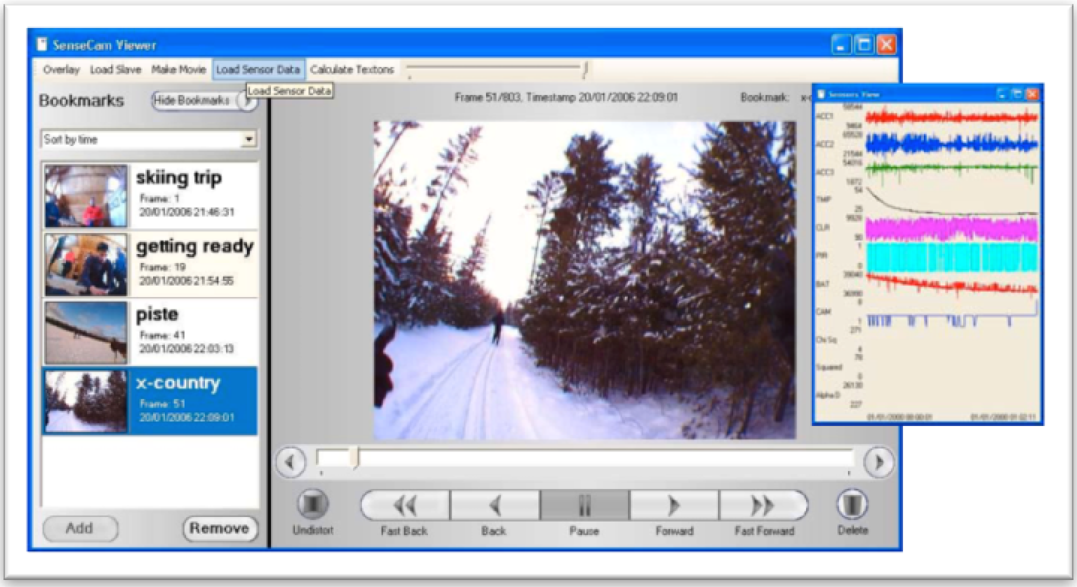


Figure 2.1: Microsoft's SenseCam Photo Viewer Interface (Hodges et al., 2006)

alongside the captured images but the data is visualised with a basic line chart and no other dissemination or analysis is performed. The only image analysis the interface provides is an option to process the entire sequence of photographs and determine which images are most similar. This process effectively splits the sequence into segments of related images which the developers state is to aid navigation of longer image sequences and resembles a rudimentary version of modern event segmentation, where objects recognised between chains of images are used to cluster images into semantic occurrences (e.g. working on a computer, eating a meal, etc.).

It is clear, especially since the prototype is labelled as a photo viewer, that Microsoft's early SenseCam application does not try to consider any specific form of remembrance, such as user reflection or user retrieval, yet still manages to provide cues which can encourage user recollection and, to a lesser extent, user reminiscence.

2.2.2 Microsoft's 'MyLifeBits' Project (Aris et al., 2004; Gemmell, Aris, and Lueder, 2005; Gemmell, Bell, and Lueder, 2006)

In contrast to some of the comparatively simpler lifelogging systems reviewed, Microsoft's MyLifeBits project is directly inspired by Vannevar Bush's Memex (Bush, 1945), seeking to create "a personal database for everything" and therefore attempts to accommodate a significantly larger variance of data. The developers state that "MyLifeBits has at its heart a SQL Server database that can store content and metadata for a variety of types, including contacts, documents, email, events, photos, songs and video". Once the schema was developed and relevant information was added to the database, the project became a quest for useful tools to organise, access, enrich and report about the data. This patchwork approach to the MyLifeBits user interface may have had a negative impact on the application's overall design and user experience and suggests that the primary value of the system is in the database rather than in the expression of its data.

The main section of the MyLifeBits user interface (see Figure 2.2) allows database queries to be viewed as a list with variable sized thumbnails in a timeline and enables refinement or pivoting according to metadata or links. The system also allows creation of text and voice comments; any number of selected items

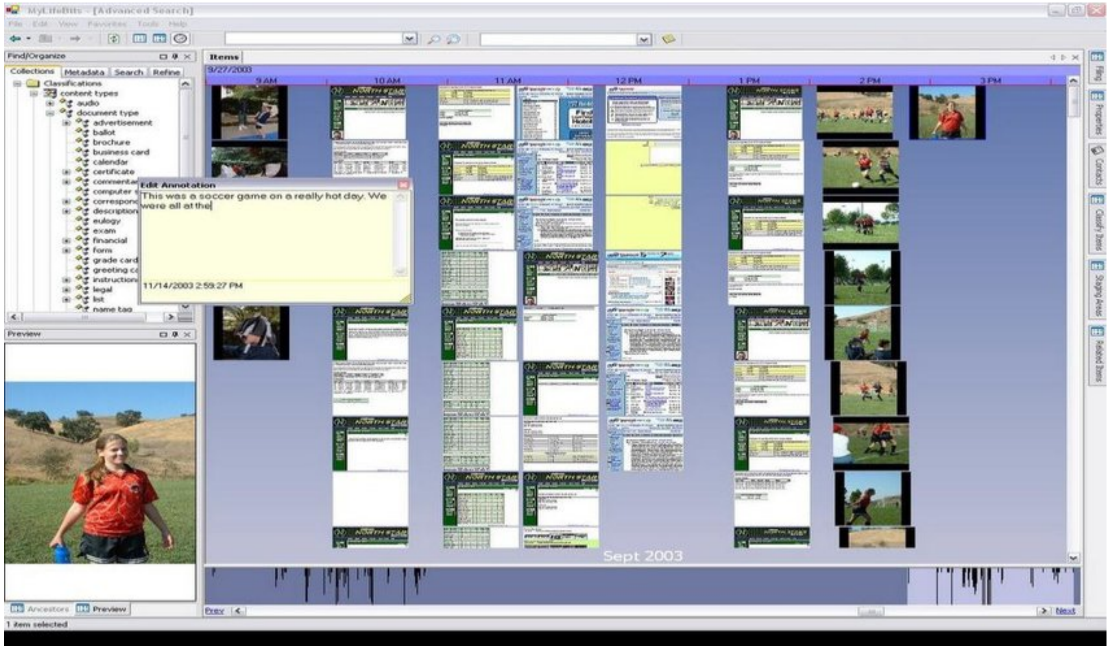


Figure 2.2: Microsoft's 'MyLifeBits' Project Interface (Gemmell, Bell, and Lueder, 2006)

may be commented on using this annotation functionality. Apart from these core methods of interaction, the MyLifeBits system relies on separate tools or plugins to visualise its data. One example the authors explored included a map visualising an individual's various GPS locations and the corresponding location of each photograph that was captured by their SenseCam or other photographic device. This map interface also has the ability to detect individual trips the user has made and play the images contained in that trip in a slideshow.

The Microsoft MyLifeBits system (Gemmell, Bell, and Lueder, 2002), openly using Vannevar Bush's Memex as a blueprint, is a clearly defined step toward capturing the totality of a person's life experiences. While there is notable merit in the technical achievements of such a system, it is also a clear example of a lifelogging application employing a total-capture methodology that can lack a clear description of value for users. While the core interface shell of MyLifeBits includes visual and contextual cues like thumbnails and timelines that can support recollection, the purpose of these interaction methods are poorly defined considering the extent of the dataset the project is attempting to capture. This is further compounded by the implementation of additional interface shells that target and visualise specific parts of the dataset instead of a unified application prototype accommodating the entirety of the dataset with clearly defined use cases and lifelog benefits. It is important to recognise that Microsoft's MyLifeBits has a primary focus on total capture and as such it constantly needs to evolve as newer and better capturing techniques become available. It is likely that because the developer's primary focus is on this aspect of the system, they believe the interface is delivering benefits without being specific about what those benefits actually are.

2.2.3 A SenseCam Visual Diary (Lee et al., 2008)

The SenseCam Visual Diary is an image management system developed by Lee et al. (2008). The developers of this visual diary note that because of the vast amount of images that can be captured each day by Microsoft's SenseCam, 'an important issue arises regarding the mechanisms for the wearer to access the images later'. To address this issue Lee et al. developed their own prototype application which employs a number of content-based image analysis techniques to automatically structure and index the captured lifelog images in a way that is more conducive to searching, browsing and inferring insights. It is made clear by the authors that, at the time, there was no user base present to obtain user needs or requirements and therefore the developers' evaluation approach was to build the scenario and the physical application and then get a small number of early adopters to use the system which would then later be refined.

The SenseCam Visual Diary web interface is designed around the concept of events in a user's daily life (e.g. having lunch, travelling to work, getting ready for bed, etc). These events are automatically detected via the system using context-based sensor analysis in conjunction with content-based image analysis (Doherty, Smeaton, et al., 2007). Once the beginning and end of an event is located, then a 'landmark photo' is extracted and used as a thumbnail representation of the event. In Figure 2.3 we can see an example of an event breakdown for one day and can observe twenty extracted events; the developers note that the amount of events extracted over the user's targeted number of days can be adjusted depending on a user's preferred level of granularity. It was also noted that the interface paradigm was not conducive to search, however, a secondary system was developed to accomplish that task (Doherty, Pauly-Takacs, et al., 2012) which was shown to be more effective.

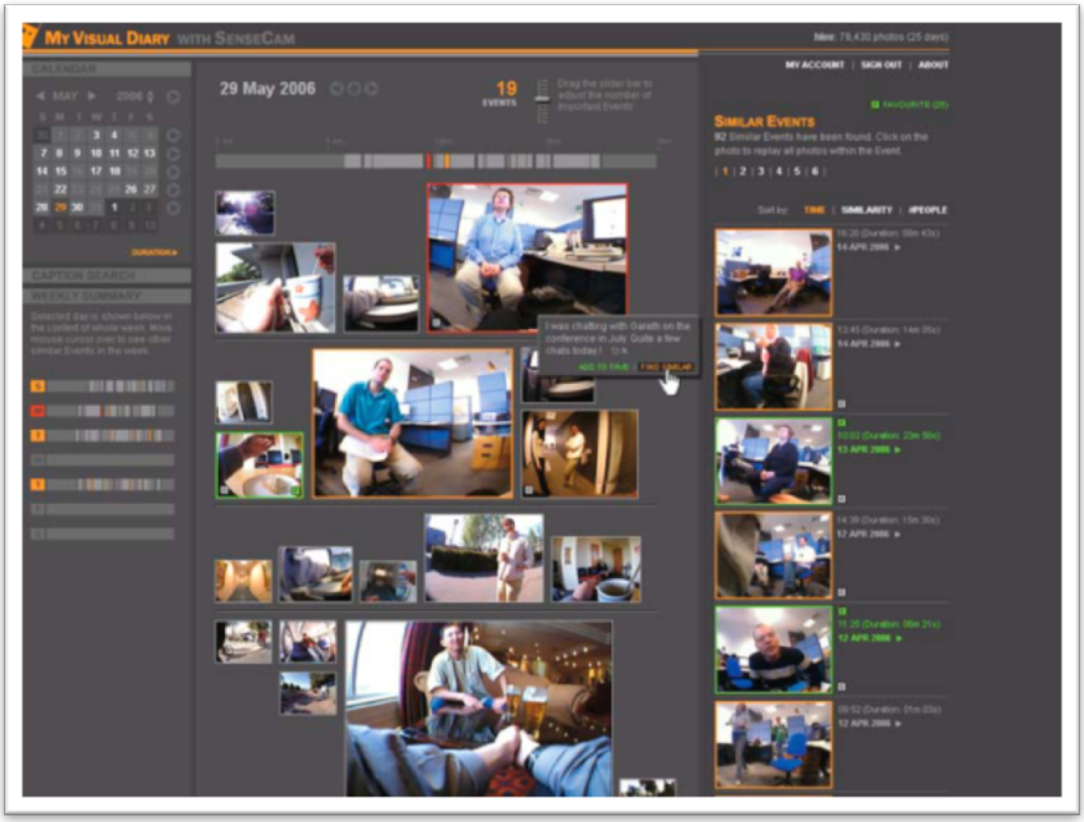


Figure 2.3: SenseCam Visual Diary interface

It is immediately apparent that the landmark photos corresponding to each of the twenty events in the web interface are of varying sizes or resolutions. These sizes correspond to the calculated novelty of each event. The novelty reflects how different an event is from the average events experienced by a user; the concept being that meeting an old friend or giving a presentation is more important than sitting at a desk or standing in an elevator. In this early work novelty was a function of visual uniqueness (colour) and the presence of faces. A user can interact with these events by hovering a cursor over a landmark photo that they are interested in and the photo will be replaced by a slideshow of the images contained within that event. To further contextualise the event in question, a timeline of the selected day is visualised at the top of the interface and reflects what time the current slideshow being played occurred at (highlighted in red) alongside similar looking events that happened throughout the day (highlighted in orange). These similar events can be further examined on the rightmost part of the interface if necessary. Navigation between different day selections is controlled via a calendar in the top-left corner of the screen. Finally, each event can be annotated, edited or deleted from the interface should the user wish to do so.

It is clear that the SenseCam Visual Diary interface touches on many beneficial aspects of lifelogging but without committing to any one benefit in particular. This is likely due to a lack of user requirements and an application primarily targeting early adopters. The most effectively addressed lifelogging benefits of the web interface appear to be recollection and reminiscence. The abundant usage of 'point of view' captured imagery throughout the system's event visualisations complements our autobiographical memory and work as effective cues to recall practical and sentimental memories (Conway, 1990). The developers also touched upon the benefit of reflection with the addition of their 'similar events' concept, featured most prominently on the rightmost part of the interface. This section, especially when used in conjunction with the system's day timeline, has the potential to provide insight into a user's patterns and behaviour. Finally, the web interface's ability to summarise and efficiently browse through thousands of photo documents in a short amount of time, alongside the system's feature to search any event once it has been annotated by a user, correlates well with the retrieval benefit of lifelogging.

2.2.4 Lifelog in a Lean-back Environment (Gurrin, Lee, et al., 2010)

This lifelog interaction prototype presents itself as a personal multimedia browser in a lean-back environment (Gurrin, Lee, et al., 2010). The developers use this terminology to separate the "enjoyment oriented (lean-back) living room environment from the task oriented (lean-forward) environment of the office computer". To reflect this criteria, the primary input device for the prototype interface is an infrared remote control, specifically a Wiimote motion controller by Nintendo. Identical in many ways to a standard television remote, except with an added accelerometer, the Wiimote is generally unsuited to standard interaction mechanisms such as complex menu hierarchies, scroll bars or very detailed iconography. Because of this the developers state that any effective lean-back lifelog browser must tailor its design to minimise user input, engage the user with simple interaction methodologies and to represent complex digital multimedia visually.

Similar to the SenseCam Visual Diary already discussed, this lifelog browser relies on extracted events for its visualisation and analysis. In Figure 2.4 we can see a screenshot of the interface which contains one large primary image at the top and several smaller secondary images at the bottom. The secondary images correspond to events extracted from the targeted day which are presented temporally across the screen, whereas the primary image is an interactive slideshow of the images contained in the currently selected event. Compliant with the restrictions imposed on the interface's design by its developers, the only necessary interaction required by the user is a date selection and then an event selection. The user has the additional option of pausing or adjusting the speed of an event's slideshow by twisting their Wiimote but other than that the prototype only engages the user with low overhead and intuitive interaction methodologies.

Figure 2.4: Lifelog browser in a lean-back-environment

This lean-back lifelog browser was developed in order to do a comparative analysis alongside a more conventional lean-forward interaction scenario. Six users were allocated several information retrieval tasks to perform and the length of time each took was recorded. The results of the experiment suggested that the gesture-based prototype was as effective, sometimes even more so, than a functionally and visually similar lean-forward desktop prototype. The users seemed comfortable with the gesture-based interface which they found easy to learn and more satisfying to use than a point-and-click mouse equivalent. The evaluation methodology implemented by the developers of this interface would suggest its primary benefit is in the retrieval aspect of lifelogging, though there could be other benefits as well which remain untested. The lean-back focus of the interface needed to be evaluated comparatively with a more traditional lean-forward focused interface and it is likely information retrieval tasks were considered the most reliable use case across both types of platform. It is clear that the prototype could also effectively expose recollection and reminiscence cues for memory recall but these are much more difficult to evaluate with test users when they are not the owners of the data being tested on.

2.2.5 Touch-Screen SenseCam Browser for an Ageing Population (Caprani et al., 2010)

As seen in previous SenseCam image browsers reviewed in this paper, this interface relies on event segmentation as a fundamental basis for user context and comprehension. This segmentation is accomplished via an identical image recognition process (comparing each image and establishing primary colour and uniqueness). However, the user requirements for presenting this information are notably different because Caprani et al. (2010) designed this lifelog interface specifically for the ageing population. The developers describe a number of compelling reasons for this research. Given that memory defects are more common and severe in the elderly, the benefit of an effective lifelog tool could have a positive impact on memory recall and quality of life. In addition to this, it is important to note that as older adults become increasingly computer literate, we must consider how the effects of ageing may influence their user interaction and experience. These design considerations are summarised by the developers under three primary processes: sensory, psychomotor and cognitive. The sensory process refers to a user's senses like their visual acuity or difficulty hearing. The psychomotor process refers to a user's depletion in muscle strength or general fatigue as they age. Finally, the cognitive process refers to a user's memory, attention or ability to process complex tasks. All these concepts must be considered in order to effectively develop an interaction tool for the elderly with tangible lifelogging benefits.

The researchers employed a user-centric design approach to the development of their SenseCam browser. This involved an evaluation of two previous SenseCam browsers before any new prototype was developed. Volunteers of the appropriate age were asked to interact with these browsers and were then asked questions about their interactions. This evaluation resulted in several key design methodologies being incorporated into the prototype's final iteration. These considerations included enlargement of image thumbnails, buttons and font size, removal of any scrolling interactions as they were deemed too complex, and also the decision to produce the interface exclusively for touch-screen as it supported novice computer usage. In Figure 2.5 we can see the prototype's primary screen and it is immediately apparent the enlarged image thumbnails are the most prominent feature. These images are representative of the extracted events from an individual's day, similar to previous lifelogs we have seen in this review. The user may navigate between days of data via a calendar and can navigate between events in a day using the previous and next buttons to the left and right sides of the interface. Additional features include the ability to view images individually or in a slideshow and the option to label events.

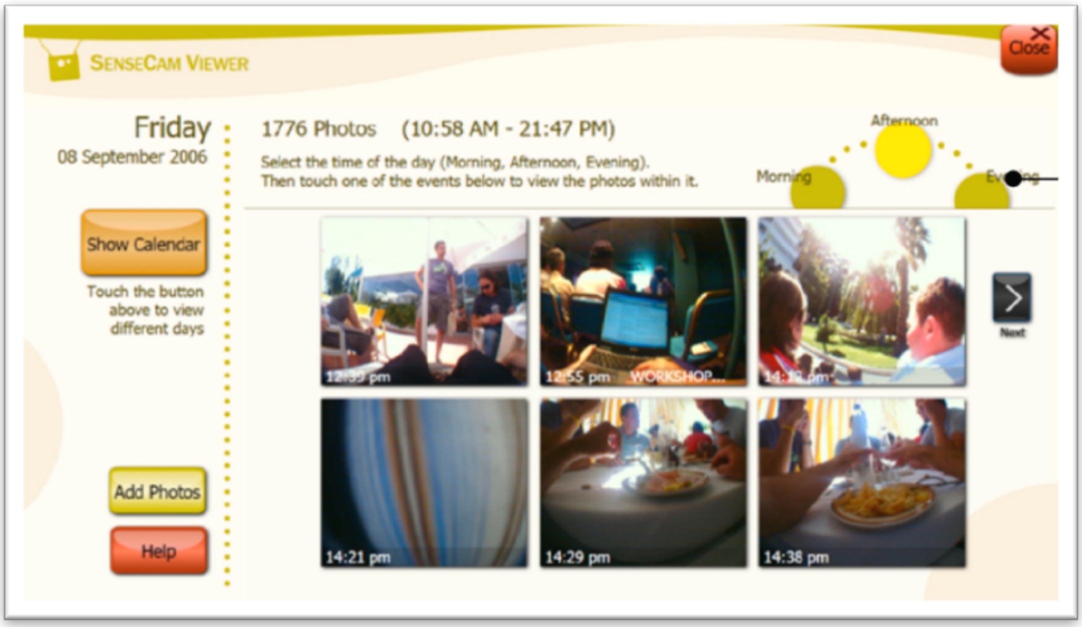


Figure 2.5: Touch screen SenseCam browser for the elderly

Due to the inherent restrictions associated with these target users, it is unsurprising that the lifelog's primary focus is on recollection and reminiscence cues due to their relative simplicity and highly visual nature. It is likely that any attempt to introduce features related to retrieval or remembering intentions would require more complex interactions that would negatively impact any benefit an ageing user might experience using a lifelog interface. However, there is potential to introduce a feature conducive to reflective cues without excessive interaction but this would require the capturing of additional data and the introduction of more automated processes which could detect concepts like physical activity or social context.

2.2.6 Lifelog Interaction on Different Platforms (Yang, Lee, and Gurrin, 2013)

Yang, Lee, and Gurrin (2013) explored visualising lifelog data on three different interaction platforms, specifically mobile, tablet and desktop. The developers cited Shneiderman's information seeking mantra (North, Shneiderman, and Plaisant, 1996) as their inspiration and decided on using an event based approach by grouping sequences of related photos together to reduce visual complexity. This is especially relevant for the mobile platform as user interaction can often occur in a disruptive environment and the limited display size of a smartphone is unsuitable to lengthy exchanges or information overload.

Figure 2.6: Lifelog mobile interface - overview

In Figure 2.6 we can observe how the lifelog overview for the mobile platform is designed to fit the entire display, utilising all available screen space. This overview uses the captured images to summarise a period of time under certain criteria like 'most appeared face' or 'most socially active moment'. The overview screen also contains the number of unique events and faces detected in the currently defined time period. Furthermore, the summary images are colour-coded to indicate the physical activity associated when the capture occurred such as ‘running’ or ‘lying down’. By tapping on a photo of interest, the user is taken to a timeline and shown a list of similar looking photos for further exploration. In Figure 2.7 we can see an example of this showing all the photos which contain new or unique faces. The image’s position on the x-axis indicates its importance score while its position on the y-axis indicates the capture order. At this point the user can shake the phone and the interface will reshuffle with an alternative selection of images.

Figure 2.7: Lifelog mobile interface - timeline

Unlike the mobile platform, the tablet interface is typically utilised in a more leisurely environment and is most often equipped with a larger screen size. This combination of factors allowed the developers to introduce more complex visualisation and interaction methods. For example, in Figure 2.8 we can see that the primary tablet interface is divided in two, where each half corresponds to the two primary screens available on the mobile platform; the overview and the timeline. Because the two interfaces exist on the same screen, the user interaction for this iteration of the lifelog is slightly different. Similar to the mobile version, the timeline presents the user with events that are ranked on each axis by timestamp and an importance metric. However, unlike the mobile version, the overview presents the user only with the images belonging to the event the user selects from the timeline section. This provides an increased level of precision and control for the user exploring their lifelog that would be less effective in a mobile environment. Another feature introduced in the tablet interface due to the increased display size is the presence of a sparkline[1] in the upper-right corner of the screen which indicates the social activity level of the currently selected event.

[1] A sparkline is a very small line chart, typically drawn without axes or coordinates.

Figure 2.8: Lifelog tablet interface

The third and final platform this lifelog interface was adapted for was the desktop (or PC). Compared to the mobile or tablet, the desktop display is significantly larger and is most often used by people for use cases that require a lot of focus and attention. With this in mind the developers explain how they designed an interface containing multivariate content in a symbolic pattern. In Figure 2.9 we see how each day of data is represented by a horizontal segment containing circles and lines. The circles correspond to each detected event in the day and the radius of the circle indicates its significance. Each circle is labelled with a number reflecting the amount of photographs captured during that event. The lines correspond to the user's current physical and social activity levels where the colour of the line indicates physical activity and the height of the line indicates social activity. Another feature introduced in the desktop interface is the ability to drag the timeline play header horizontally to trigger a slideshow of the photos captured around the highlighted event. The user can filter the information onscreen by adjusting options like time, activity or events.

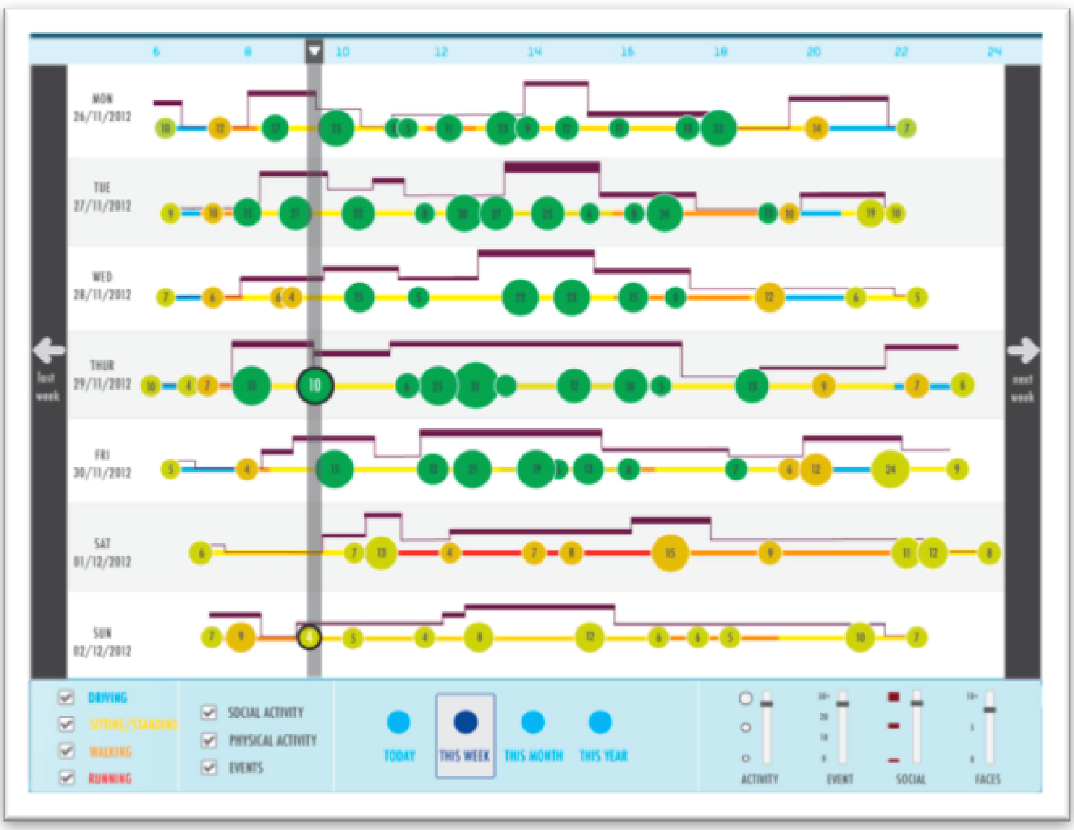


Figure 2.9: Lifelog desktop interface

It is important to note that the three interfaces proposed by Yang, Lee, and

Gurrin (2013) were design prototypes rather than application prototypes. No interface was developed so no evaluation methodology can be cited with regard to the interface's effectiveness. Despite this we can still speculate on how each platform's interface could have addressed the potential benefits of an effective lifelog. For example, the mobile lifelog seemed especially focused on reminiscence, deliberately employing cues that enable users to relive past experiences for emotional and sentimental intentions rather than purely factual ones. This is likely related to the disruptive environment a handheld device is most often used in and the improbability that a user's intent is to remember something factual or important under such conditions. This theme of reminiscence is prevalent throughout the three platforms with elements of recollection only becoming apparent on the desktop platform. The only other prevailing lifelog benefit is that of reflection which is achieved via the timeline feature present in each platform's interface. It enables the user to reframe the lifelog's detected events under a number of criteria that support reflection cues in an insightful manner. However, the effectiveness of these cues significantly improve as we move from the smaller displays to the larger displays, with the desktop platform being superior in terms of features and performance. It is clear there was very little emphasis devoted to other potential lifelog benefits such as retrieval or remembering intentions. This is not unsurprising when the primary unit of measurement employed by the lifelog are events detected via image recognition and no other retrievable document. However, with regard to remembering intentions, we have already established a recurring theme of omission with regard to this lifelogging benefit due to its inherent complexity.

2.2.7 Recent Interactive Lifelog Retrieval Systems (Gurrin, Schoeffmann, et al., 2019)

So far in this section we have utilised a set of design principles to evaluate a selection of lifelog applications in isolation which, while helpful when reviewing the progression of interactive lifelog conventions as a whole, is not useful for evaluating the relative effectiveness of these conventions. This is because without a direct comparison between lifelog systems in the context of a specific use case, it is difficult to determine which application is most successful at a given task. The reason this type of comparative analysis has not occurred to date is likely because these systems are developed by different researchers, often years apart, and operating on vastly different (or proprietary) datasets. However, recent attempts to organise such an evaluation have successfully culminated in the form of the Lifelog Search Challenge (LSC), a real-time interactive competition modelled after the more established Video Browser Showdown (VBS) (Lokoč, Bailer, et al., 2018) which takes places at the MultiMedia Modeling (MMM) conference each year. The LSC invites international researchers to participate in an interactive content retrieval challenge that is performed over a multimodal lifelog datatset, continuously recorded by one lifelogger over a number of weeks.

The first iteration of this competition took place in Yokohama, Japan in 2018 as part of the International Conference on Multimedia Retrieval (ICMR), and consisted of six interactive retrieval systems developed by six different research teams. The format of the competition challenged the participants to retrieve a selection of memorable and interesting events (referred to as topics) which occurred during the time period covered in the dataset. Each round of the competition took place in real-time and the teams were scored based on the speed of retrieval and penalised for incorrect submissions. Another key component of the LSC, which originated in the VBS format, was the requirement for the participation of both expert and novice users. The expert users were typically the system developers themselves, whereas the novice users were recruited from the audience of the conference and were not familiar with any internal details of the competing systems. We will now briefly discuss each participating system.

AAU: liveXplore

The liveXplore system, developed by a team from Alpen-Adria-Universität Klagenfurt (AAU), was a modified version of a pre-existing system built to compete in VBS (Schoeffmann et al., 2018). The primary focus of the liveXplore system was on visual exploration and metadata filtering. Since the original application was developed for processing video scenes, the lifelog image sequences were converted to video using a constant frame-rate. Pre-calculated semantic shot segmentation enabled clustering of similar images to coherent scenes and the creation of the main interface, which presented the user with an adjustable multi-level feature map grouping together similar shots according to machine learning descriptors or handcrafted features. In addition to providing shot-specific similarity search based on these features, liveXplore specifically offered the possibility of exploring individual lifelog day summaries as chronologically ordered galleries as well as videos in an overlay view enriched with metadata information. Finally, in order to search the data according to metadata information the system featured a filter view that allowed users to mix and match temporal, location-based or activity-based and machine learning concept oriented filtering.

DCU: LIFER

The LIFER system (Zhou et al., 2018), developed by a team from Dublin City University (DCU), was a first generation interactive lifelog search engine that aimed to retrieve relevant moments from lifelog data in a fast and efficient manner. For the LSC, it was developed to index only locations, visual concepts, time and activities from the target dataset. This data was converted into feature vectors over every minute and then hierarchically grouped into event nodes. The retrieval was performed by collecting moments that matched the queried criteria and presenting them on screen in a ranked list with associated metadata. Queries were submitted as sets of facets relating to datatset and merged to generate feature vectors for similarity ranking.

UPC-DCU: Interactive Lifelog Browser

The Interactive Lifelog Browser (Alsina, Giró, and Gurrin, 2018), developed by Universitat Politecnica de Catalunya (UPC) in collaboration with Dublin City University (DCU), was a multi-faceted query interface built on a trusted retrieval engine that presented its results via a ranked list. The ranking engine indexed every minute as the retrievable unit using the commonly used TF-IDF ranking methodology. The free text search implemented standard enhancements such as stop word removal and term stemming for the English language. The ranked list from the free-text search was filtered by the other data facets, such as time of day, day of week, or location. The result was a ranked list of filtered moments for presentation to the user. In order to provide additional context to a ranked moment, the preceding two images and the succeeding two images contributed (on a sliding scale) to the overall score of the main image.

VNU-HCM: Semantic Concepts Fusion Retrieval

The Semantic Concepts Fusion Retrieval system, developed by the Vietnam National University - Ho Chi Minh city, was a lifelog retrieval system that integrated recent achievements in computer vision for place and scene attribute analysis, object detection and localisation, and activity detection using image captioning. Independent images were organised into visual shots using sequences of similar images based on visual information then visually similar sequences were linked to a scene using visual retrieval with the bag-of-word framework. The system extracted the location of, as well as the scene attributes of, an image and created a textual caption of it for indexing. To generate queries, users could specify the date and time, time span, or time period (morning, afternoon, etc), scene categories (hotel, restaurant, lobby, etc.) or scene attributes (open area, camping, sunbathing, etc.), visual concepts or activities and finally biometric or computer usage information. The overall interface enabled the user to integrate all of these facets and techniques in one comprehensive system.

SIRET: VIRET

The VIRET system (Lokoč, Souček, and Kovalčik, 2018), developed by the SIRET research team from Charles University in Prague, was another system based on a modified version of a pre-existing system built to compete in VBS. The developers stated their objective for participation in the LSC was to inspect the performance of a purely content-based video retrieval tool for lifelog data. As a result, the tool did not consider any provided lifelog specific modalities, such as locations or heart-rate). The original system relied on sequences of extracted video frames so the transition to the visual lifelog repository was straightforward. Every day from the collection was treated as one 'video' represented by the lifelog images. For each image, automatic annotations were obtained from a retrained GoogleNet (with its own set of 1,390 ImageNet labels). In addition, a colour signature for sketch-based search and deep feature vector from the original GoogleNet were extracted. Based on the automatically extracted features, users could provide three types of query input (keywords, colour sketch and example images) that could be further combined by a late fusion strategy. More specifically, each modality could be used to define a subset of top relevant images and the intersection of all constructed subsets was returned as the result. The final result list was sorted by selected modalities and displayed in the presentation panel. The VIRET system supported two types of result presentation: classical grid with images sorted by relevance and a result list enhanced with nearby temporal context for each top matching frame. Whereas the grid with more images was useful for the exploration phase of the search with frequent query reformulation actions, the temporal context view helped with inspection of promising (visually similar) candidates.

UU-DCU: Virtual Reality Lifelog Explorer

The Virtual Reality Lifelog Explorer (Duane and Huerst, 2018), developed by Dublin City University (DCU) in collaboration with the University of Utrecht (UU), was a variation on the virtual reality lifelog retrieval prototype that is the subject of this thesis. The variation that competed in the LSC was adapted to target the competition dataset and lacked some additional features which were developed in the latter part of this research. As a more in-depth description of this system will occur in subsequent chapters, it will not be described here, except to acknowledge its participation. Furthermore, as the performance of the systems participating in this challenge directly contribute to our hypothesis and research questions, we will reserve discussing it here in order to provide additional context later in this work.

Feature	AAU	DCU	UPC-DCU	UU-DCU	VNUHCM-US	VIRET
Facet Filters	Y	Y	Y	Y	Y	Y
Event/Scene Organisation	Y	Y	N	N	Y	Y
Visual Clustering	Y	N	N	N	Y	N
Novel Ranked List Visualisation	Y	N	Y	Y	N	Y
Enhanced Visual Analytics	Y	N	N	N	Y	Y
Integration of Biometric Data	Y	Y	N	N	Y	N
Non-textual/faceted Querying Mechanism	Y	N	N	N	Y	Y
Based on Existing Video Search Tool	Y	N	N	N	N	Y

Figure 2.10: Summary of the features used by all six systems participating in LSC 2018)

Lifelog Search Challenge 2019

Since the completion of the primary research carried out for this thesis, the second iteration of the LSC has taken place at ICMR 2019 in Ottawa, Canada. Though the virtual reality prototype developed for this work could not participate, and the systems which did could not directly contribute to the research carried out in this work, these systems still represent the latest in the state-of-the-art in this area of research. Only two systems returned to participate in LSC 2019 from the previous year, which were the lifeXplore system (Leibetseder et al., 2019) and the VIRET system (Lokoč, Souček, Čech, et al., 2019). The researchers behind both systems stated that most of their primary improvements from the previous LSC were based on improvements already implemented in preparation for the VBS at MMM 2019, which both systems also compete in. Primary changes made to the systems for LSC 2019 comprised mainly of adjusting the user interfaces to support better filtering of lifelog specific modalities. However, their previous experience in the LSC in 2018 did not appear to contribute a major advantage as the overall winner in 2019 was a newcomer called Vitrivir (Rossetto et al., 2019). This system was another adapted VBS system that was originally developed to retrieve video data. It is clear from the success of these adapted VBS systems in the LSC that there is some overlap in their effectiveness between both competitions, which is likely related to the similar structure of each of the competition's retrieval tasks.

Other newcomers who participated in the LSC 2019 included the Exquisitor system (Khan et al., 2019), described as a highly scalable interactive learning system which uses semantic features extracted from visual content and text to suggest relevant media items to users; the VieLens system (Nguyen et al., 2019), described as an interactive retrieval system enhanced by natural language processing techniques to extend and improve search results in the context of a user's daily activities; the LifeSeeker system (T.-K. Le et al., 2019), described as a faceted search and browsing interface with query expansion to help solve the lexical-gap between novice users and the target dataset's concept metadata; an unnamed system developed by N.-K. Le et al. (2019) described as a smart lifelog retrieval system using habit-based concepts and moment visualisation; an unnamed system developed by Nguyen Van Khan et al. (2019), described as a search engine with two levels of search; an unnamed system developed by Chang et al. (2019), described as an interactive approach to retrieval integrating external textual knowledge.

It is clear from the nature of the systems which have participated in both the LSC 2018 and the LSC 2019 that much of the state-of-the-art lifelog research is focused on non-interactive forms of information retrieval, such as improvements to content analysis, visual clustering or natural language processing, where the user's relationship with the system itself is not a priority. It is our perspective that while such improvements are always valuable, the user's relationship with the system should always be a priority, and we intend to motivate this perspective throughout this work.

2.3 Designing for Virtual Reality

In this section we transition from design principles for lifelogging applications to design principles for virtual reality applications. However, we note that these principles are not explicitly tailored to virtual reality. This is because, at the time of research, virtual reality had not fully matured as an interactive platform so there was no explicit conventions to build upon. Furthermore, there was no prior research on virtual reality applications targeting lifelog data. Without a pre-existing academic foundation, it was our intention to merge well-established lifelogging conventions with high-level design principles to produce an appropriate baseline for study. It should be acknowledged that for future work if more low-level design principles can be utilised in the context of this research, we would advocate their usage over the high-level principles discussed here.

2.3.1 Principles of Interaction

One of the most important aspects when interacting with any system is its discoverability, which relates to the user's ability to learn how the system works and what operations are possible. This is especially important in a virtual reality system because the user is disconnected from the outside world (Jerald, 2015). D. Norman (2013) defined a set of fundamental principles to maximise discoverability in a wide variety of interactions, from operating light switches and door handles, to navigating websites or smartphone applications. The broad application of these principles makes them suitable for assisting us in the development of our novel virtual reality prototype to support lifelog retrieval, and in this section we discuss each of them in that context.

Affordances

An affordance describes the relationship between the capabilities of our user and the properties of our system. For example, user interface elements afford interaction in the same way a virtual controller affords selection. D. Norman (2013) emphasises that an affordance between one object and one user may be different between that same object and another user. This can be best illustrated by imagining a light switch on a wall which affords the ability to control light in a room, but only to those who are tall enough to reach it. Effective interaction design should create appropriate affordances for the desired actions which are possible with the technology being used. This is of course highly relevant for virtual reality where something like a user's height or reach is a very important component. We could design an excellent user interface to support lifelog exploration but then neglect its affordance in a virtual environment thus rendering that interface inaccessible for many users.

Signifiers

A signifier communicates where an action should take place. A good signifier informs a user what is possible before they interact with its corresponding affordance. Signifiers are most often intentional, for example a sign or label, but can also be unintentional, for example the use of a visible trail made by previous people walking through a forest. We might think all signifiers in a virtual environment are intentional because they were placed there by a designer, but this is not necessarily the case. A virtual object might be designed to be picked up for use in a puzzle, but it could be perceived as an object to be picked up and thrown. A good signifer in a virtual reality system should be well communicated and intelligible as the potential for misinterpreting VR signifiers is often greater than in more conventional systems.

Constraints

This refers to interaction constraints within a system, namely limitations on user actions and behaviours. Such constraints can include logical, semantic and cultural limitations to guide actions, but the most relevant to a virtual reality system are physical and mathematical constraints. Properly implementing such limitations can ease the interpretation of a user interface and simplify user interactions, which will improve accuracy, precision and overall efficiency (Bowman, Kruijff, et al., 2004). Such constraints are once again very important within a VR system due to the three-dimensional nature of the environment, which affords a larger degree of freedom to users and thus a larger margin of error. D. Norman (2013) also reminds us that consistency of constraints can be very important as learning can be transferred across tasks. Most users are highly resistant to change, so if a new way of doing something is only slightly more effective, then it is better to be consistent.

Feedback

Feedback is a well-established concept in design; it communicates to a user the results of an action or the status of a task. In a virtual reality system, timely feedback is very important. The simple act of moving one's head requires immediate visual feedback or else the illusion of reality is broken. D. Norman (2013) reminds us, however, that too much feedback can be as ineffective as too little feedback. If a user is overloaded with information, they are likely to disregard it completely. The nature of virtual reality also creates the issue of where appropriate feedback should be positioned in the environment. Displaying all information in front of the user's vision can cause disorientation and information overload, but displaying information elsewhere increases the chance it will be overlooked while navigating the virtual space. Organising and exposing feedback with respect to importance and urgency are key in the development of VR systems.

Mappings

A mapping refers to the relationship between two or more things within a system. The simplest example of this is the relationship between a control, such as a button, and its results. In virtual reality, mappings from hardware to software are especially important. For example, hand-tracked devices (such as for the HTC Vive) work well for pointing interactions, but would be unsuitable for a driving simulator where a steering wheel would be more appropriate. Being aware of such mappings is an important aspect to remember when designing for a VR system.

Compliance

A final principle, described by Jerald (2015), refers to the matching of sensory feedback with a user's input devices both spatially and temporally (Bowman, Kruijff, et al., 2004). For example, if an individual in virtual reality picks up a virtual object, the expected compliance to move that object is to move the hand you used to pick it up. Where the expected compliance is not met, especially within VR, it can cause user confusion and even motion sickness. Yet this does not mean the expected compliance must always be adhered to. For example, if a user attempts to touch a virtual object with their input device, it might be preferable to prevent their input device from passing through the object in the virtual space, even though it does not precisely match the position of the actual input device they are holding. Determining when and where to maintain or break these expected compliances is a difficult challenge designers must overcome when designing a VR system.

2.4 Summary

In this chapter we have introduced a set of design principles for evaluating lifelog applications focused on total capture. Though the primary focus of this work is on the principle of lifelog retrieval, all of the principles are an integral component of the lifelog design framework which will be introduced in the next chapter of this work. We also reviewed a selection of seminal lifelog applications in the context of their adherence to these design principles to better establish a foundation for our own prototype. This included the introduction of five state-of-the-art lifelog retrieval systems which will be further discussed in chapter six. Finally, in preparation for the development of a virtual reality system, we introduced a set of broad interaction principles which will be incorporated into our design methodology. This will be described in our next chapter alongside the introduction of our governing design framework and evaluation methodology.

Chapter 3

Methodology

In the previous two chapters we alluded to a design framework we created to support the development of lifelog applications. To promote the most utility, this framework needed to be agnostic to hardware constraints and compatible with a broad range of multimedia data. In this chapter we introduce this framework and describe how it supported the development of our novel lifelog retrieval system in virtual reality. The development of this lifelog design framework is also one of the contributions of this work and directly assisted in the investigation of our proposed hypothesis and first research question. The latter section of this chapter outlines the techniques and practises utilised throughout this work which served as the foundation of our evaluation methodology.

3.1 Lifelog Design Framework

Our first research question asks what is an appropriate design framework to support the development of interactive lifelog applications. Developing such a framework is inherently difficult due to the highly diverse nature of the data being captured and the wide variety of platforms that lifelog systems can now be developed for. Explicitly tailoring a lifelog application based on the nature of its content and the capabilities of current technology is a suitable way to produce an effective system, but without a broader governing framework, the relationships between these common systems can become difficult to track. This often results in new lifelog systems being developed with little consideration of previous iterations and how they addressed specific lifelogging criteria. Our governing design framework is intended to unify these concepts under one model and support the development of lifelogging systems as a whole. To assist in its development, we borrowed concepts from another established framework in a separate domain; the technological pedagogical and content knowledge (TPACK) framework (Koehler and Mishra, 2009) which was developed by researchers to integrate technology with education. Research in this domain shares many similarities with lifelogging, specifically the emphasis on integrating technology with a diverse range of content under a specific set of principles.

While some critics claim the TPACK framework is only focused on three major areas, namely content, pedagogy, and technology, and therefore does not represent the causative interaction or the direction of the relationship between and among these areas (Graham, 2011), the researchers who developed TPACK argue that their framework does not reflect a specific combination of knowledge related to technology, content and pedagogy but rather represents an understanding that emerges from how these three bodies of knowledge interact. There is evidence suggesting notable benefits from using TPACK's principles in the teaching domain (Schmidt et al., 2009; Burgoyne, Graham, and Sudweeks, 2010) and we can adapt a similar model to convey our own principles with respect to lifelog systems, where instead we are integrating technology, personal data and our set of design principles which we will refer to as lifelogging criteria. In Figure 3.1 we can observe an overview of this framework with some examples highlighting the relationship between its components. The topmost level corresponds to the intersecting domains which reflect the primary relationship model at the core of this framework and represents the three bodies of knowledge we have just discussed: technology, data and lifelogging criteria. The central level of the framework corresponds to the underlying design methodology utilised in the development of the application itself. The bottom level of the framework represents the prototype output which can be interpreted as any single design iteration throughout development. We will now discuss each of the framework components in more detail.

3.1.1 Technology

This part of the domain relationship model represents the access mechanism of a lifelog interaction system. Within the research community this has most often been a desktop or laptop computer paired with a mouse and/or keyboard, though

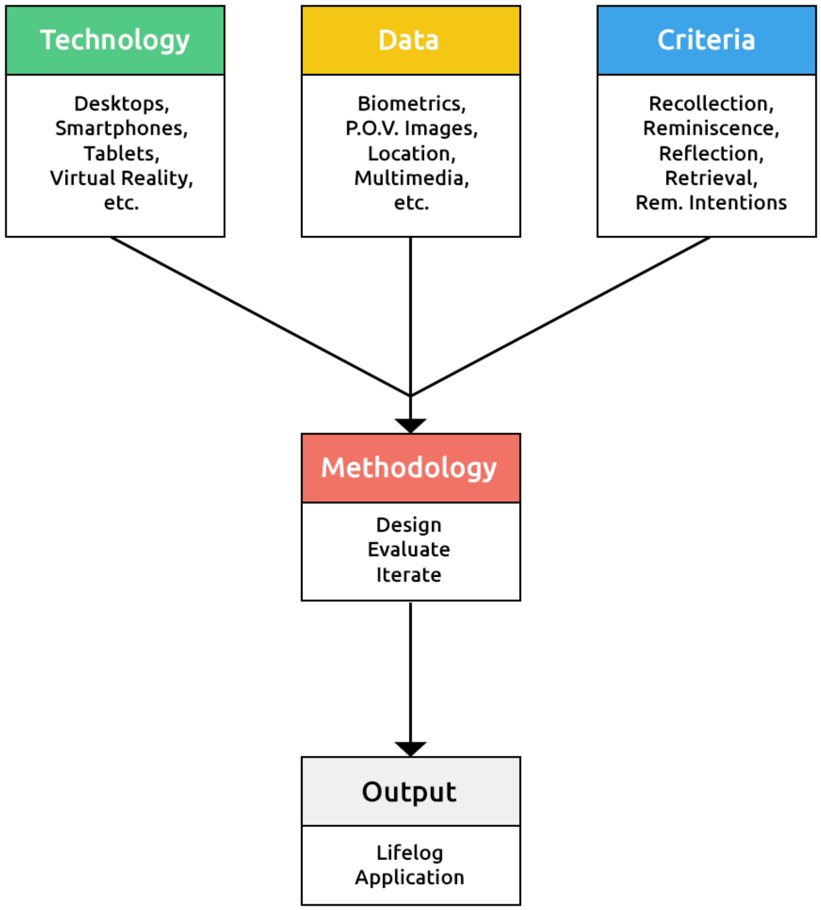


Figure 3.1: Overview of the Lifelog Design Framework

there has been some effort at exploring other access mechanisms (Yang, Lee, and Gurrin, 2013; Qiu, Gurrin, and Smeaton, 2016). Yet despite the popularity of these conventional mediums, it is important to note that any form of technology can be utilised in the development of a lifelog system, where the limitation should not be the technology itself, but rather its relationship to the data and lifelog criteria the developers intend to target. For example, a television and remote control do not efficiently support complex interactions like typing or selecting, but the often larger screen and focus on minimal user input supports longer and more comfortable interaction experiences (Gurrin, Lee, et al., 2010). Therefore an example lifelog system in this domain might focus on visual data like images or infographics and target lifelogging criteria such as reminiscence or reflection.

3.1.2 Data

This part of the domain relationship model represents the personal data being explored by the lifelog interaction system. The nature of lifelog datasets means this data can appear in a wide variety of forms, and because it is captured passively and continuously, there is often a large quantity of it to organise. The data component of our model has seen the most attention within the lifelogging community as it is perceived as the most valuable asset in this area of research (Gurrin, Smeaton, and Doherty, 2014). Yet without proper consideration for the technology being used to access it and the specific lifelogging criteria it should address, the true value of lifelog data is often not realised. For example, a lifelog dataset where the primary focus is on images should target a technology which supports easy browsing of visual material at an appropriate resolution and target lifelog criteria such as recollection and reminiscence due to the connection between images and autobiographical memory (Conway, 1990).

3.1.3 Criteria

This part of the domain relationship model represents the lifelogging benefits which describe the practical value of lifelog applications. These benefits are described in detail in Sellen and Whittaker (2010) and, for the purposes of our research, we refer to these benefits as specific criteria which a lifelog application should target in order to promote its utility. We currently recognise five main criteria: recollection, reminiscence, reflection, retrieval and remembering intentions, but this list is not necessarily exhaustive as new criteria could emerge in the future. We have already discussed the implicit relationship between these criteria and lifelog data in our previous chapter, but their relationship to technology is also important to consider. For example, if a lifelog system's focus is on the retrieval of images or videos, a smartphone might not be an ideal access mechanism due to the small screen size and the difficulty inputting complex search and filter queries on a small touchscreen.

In Figure 3.2 we can see a clearer overview of the domain relationship model which we have just discussed. The intersections visualised in this model are intended to reinforce the concept that it is not about a specific combination of knowledge with respect to technology, data and lifelog criteria, but rather being cognisant of the understanding that emerges from how these three domains interact with each other. The overlap between data and technology highlights the relationship between the capabilities of the access mechanism and the nature of the personal data, for example using lifelog summary techniques on smartphones to prevent visual clutter on a small screen (Qiu, Gurrin, and Smeaton, 2016). The overlap between criteria and data highlights the relationship between the nature of the personal data and the lifelog criteria being targeted, for example using images to promote recollection due to the strong connection between visual images and

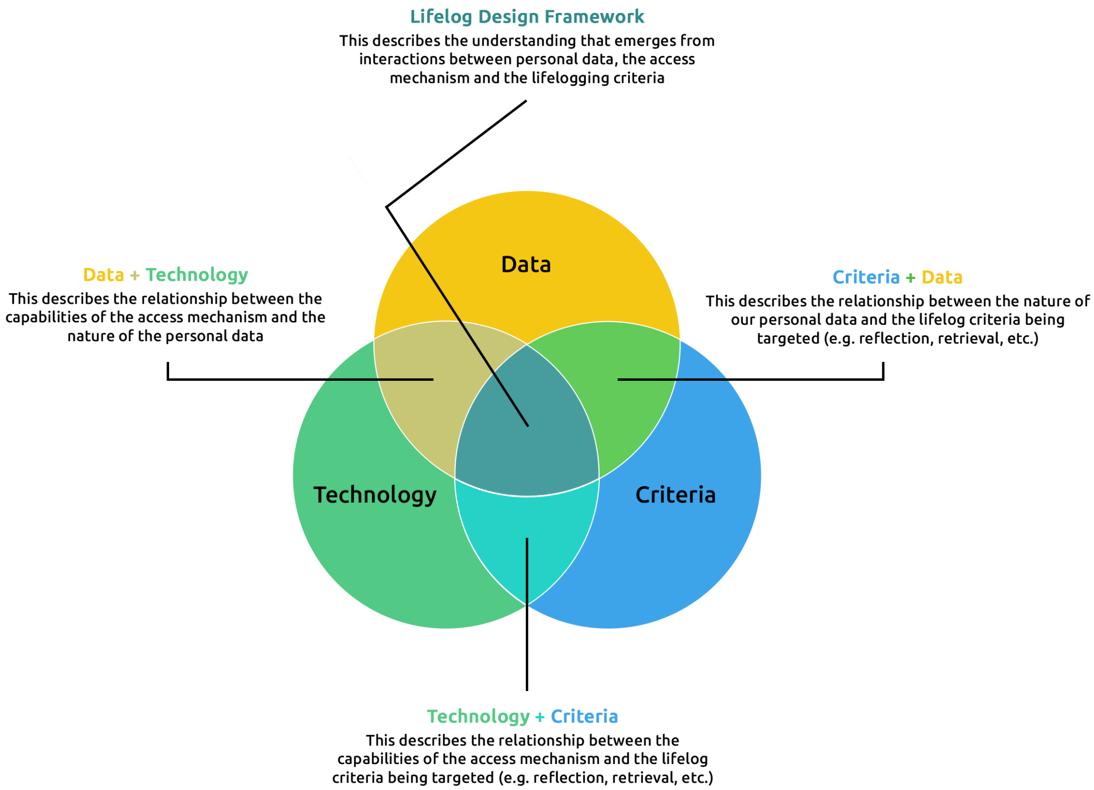


Figure 3.2: Overview of the domain relationship model in the Lifelog Design Framework

autobiographical memory (Conway, 1990). The overlap between technology and criteria highlights the capabilities of the access mechanism and the lifelog criteria being targeted, for example using a television for reminiscence because it benefits from displaying many images in a comfortable environment (Gurrin, Lee, et al., 2010). The focus of the Lifelog Design Framework is to unify these three overlapping domain relationships into a single framework to convey the understanding which emerges when these three bodies of knowledge interact.

3.1.4 Methodology

This section of the framework corresponds to the design and evaluation methodology utilised in the development of the lifelog system. We position this component after our previous section because it can be observed as the next fundamental step after the technology, data and criteria have been identified and their specific relationships understood. Due to the significant amount of design theory that exists, it must be emphasised that this component of the framework does not favour a single design or evaluation methodology but rather serves as a summary of the process to illustrate its relationship to the framework as a whole. The specific design and evaluation methodology utilised should be inferred by the previous section of the framework (the technology, data and criteria) and the context the system is intended for. For example, if a commercial entity were trying to develop a lifelog system for its customers, they might create user profiles, conduct market research or perform rigorous unit testing as part of their design and evaluation process. In contrast, the processes implemented in this work are focused more on researching effective conventions and baseline interactions for lifelog retrieval in a virtual environment, which we will elaborate on in the latter part of this chapter.

3.2 Methodology for a Lifelog Retrieval System in Virtual Reality

Now that we have introduced a framework to govern our actions, we can begin to outline the specific implementation of this framework which was utilised in the development of our lifelog retrieval system in virtual reality, summarised in Figure 3.3.

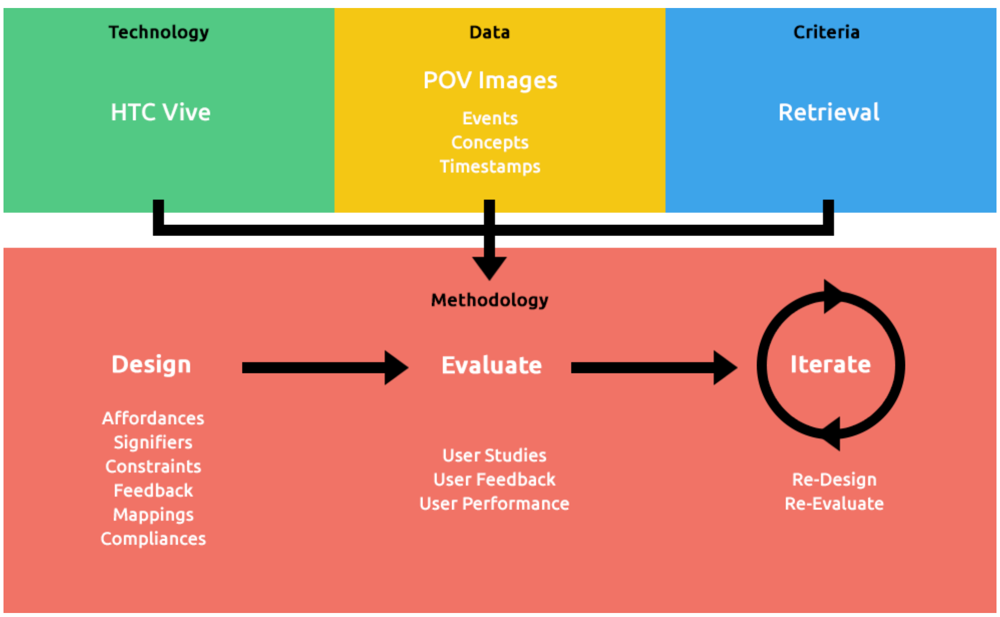


Figure 3.3: Framework implementation for lifelog retrieval system in virtual reality

3.2.1 HTC Vive (Technology)

The technology selected for this work was the HTC Vive because, at the time of researching, this was identified as one of the most sophisticated virtual reality hardware platforms available, and included a pair of wireless input devices (referred to as Vive wands). We are aware that due to the relative infancy of the VR medium as a whole, it is likely that subsequent generations of VR hardware will diverge from their predecessors over time, similar to how modern smartphones have evolved and matured in the years since their initial release. It is for this reason that we refer to our research using the HTC Vive as defining baseline interaction methods for a lifelog system in virtual reality. This is because, as VR hardware matures and potentially converges with AR (augmented reality) technology, we expect these baseline interactions to serve as the foundation for future lifelog systems which will be better tailored to the technology available at that time.

3.2.2 NTCIR-13 Test Collection (Data)

The primary data chosen for this work was the NTCIR-13 test collection (Gurrin, Joho, Hopfgartner, Zhou, Gupta, et al., 2017) released for the lifelog workshop at NTCIR-13[1] in 2016. This dataset consisted of 90 days of continuous images captured via a wearable camera from the perspective of two lifeloggers, where each image was anonymised via face blurring to alleviate data privacy concerns. In total the datatset consisted of 114,547 images which were enriched via computer vision techniques, using feature extraction to provide content analysis and event segmentation. This resulted in each image in the dataset being semantically grouped and automatically labelled with a set of words describing the image's

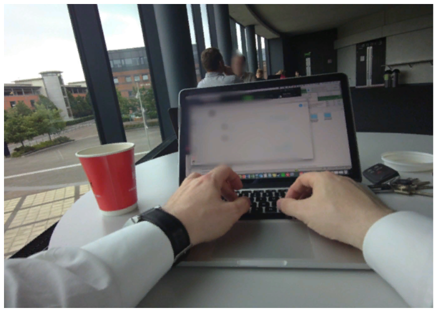
[1]NTCIR-13 Lifelog website: http://ntcir-lifelog.computing.dcu.ie/NTCIR13/

content, referred to as visual concepts, but which we refer to as *lifelog concepts* (see Figure 3.4). The dataset also included a selection of biometric, multimedia and activity data, but this data belonged primarily to only one of the two lifeloggers.

The NTCIR-13 test collection was chosen because it was one of the few lifelogging datasets focused on total capture to be publicly released by the research community. This is likely due to privacy concerns, the expense necessary to analyse a sufficiently large corpus of images, and the relatively niche research area lifelogging exists within. A different version of this test collection targeting different data was also made available for NTCIR-12 (Gurrin, Joho, Hopfgartner, Zhou, and Albatal, 2016) but the content analysis in this collection was not as accurate and was also less comprehensive. To avoid unnecessary complexity navigating between lifeloggers and their corresponding data, we elected to target only one of the two available lifeloggers, who constituted two thirds of the NTCIR-13 datatset, or 60 of the total 90 days. Furthermore, as the biometric, multimedia and activity data was not as comprehensive or consistent, a decision was made to initially target only the images, events, concepts and timestamps. This did not preclude the possibility of introducing the remaining data at a later stage but time constraints and project scope restricted the possibility of including it for this research.

3.2.3 Lifelog Retrieval (Criteria)

The primary criteria selected for this work was lifelog retrieval, or specifically the system's ability to retrieve specific digital information the lifelogger has encoun-



```
concepts: {
        "person",
        "indoor"
        "coffee",
        "laptop",
        "computer"
}
```

Figure 3.4: Example of lifelog concepts for one image

tered. We chose retrieval because it is the most capable of supporting comparative evaluation and is also intrinsically linked to some of the other lifelog criteria and therefore can support broader application. For example, a lifelog system which can efficiently retrieve a specific image can also use that image as a mental cue to support recollection or reminiscence. It is also much more feasible to evaluate a lifelog system specifically targeting retrieval as it does not require the owner of the lifelog dataset to be a test subject. This is because the most commonly used method to evaluate lifelog retrieval is via a known-item search task (Gurrin, Schoeffmann, et al., 2019) which can be performed by any user. In this context, a known-item search task refers to a situation where the user is provided with a specific description of an item, for example an image or event, and are then asked to retrieve it using the lifelog system. In contrast, if researchers want to evaluate a system targeting reminiscence, they would require each test subject to generate their own personal lifelog of sufficient length and then test them all individually. This decision to focus on retrieval was further supported by the fact that a known-item search task was already prepared by the researchers and released alongside the test collection for NTCIR-13. The task was referred to as the Lifelog Semantic Access Task (LSAT) and provided 24 test topics describing instances which occurred in the test collection that a retrieval system could attempt to retrieve.

3.2.4 Design and Evaluation Methodology

The foundation of our design and evaluation methodology was based on user-centred design, where the focus was on producing improved iterations of our prototype based on feedback from users and their performance completing specific tasks. As we mentioned in our previous chapter, at the time of research, there was no prior academic work on virtual reality applications targeting lifelog data. Without a preceding foundation to build upon, it was our intention to merge well-established interaction principles (affordances, signifiers, constraints etc.) by D. Norman (2013) with well-established lifelogging conventions and emerging standards in VR interaction to produce an initial system design which we could evaluate.

The evaluation itself was conducted via a series of user studies which targeted specific aspects of the lifelog retrieval system and compared various interaction modalities. These experiments consisted of 16-18 participants of varying age and technical background but, due to the descriptive nature of the lifelog concepts in our target dataset, required sufficient proficiency in the English language. To ensure this, before participating in an experiment, each user was given a survey to ascertain their language proficiency. We also asked each user their age range, technical background (i.e. general computing skills) and experience with virtual reality. While it would have been preferable to recruit several hundred volunteers of varying experience to produce more robust quantitative data, time constraints and project costs made this unfeasible. As a result, an emphasis on qualitative data became a primary focus, utilising open-ended questions and informal interviews to establish user feedback, which could then be informed by the quantitative data, such as the user's performance.

3.2.5 Experiment Configuration

Experiment	Chapter	Users	Topics	Variables	Scope
User Interaction	4	16	16	4	RQ 1
Data Visualisation	5	18	15	3	RQ 2
Comparative Analysis	6	16	16	2	RQ 1-3

Figure 3.5: Breakdown of the three primary user studies presented in this work

Each user study consisted of a multivariate experiment centred on known-item search tasks, and while the specific configuration of each experiment varied slightly (see Figure 3.5), the overall principle of testing remained the same. The goal was to evaluate a number of variations to our prototype which supported lifelog retrieval in different ways. Each user performed a specific number of the known-item search tasks on each system variant to provide feedback on their experience and determine that variant's effectiveness as a whole. This configuration is illustrated in Figure 3.6 with an example consisting of 4 system variants and 20 known-item search topics.

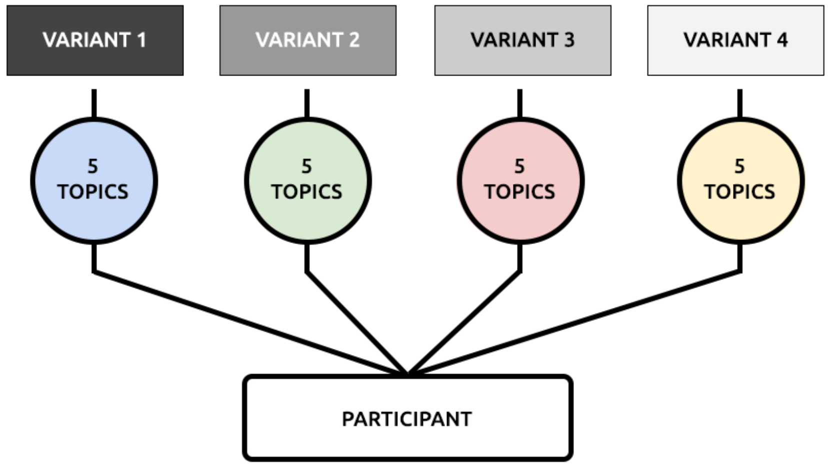


Figure 3.6: Individual participant experiment configuration for 4 variants and 20 topics

To address the bias where topics and system variants performed in the latter part of a user's experiment will benefit from learning, it is necessary to determine a specific configuration in advance for every user. Simply randomising the order of the experiment variables is not a feasible solution because the relatively small number of available participants (16-18) means a bias could still be present. To further complicate matters, the difficulty of each known-item search topic often varies significantly, so it is also necessary to ensure each system variant performs each of the topics an equal number of times across all users, or else there will be a bias toward system variants which encountered easier topics.

SYSTEM VARIANTS

V1	V2	V3	V4
V2	V3	V4	V1
V3	V4	V1	V2
V4	V1	V2	V3

TOPIC GROUPS

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

Figure 3.7: Latin square arrays for four system variants and four topic groups

To address these issues, we utilise a type of array called a Latin square[2] which will enable us to produce a configuration where each experiment variable appears evenly across all users. First, the number of available topics is divided by the number of system variants to create a set of topic groups; in the example case of 20 topics and 4 system variants, this will produce 4 groups containing 5 topics each. We then generate two Latin squares, one array for the system variants and one array for the topic groups (see Figure 3.7).

[2]A Latin square is an n x n array filled with n different symbols, each occurring exactly once in each row and exactly once in each column.

USER	V1	V2	V3	V4
1	A	B	C	D
2	B	C	D	A
3	C	D	A	B
4	D	A	B	C

USER	V3	V4	V1	V2
9	A	B	C	D
10	B	C	D	A
11	C	D	A	B
12	D	A	B	C

USER	V2	V3	V4	V1
5	A	B	C	D
6	B	C	D	A
7	C	D	A	B
8	D	A	B	C

USER	V4	V1	V2	V3
13	A	B	C	D
14	B	C	D	A
15	C	D	A	B
16	D	A	B	C

Figure 3.8: Merged Latin square arrays used to produce final user configuration

To produce the final experiment configuration these two arrays must be unified, which can be achieved by staggering them together to create four new arrays. In Figure 3.8 we can observe how these four configuration groups correspond to the intended number of users in our experiment. Each group has 4 rows, one per user, creating a configuration for 16 users in total. This can be easily adapted for more than 16 users as long as the final number of users remains divisible by 4 to prevent an uneven bias. Each of these rows indicates the order that user will perform the topic groups in and on which system variant the topics in that group will be performed. Finally, within each topic group, the order of the five topics is randomised as it is impractical to try and account for every single potential topic arrangement.

3.2.6 Summary

In this chapter we have introduced a governing framework to support the design and evaluation of lifelog interaction systems. We then described how we implemented this framework as the basis for the development of a lifelog retrieval system in virtual reality, which included a summary of our intended design and evaluation methodology. The success of implementing this framework directly corresponds to our first research question and will be readdressed later in this work. In the remaining chapters of this thesis we will discuss the specific implementation of the methodology we have just described and detail how it impacted the design of our lifelog application prototype. These chapters are divided with respect to the three major user studies which were conducted during the course of our research and directly correspond to our remaining research questions, and subsequently, our proposed hypothesis.

Chapter 4

A Virtual Interface for Lifelog Retrieval

Our second research question focuses on the generation of lifelog retrieval queries in virtual reality and how it might compare to a conventional alternative. However when considering this, we must also consider the design of the user interface which will support these queries and its relationship to the input modality. The most common input devices for conventional lifelog access mechanisms, often some combination of a screen, mouse or keyboard, are typically well understood within the research community. Even lifelog systems utilising touchscreens have established baseline interaction standards that are becoming more ubiquitous. In contrast, virtual reality as a lifelog access mechanism has seen little attention, and there remains no standardised input modality despite the numerous virtual reality platforms that are now available. As a result, it is unproductive to discuss the design of a user interface for virtual reality without also addressing its relationship to the specific input modality utilised and its impact on user interaction. In the following chapter we discuss this relationship and how it informed the design and evaluation of our prototype virtual interface in supporting effective lifelog retrieval. Our objective was to utilise the insights garnered from this work to help inform our design decisions, with the intent that the comparative analysis with a conventional analogue would directly address our second research question and contribute to the validation of our overall hypothesis.

4.1 Designing a Virtual User Interface

We define a virtual user interface as a type of 3D user interface characterised by the use of spatial input in a three-dimensional virtual environment (Bowman, McMahan, and Ragan, 2012). As we established in our previous chapter, the target virtual reality platform for our lifelog retrieval prototype is the HTC Vive. The Vive comes with two wireless input devices, also known as Vive wands, which contain 24 infrared sensors that allow them to be tracked in real-time. As long as the controllers remain synchronised with the tracking system, this enables a direct mapping between the physical controllers and their digital counterparts within the virtual environment. In Figure 4.1 we can observe that each device contains five input methods, including a trigger and multipurpose touchpad, and provides haptic feedback via controller vibrations.

If we recall the design principles by D. Norman (2013) we discussed in chapter

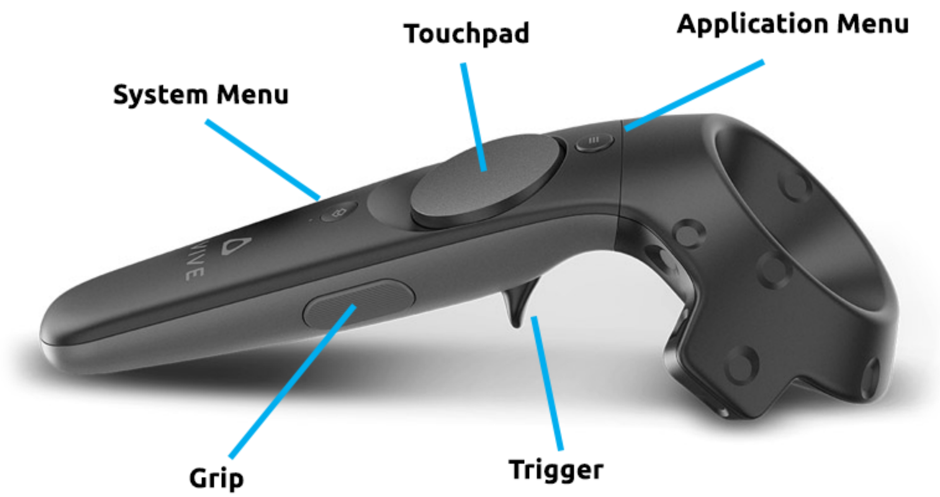


Figure 4.1: Inputs on a HTC Vive controller

2, we know that we can inform the design of our preliminary user interface by considering the affordances provided by the Vive controller. For example, its direct physical mapping to a typical user's hand affords accurate spatial compliance which in turn supports many common interactions such as picking up objects, throwing objects, pointing at objects, etc. In contrast, the presence of only one pressure-sensitive finger input (the trigger) means there is poor interaction fidelity when it comes to other activities such as gripping or typing. When we refer to interaction fidelity, we are referring to the degree to which physical actions used for virtual tasks correspond to the physical actions used in the equivalent real-world tasks (Bowman, McMahan, and Ragan, 2012). Interactions that have high fidelity are realistic interactions such as swinging a virtual bat similar to how you would in the real world, whereas if this interaction was accomplished simply by pressing a button on the controller, this would be described as a low fidelity interaction. High fidelity interactions can afford a strong sense of presence to users where their actions feel familiar and intuitive. This concept is referred to as proprioception and corresponds to a person's sense of the position and orientation of their body and its various parts (Boff, Kaufman, and Thomas, 1986). It is important to note that low fidelity interactions can become just as intuitive over time, whilst also reducing fatigue, promoting wider usability with less physical demand on the user. There is no ideal level of fidelity when it comes to virtual reality systems; it should instead be focused on application goals, and often a combination of high and low fidelity interactions is most appropriate (Jerald, 2015).

When we consider these insights, we begin to understand that while it may be tempting to design a virtual user interface for lifelog retrieval with familiar objects like levers and sliders which afford high fidelity interactions that are comfortable for users, these may not be the most efficient modes of interaction. This type of virtual reality interface would be very appropriate for a driving or flight simulator, but for our lifelog retrieval prototype it would only serve to add unnecessary complexity. D. Norman (2013) has reminded us that consistency of constraints is very important, as learning can be transferred across tasks, so in this respect we can begin to consider what users are already familiar with, and use that as a template for our initial design. It may seem counter-intuitive to suggest a two-dimensional user interface when virtual reality affords access to three spatial dimensions, but operating primarily within two dimensions has proven widely effective in conventional digital mediums to date. Grouping user interface elements together and constrained to familiar dimensions can support more efficient interactions that are still very familiar to users. This is not to suggest that these constraints must be strictly adhered to, or that other spatial constraints could not be even more efficient, but there is sufficient evidence to support what we have just described as an initial design strategy.

4.2 Reference Frames for Virtual Interfaces

We have established that our initial prototype should utilise a relatively familiar style of user interface, constrained primarily to two dimensions within the virtual environment, but we have not considered how exactly the user should interact with this interface and where it exists within that environment. Virtual reality affords us the ability to position elements at any height, distance, angle or scale in relation to our user, and a change in even one of these can have a notable impact on the user experience. This is further complicated when we consider that the lifelog data retrieved by our prototype is also visualised in this same space and should not be obscured by any obtrusive interface elements that are simultaneously available to the user.

To address these issues, we divided our virtual environment into semantic reference frames which correspond to different user interactions. Jerald (2015) defines some of the major reference frames in virtual reality as the hand, the head, the torso, the virtual world and the real world. Apart from the real world, which is a reference frame which must be implicitly considered for every interaction, each of these reference frames must be evaluated in relation to how they support user interactions with respect to our lifelog retrieval prototype. The virtual world reference frame matches the layout of the virtual environment and includes geographic directions and global distances independent of our user. The torso reference frame is defined by the body's spinal axis and the forward direction perpendicular to the torso. The hand reference frames are defined by the position and orientation of the user's hands. The head reference frame is based on the point between the two eyes in the direction perpendicular to the forehead. Finally, the real world reference frame is defined by the physical space our user exists within while operating in virtual reality.

There is no preferred convention when it comes to favouring one reference frame over another in virtual reality as each application has different requirements. However, interactions performed relative to the body's reference frame, also referred to as egocentric interactions, are known to be highly proprioceptive and therefore more intuitive to users. Mine, J. Brooks, and Sequin (1997) describe three primary forms of body-relative interaction. First, there is direct manipulation, which corresponds to methods of using body sense to help control manipulation. Second, there is physical mnemonics, which corresponds to methods of storing or recalling information relative to the body. Finally, there is gestural actions, which corresponds to methods of using body-relative actions to issue commands. Using these egocentric forms of interaction in conjunction with appropriate reference frames, we determined a selection of potential lifelog interaction methodologies which we intended to evaluate. As there was no precedent for lifelog retrieval in virtual reality, the overall goal was to assess the viability of direct, indirect and semi-direct interactions (Hutchins, Hollan, and D. A. Norman, 1985) in various virtual reference frames to help inform our subsequent design decisions.

4.2.1 Distance-Based Interaction

The first interaction methodology exists primarily within the virtual world reference frame and is characterised by interactions made by the user from a distance. This style of interaction is reminiscent of the relationship between a television and remote control where the user points and clicks at what they want to interact with. Despite the user interface existing in the virtual world reference frame, this is still an egocentric style of interaction because the user directly manipulates the interface using their hand via the wireless controller. This, paired with the experience many users have operating televisions or other pointing devices, means this interaction method imparts a strong sense of proprioception and intuitiveness. We have developed two variants within this distance-based interaction methodology to further explore its impact supporting lifelog retrieval.

Billboard

We refer to the first of the distance-based interaction modes as the *billboard* variant, as the scale and relative positioning resemble a large billboard. In Figure 4.2

we can see how in this variant the virtual user interface is arranged in a vertical orientation and positioned at a distance in front of the user. To compensate for this distance, the scale of the interface is notably increased to ensure the text present on the interface remains clear and legible. In addition, this increase in scale also ensures that interface elements remain large enough to be easily targeted by the user. To further accommodate selection of these elements, a glowing beam is automatically rendered (see 4.9) when the user points either controller at any part of the user interface. These beams emerge from the top of the wireless controllers and will highlight any interactive element within the interface that they connect with.

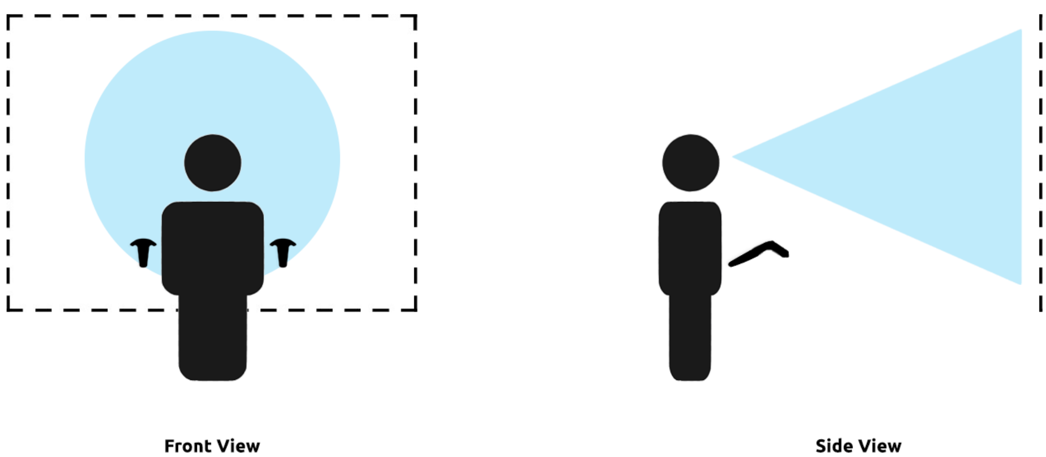


Figure 4.2: Billboard variant of the distance-based interaction methodology

Floorboard

The second variant of the distance-based interaction modes is very similar to the previous variant where the scale and positioning are identical, but the interface's orientation with respect to the user is now horizontal instead of vertical. In Figure 4.3 we can also see how instead of positioning the user interface in relation to the user's head, we position it in relation to the user's feet, which is why it is referred to as the *floorboard* variant. The selection of elements within the interface remains identical however, where the user simply points at an element in the user interface using either controller and a beam will signal to the interface to highlight that element.

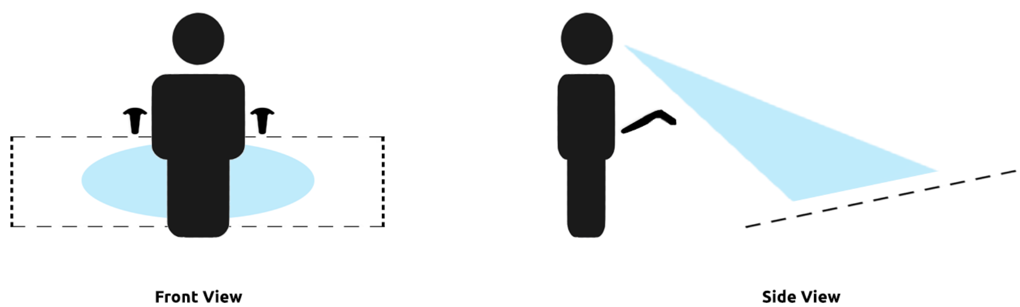


Figure 4.3: Floorboard variant of the distance-based interaction methodology

The *billboard* and *floorboard* variants of the distance-based interaction methodology were chosen deliberately to investigate their impact on user performance and experience. It was initially planned to provide users with the ability to relocate the virtual user interface within the environment at any time, but instead we took the opportunity to try and establish how user's felt about exposing user interface elements in specific locations. The goal was to increase our knowledge with respect to user preferences regarding the visibility of virtual information but also the potential impact on neck and arm strain.

4.2.2 Contact-Based Interaction

The second interaction methodology exists primarily within the torso reference frame and is characterised by the requirement for the user to directly contact the virtual user interface with their wireless controllers. This style of interaction is reminiscent of a touchscreen, where the user identifies what interface element they are interested in and simply reaches out and touches it. Similar to our previous methodology, this was deliberately designed as an egocentric style of interaction to promote a strong sense of proprioception and intuitiveness, and have developed two variants within this contact-based interaction methodology to further explore its impact supporting lifelog retrieval.

Dashboard

We refer to the first of our contact-based interaction modes as the *dashboard* variant, as its scale and relative positioning within arm's reach of the user is reminiscent of the dashboard on a car. By default the virtual interface is positioned based on the user's height, but the necessity for the user to make direct contact with it means we included the ability to adjust the angle and positioning of the interface at any time to accommodate varying arm length and reduce potential discomfort.

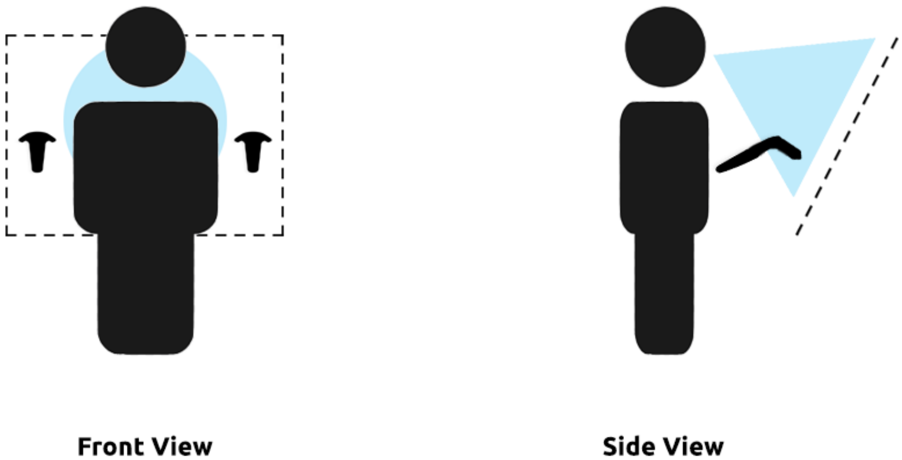


Figure 4.4: Dashboard variant of the contact-based interaction methodology

Clipboard

The second of the contact-based interaction modes is referred to as the *clipboard* variant, as the virtual user interface is directly attached to one of the user's wireless controllers and is interacted with via the opposing wireless controller, reminiscent of a person using a clipboard and writing implement. To accommodate hand preference, the user can select which input device the virtual interface is attached to and which input device is used for interaction.

The *dashboard* and *clipboard* variants of the contact-based interaction methodology represent a more tactile approach to interacting with user interfaces in a virtual environment. It was important to establish users' feedback on the experience of controlling the exact whereabouts of the user interface at all times. In the physical world we are quite familiar with egocentric tasks such as holding

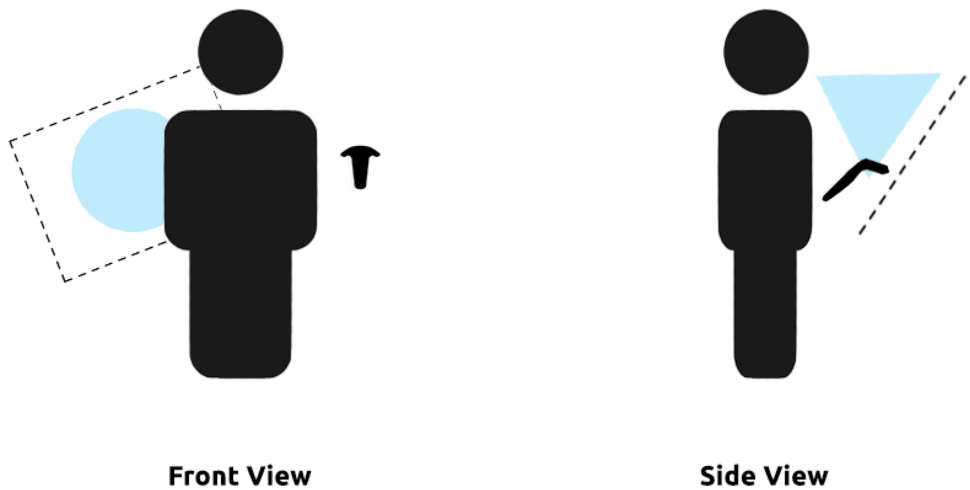


Figure 4.5: Clipboard variant of the contact-based interaction methodology

clipboards or electronic tablets, but this does not mean it is a preferred method of interaction within a virtual setting. Our goal was to increase our knowledge with respect to users' more direct control of the interface elements and its potential impact on efficiency.

4.3 Virtual User Interface Design

As we have already established in section 4.1, the design of the virtual user interface for our lifelog retrieval system was constrained primarily to two dimensions. The intention was to develop an interface which could be easily adapted to the various interaction methodologies we wanted to evaluate. If we recall from the previous chapter, the NTCIR-13 dataset being targeted by our lifelog retrieval system contained images captured via a wearable camera and annotated with events, visual concepts and timestamps. To retrieve relevant images matching the topics of our known-item search tasks, the user needs to generate faceted search queries based on concepts and, optionally, a specified range of time.

4.3.1 Concept Selection

Figure 4.6: Concept selection

In Figure 4.6 we can observe the user interface within our virtual setting. The environment is designed to be featureless and without edges to focus user attention on the interface and the lifelog data being retrieved. The user can navigate between the two sections of the interface, one for concept selection and one for time selection, by selecting the corresponding buttons at the top of the main menu. It is important to note that we refer to lifelog concepts as tags within our lifelog system as many users are unfamiliar with the academic terminology. There was a total of 638 unique concepts contained in the lifelog dataset which the user could use to generate their retrieval queries. These concepts were organised and separated alphabetically and could be viewed by selecting the relevant letter at the top of the concept selection menu. This can be observed in Figure 4.6 where a user is in the process of selecting concepts beginning with the letter A. When concepts are selected, they are highlighted in white to indicate their selection and are added to an active list at the bottom of the interface. The user can remove an active concept by selecting its corresponding element a second time, either from the active list or its original position.

4.3.2 Time Selection

If the user navigates to the time selection menu, they are presented with the 7 days of the week and the 24 hours of the day. In Figure 4.7 we can observe a user who has selected Monday to Friday and 09:00 to 18:00, which will limit any retrieved images to times occurring within these days and hours. If no days are selected, then no limitation will occur and all days will be retrieved by the system. Similarly, if no hours are selected, then all hours will be retrieved by the system. This method of generating the temporal facet of the lifelog retrieval query was intended to be simple and efficient. An initial design utilised a selection method resembling a calendar, but it was quickly determined this added unneces-

Figure 4.7: Time selection

sary complexity to the system as the days contained in the dataset where spread across several months and indicating the specific dates within a calendar view was cumbersome to navigate.

4.3.3 Selection Mechanisms

We have established that both the distance and contact interaction methodologies engage with the user interface in fundamentally different ways. This required subtle changes in the user interface design to accommodate both modes of interaction. For example, if we recall the shape of the Vive controller in Figure 4.1 we can observe that the broad end of the device does not afford precise selection when attempting to make direct contact with virtual elements in the environment. To alleviate this issue when interacting with the system using the contact-based interaction methodology, we modified the virtual representations of the Vive controllers with an additional appendage resembling a drumstick (see Figure 4.8) which supports much more precise connections with virtual elements. To prevent unintended interactions, the user must hold down the trigger on their respective controller to enable selection with that device. This is indicated by the colour of the drumstick appendage which will change from black to white. Upon successful selection of an interface element, the user receives haptic feedback via a short controller vibration, mimicking the sensation of physically interacting with an object in the real world.

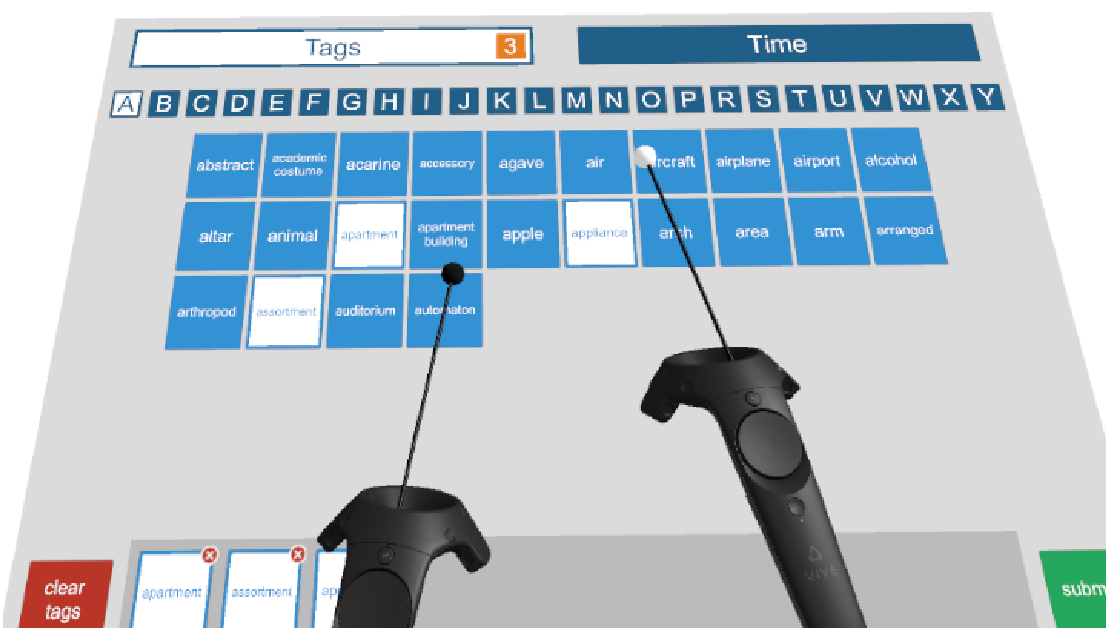


Figure 4.8: Drumstick appendage designed to support contact-based interactions

In contrast to contact-based interactions, the distance-based interaction methodology requires no additional appendage. In Figure 4.9 we can observe that a user simply has to point one of their controllers at the virtual user interface and a glowing, blue beam will appear. This beam indicates precisely what the user is pointing at and highlights the relative element accordingly. To interact with the element being pointed at, the user simply presses the trigger and it will signal a valid selection.

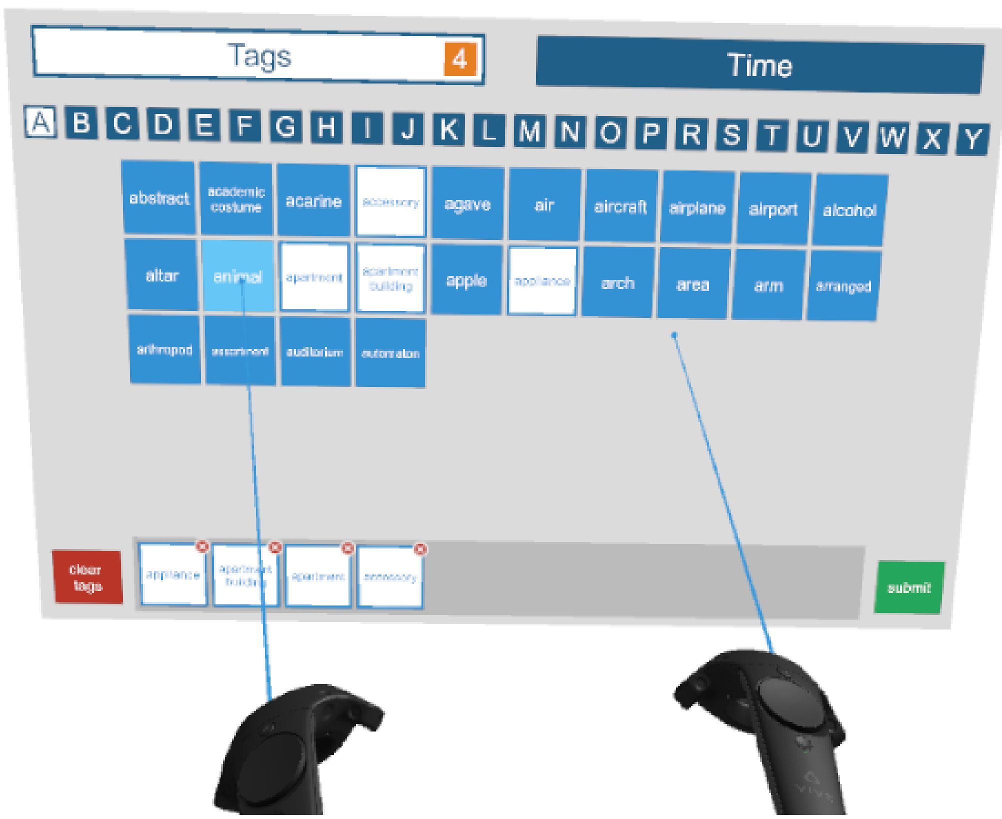


Figure 4.9: Glowing beam designed to support distance-based interactions

4.4 Data Visualisation

The content of this chapter has deliberately focused on user interaction and the design of the virtual user interface for our lifelog retrieval prototype. It is our intention to expand on the visualisation of the lifelog data retrieved by the system in the following chapter of this work. However, to enable us to effectively evaluate the virtual user interface and its corresponding interaction methodologies, it is necessary to implement some form of data visualisation so that users are actually able to complete a retrieval task. It should be acknowledged that more effective data visualisation is likely to have a significant impact on the quality of lifelog retrieval, therefore any evaluation of the virtual user interface at this time will not accurately reflect the effectiveness of lifelog retrieval in virtual reality as a whole.

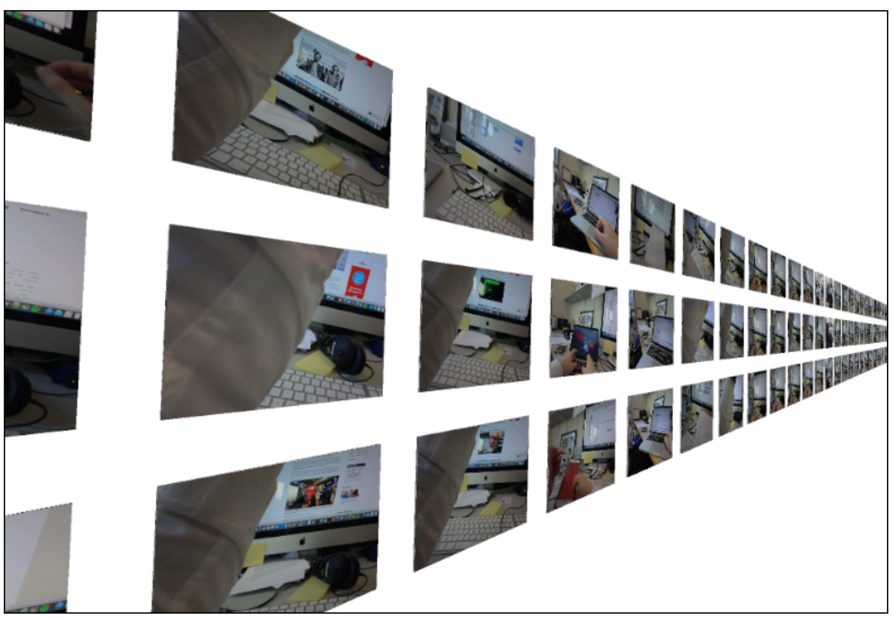


Figure 4.10: Images retrieved by the system visualised in the virtual environment

Upon submitting a query to the system, the virtual user interface fades from view and the user is presented with the query results. This decision to remove the interface while visualising the lifelog data was made to focus the user's attention immediately on the retrieved data and to prevent certain variants of the virtual interface from obscuring the data more than others, thereby negatively impacting the performance of that variant during evaluation. The image results retrieved by the system are displayed in a standard grid formation perpendicular to where the user is facing and positioned based on the user's height. The design of this formation was based on the size of each lifelog image, which was in turn based on guaranteeing image clarity within a reasonable distance. The images are ranked from left to right based on their concept score, which is a simple calculation of how many concepts the user queried are present on an individual image. If multiple images contained the same number of submitted concepts, they are ranked temporally from earliest to latest. In Figure 4.10 we can see a user observing a selection of results at an angle which trails off into the distance. We refer to this as the *memory wall* and it can be scrolled left and right by squeezing the grips on either controller and performing a throwing gesture in the direction the user wishes the wall to move. The speed of this gesture impacts the speed at which the memory wall will scroll and is reminiscent of the momentum scrolling interaction which exists on all modern touchscreens.

The goal of this data visualisation was to implement a type of virtual museum where the user could examine the images as if they were hung on a wall. If a user wanted to get a closer look at a particular image, they would not need to locate a

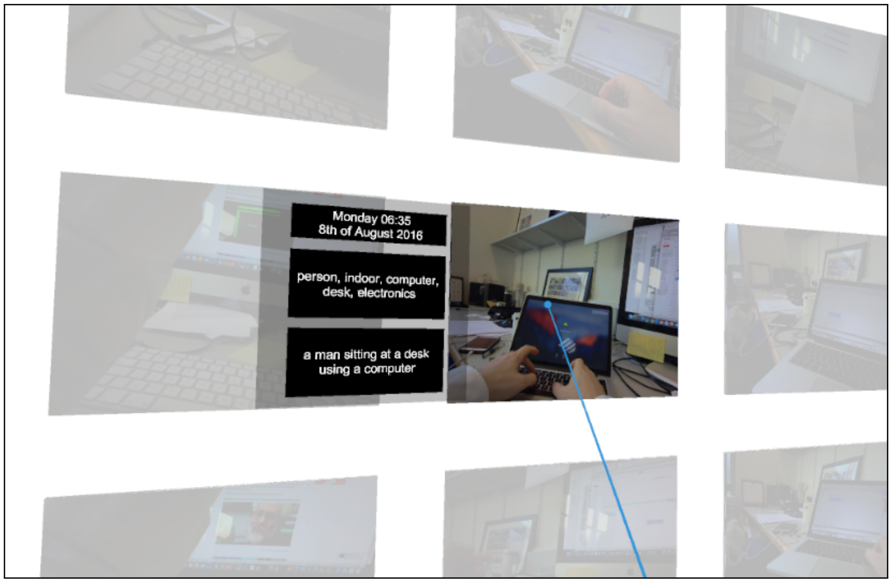


Figure 4.11: Additional metadata can be exposed by targeting a specific image

button or discover a feature, they simply had to step forward and literally take a closer look. However, to provide additional context for each image, we determined it would be necessary for the user to somehow expose the image's metadata (e.g. its timestamp, labelled concepts, etc.). Initially this information was displayed alongside each image by default, but this had a very negative impact on the application's frame-rate and loading time. To address this, any relevant metadata was concealed until the user targeted a specific image they were interested in. This was accomplished by activating a beam by holding down the controller's trigger and pointing it at the relevant image (see Figure 4.11). The user can clear the memory wall at any time and return the virtual user interface by pressing the application menu button on their controller.

4.5 User Interaction Study

After we established four potential user interaction approaches to support lifelog retrieval in virtual reality (billboard, floorboard, dashboard and clipboard), it was necessary to evaluate their effectiveness. As we outlined in our previous chapter, this was achieved via a user study involving known-item search tasks. For this experiment we recruited 16 volunteer participants and used 16 topics provided by the NTCIR-13 lifelog semantic access task (LSAT). We did not use all 24 topics as some of them were specific to the second lifelogger in the dataset which our prototype was not targeting. Following the methodology described in our previous chapter, the total number of topics was divided evenly by the total number of variants to produce 4 topic groups containing 4 topics each. These groups are shown in Tables 4.1, 4.2, 4.3 and 4.4 where we can observe the topic ID, title and accompanying descriptions which were provided to users before beginning each retrieval task.

ID	Title	Description
1	Eating Lunch	Find a moment when the lifelogger was eating lunch
5	Graveyard	Find a moment when the lifelogger visited a graveyard
9	Working Late	Find a moment when the lifelogger worked at home late at night
17	Television Recording	Find a moment when the lifelogger was being recorded for a television show

Table 4.1: User Interaction Study - Topic Group A

ID	Title	Description
2	Gardening	Find a moment when the lifelogger was gardening at their home
6	Graveyard	Find a moment when the lifelogger visited a graveyard
10	On the Computer	Find a moment when the lifelogger was working on the computer at their office desk
14	Photo of the Sea	Find a moment in which the lifelogger was walking by the sea

Table 4.2: User Interaction Study - Topic Group B

ID	Title	Description
3	Coffee	Find a moment when the lifelogger was drinking coffee in a cafe
7	Presenting or Lecturing	Find a moment when the lifelogger was lecturing to a group of people in a classroom environment
11	Cooking	Find a moment when the lifelogger was cooking at home
20	Eating Pasta	Find a moment when the lifelogger was eating pasta

Table 4.3: User Interaction Study - Topic Group C

Each user performed their group of four topics on a different system variant and in a different arrangement to address the learning bias across all users (see example in Figure 4.12) utilising the Latin square technique described in chapter 3. In addition, each time a user switched to a different system variant, they received a thorough walkthrough of the new system and its features as well as being required to complete a practice known-item search task using a test topic. The user was instructed that they had a total of 180 seconds to submit an image

ID	Title	Description
4	Sunset	Find a moment when the lifelogger was viewing a sunset
8	Grocery Shopping	Find a moment when the lifelogger was grocery shopping
15	Having Beers in a Bar	Find a moment when the lifelogger had more than one beer in a bar
22	Benbulbin Mountain	Find a moment when the lifelogger was looking at Benbulbin mountain

Table 4.4: User Interaction Study - Topic Group D

matching the criteria described by the topic, and had an opportunity to clarify their understanding of the topic description before they began. This time limit was chosen because we determined from ad-hoc testing that it provided sufficient time for most users to make a strong attempt at retrieval without burdening the allotted experimentation time. When the user identified an image they wished to submit, they were instructed to notify the examiner who would inform the user if their submission was correct. There was no penalty for submitting an incorrect image and the user would continue to submit until time ran out.

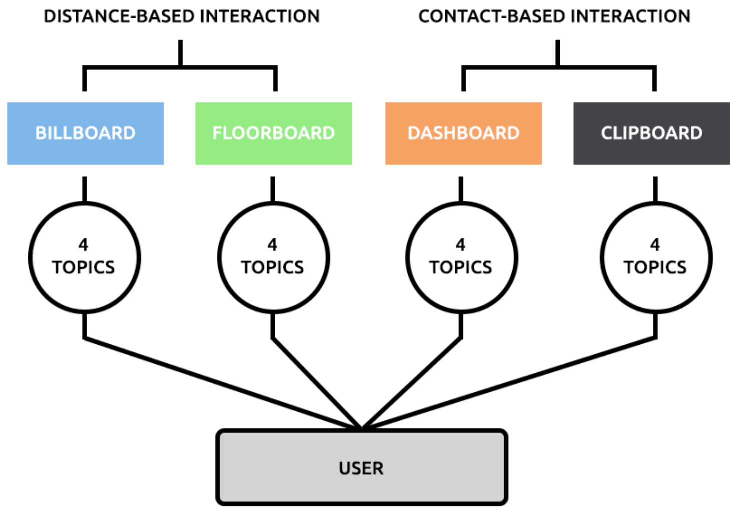


Figure 4.12: Example configuration for one participant in the user study

4.5.1 User Performance

During the experiment, the length of time taken to complete each known-item search task was recorded along with the number of times the user could not complete the task in the allotted time frame. After all the users had participated in the experiment, each topic had been searched for a total of 16 times, 4 times on each of the 4 virtual interface variants. In an ideal experiment, we would recruit several hundred participants with identical technical experience who could perform the exact same topics on each of our experiment variants. However, since it was impossible to source this many experienced users, and topics could only be used once per user before becoming useless, the focus shifted to using the quantitative assessment in conjunction with the qualitative feedback to help inform our design methodology. This is especially relevant when we consider that for each topic, a single poorly performing user can notably impact the perceived average performance of one of the experiment variants. As we mentioned briefly in our previous chapter, it is likely that this study would be more reliable if the number of participating users was increased.

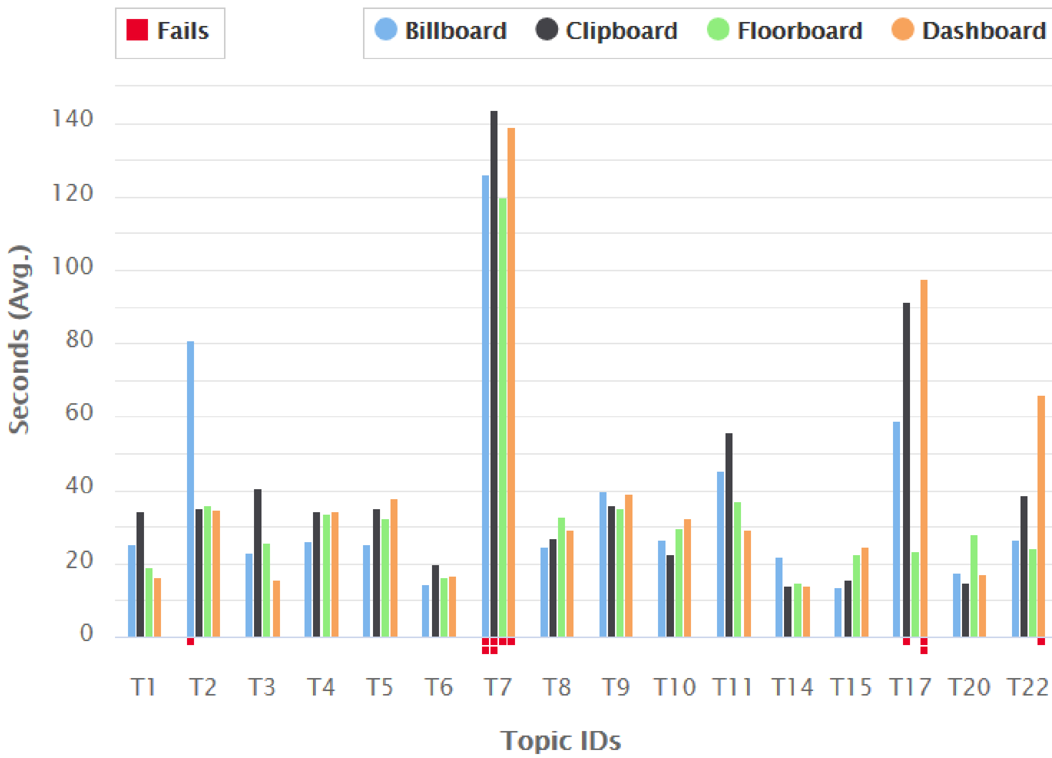


Figure 4.13: Average seconds taken per topic for each system variant (with total failures below axis)

In Figure 4.13 we can observe a bar chart displaying the average time taken to successfully identify a topic on each of the four interaction variants. The 16 topics are labelled on the horizontal axis and the average time taken in seconds is labelled on the vertical axis. The four interaction variants are represented by four coloured bars for each topic and there is an indication in red beneath each topic of how many times a user exceeded the 180 second time limit on an interaction variant and failed the task. The total average seconds over all topics with error margins is shown in Figure 4.14.

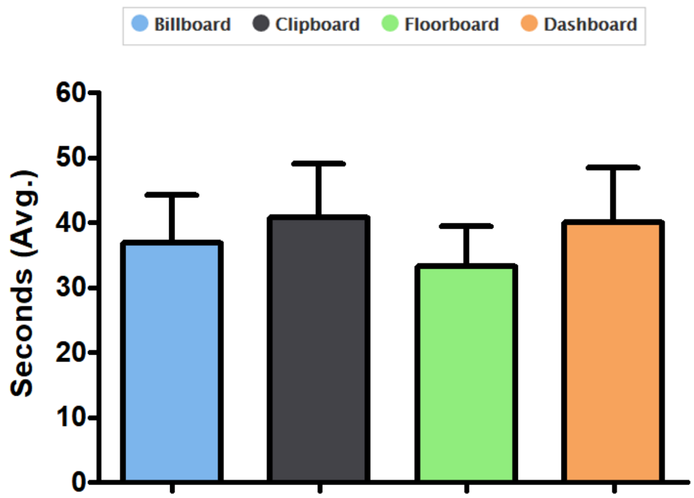


Figure 4.14: Total average seconds for each user interaction method with error margins

For the majority of topics, the interaction approaches performed similarly, suggesting that there is no definitively superior interaction approach. However there was some inconsistency in topics T2, T17 and T22. The fact that these are also the only topics which were failed by a number of users suggests this inconsistency is less related to the interaction type and more likely the result of how some users interpreted the topic description. For example, T17 was a topic describing the lifelogger being recorded for a television show, and many participants correctly used 'camera' as a concept to filter with, as logically the lifelogger's personal camera would capture the television camera recording them. However, many participants failed to make this connection and instead used the 'television' concept which resulted in a significant number of false positives being returned.

It is immediately apparent that T7 proved the most difficult topic for participants, having the highest average time across all interaction types and the most failed attempts. For this topic, the users were asked to locate an image where the lifelogger was presenting or lecturing to students in a classroom environment. However, there were no obvious concepts in the test collection related to this topic ('classroom', 'presentation', etc. did not exist), so it was universally challenging for all users across all interaction types.

4.5.2 User Feedback

The participants were asked to fill out a user experience questionnaire after each group of topics regarding the interaction approach they had just used. Each questionnaire contained usability statements which the users needed to state their level of agreement with on a five-point Likert scale; the answers to which are visualised in Figures 4.15, 4.16, 4.17 and 4.18. Most importantly, the users were asked a series of open questions as part of an informal interview regarding the usability of each interaction approach and their thoughts on the system as a whole. At the end of the experiment, the participants were asked to rank the four interaction types in order of their preference. The full questionnaire can be viewed in Appendix A.1.

Billboard

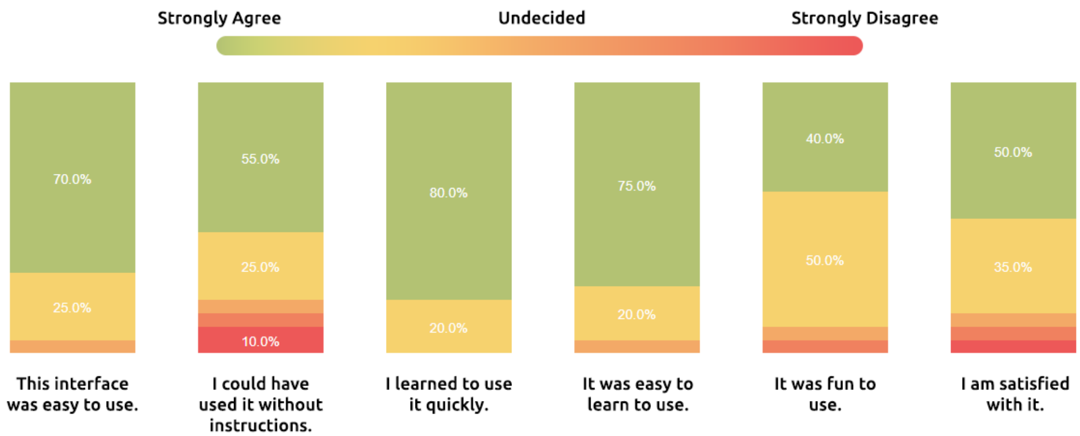


Figure 4.15: User feedback for the Billboard style interaction

Sentiment toward the billboard style of interaction was generally quite positive, with many users discovering how to interact with the virtual interface prior to even being instructed. This is likely related to users' familiarity with operating a television using a remote control which also served as the inspiration for this approach. Some users stated the upright positioning of the interface caused minor neck strain toward the end of the experiment, but it was not severe. Additionally, many users elected to operate the virtual user interface using only one hand and suggested the opposing hand could serve another function rather than just another selection mechanism. Finally, despite the option to move farther or closer to the virtual interface at any time, users stated they would have preferred the ability to adjust its position without having to move themselves.

Floorboard

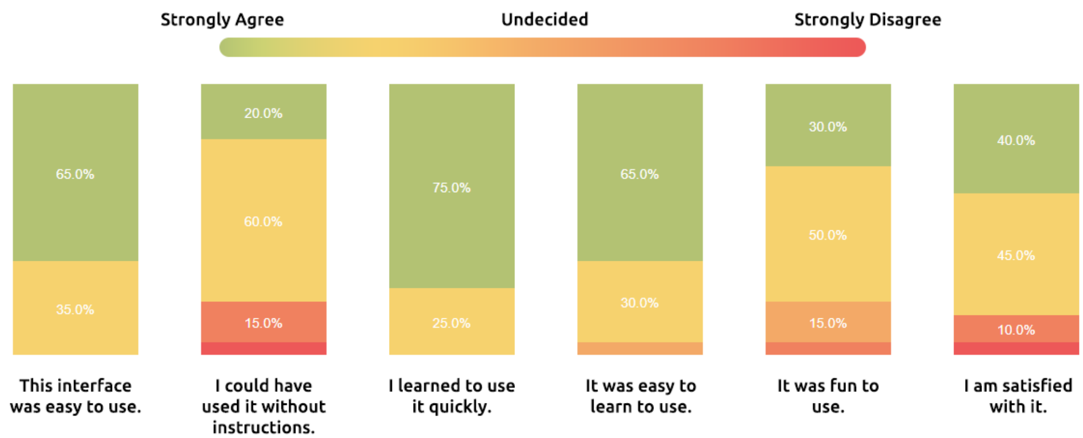


Figure 4.16: User feedback for the Floorboard style interaction

Sentiment toward the floorboard style of interaction was largely identical to the billboard style, suggesting the precise positioning of the virtual interface for distance-based interactions did not have a significant impact on usability. However, a notable portion of users stated they preferred operating with the interface at a lower position as it felt more comfortable for extended periods. It should be noted that a minority of users stated the exact opposite of this so a definitive preference can not be determined. This reinforces our view that providing a mechanism for the user to adjust the positioning of the virtual interface at any time is necessary to adapt to every user's preference. However, deliberately omitting this feature has helped us determine that occasionally exposing virtual information at different heights is not particularly disruptive or uncomfortable for many users, especially for short periods.

Dashboard

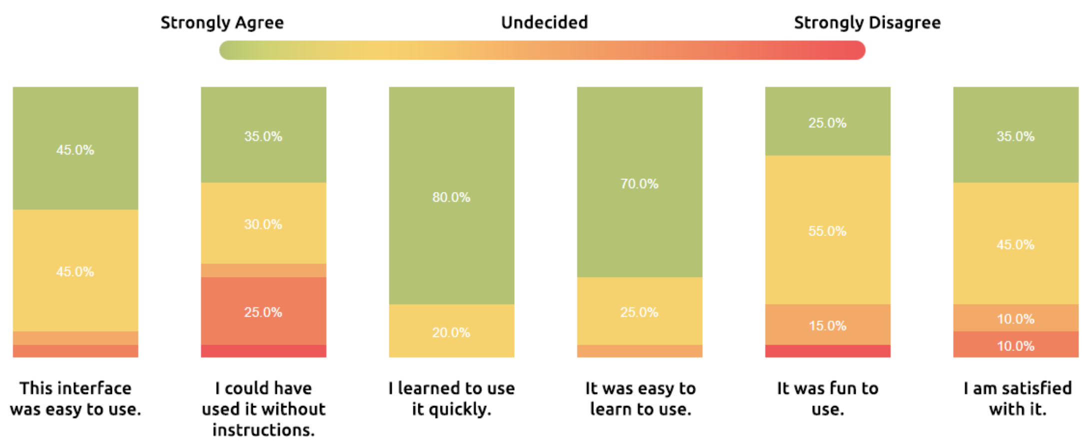


Figure 4.17: User feedback for the Dashboard style interaction

The feedback regarding the dashboard style of interaction was divisive. Some users disliked having to directly connect with the virtual interface, citing difficulties with depth perception and minor arm strain over extended periods. However, in contrast other users adapted very quickly to this style and felt they were performing interactions faster and more precisely. We identified that the former users were generally less experienced or comfortable with virtual reality and the latter users were the opposite, either having some experience, or simply a high degree of comfort, operating in the virtual environment. This suggests that the dashboard style of interaction is potentially quite efficient but is not necessarily suitable for novice users as directly contacting virtual objects requires some practice to become accustomed with.

Clipboard

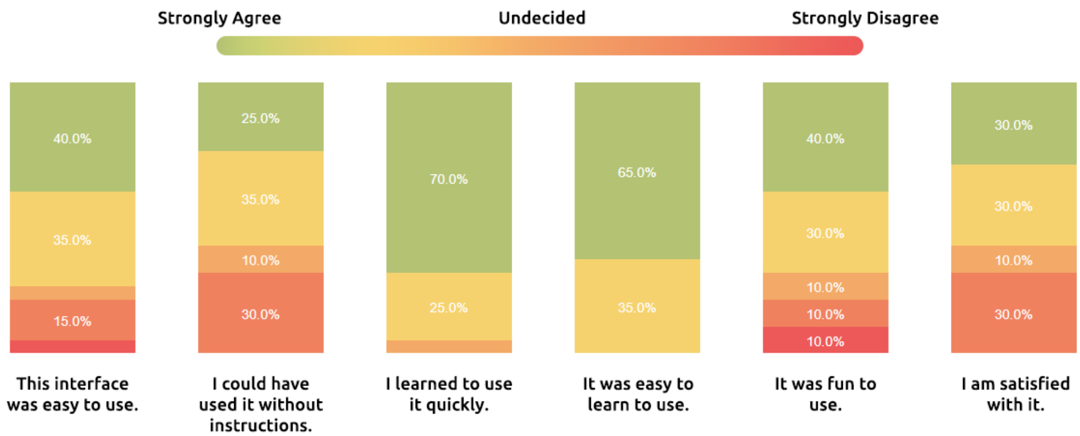


Figure 4.18: User feedback for the Clipboard style interaction

Out of all the interaction styles the clipboard received the most negative feedback. Similar to the dashboard, some users described issues related to depth perception and arm strain but it was notably worse using the clipboard approach. Any users who felt faster or more precise with the dashboard generally felt that the clipboard accomplished this to a lesser extent. The primary criticisms came from the necessity to coordinate both hands to perform any interface interaction. Furthermore, users felt obligated to hold the virtual interface upright and ready at all times, rather than just when required, which significantly increased arm strain especially after an extended period.

General

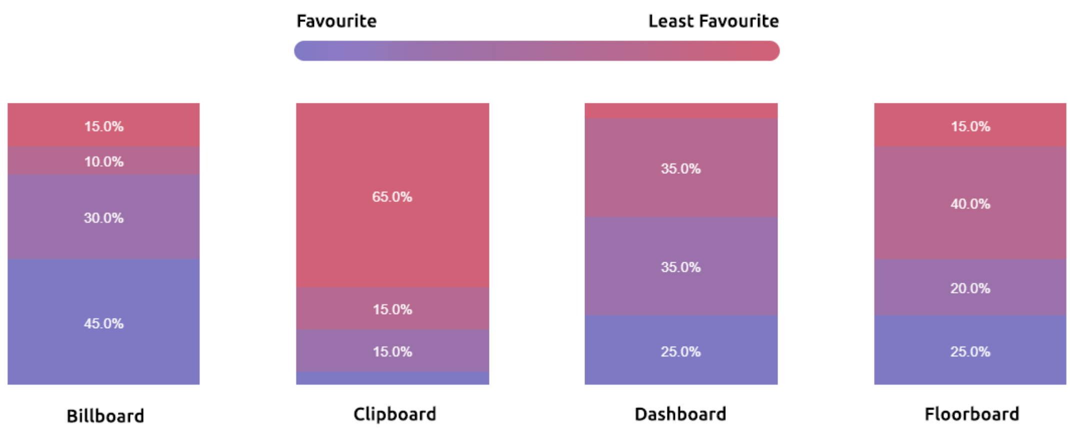


Figure 4.19: Order of preference for each interaction type

At the end of the experiment, each of the participants were asked to rank the four interaction types from their favourite to least favourite. In Figure 4.19 we can observe that, as we might expect from the previous feedback, the clipboard was definitively the least favoured approach. However, preference regarding the remaining three interaction methods was comparably mixed. Overall the billboard style of interaction was favoured most by the majority of users, but the dashboard and floorboard approaches were not far behind, often being selected as a user's second preference. It is also interesting that despite notably positive feedback regarding the floorboard, it was generally favoured less than the dashboard approach overall. A number of users stated that their preference was heavily impacted by the inability to adjust the position of the distance-based interaction approaches versus the contact-based counterparts. The option to adjust the position of the dashboard style, along with its affordance of faster and more precise interaction for some users, has likely contributed to its more positive reception.

Recurring feedback toward the lifelog retrieval system across all interaction types was centred primarily on the lifelog concepts themselves. Many users wanted more concept options and found some of the existing concepts to be inaccurate. Further confusion arose when concepts were labelled accurately but were not what users expected. For example, a user trying to identify an image that occurred inside the lifelogger's home might select the 'house' concept, not realising this concept only applied to exterior buildings resembling houses. To address this issue, some users attempted to pair together 'house' and 'indoor' but this only produced similarly confusing results. This is because the underlying content analysis identifying these two concept labels is searching for opposing features, namely interior versus exterior features, so it is unlikely that these two concepts would ever appear labelled on the same image. Adequately conveying this information to users proved difficult and it is clear this confusion had a notable impact on user performance. Finally, users expressed a desire to filter concepts using a search feature where they could type out potential concepts to see if they exist, rather than searching for them in an alphabetical list.

4.5.3 Analysis

Due to the relatively small scale nature of this user study, both in scope and number of participants, it is unwise to make any definitive claims regarding lifelog retrieval in virtual reality at this time. The primary goal of this study was to evaluate our various interaction variants and how they might support lifelog retrieval. In that respect, we determined that the most effective interaction approach to support lifelog retrieval is likely a hybrid methodology, utilising aspects of both the distance and contact-based interaction methodologies. For example, interactions requiring low precision, such as pointing and selecting, benefit most from a distance-based approach as it appears to be efficient, comfortable and intuitive for many users. When interacting with a virtual interface in this context, it is important to allow the user to adjust the position of the interface so it can adapt to the user's preference. When it comes to higher precision tasks, such as targeting smaller elements or gripping virtual objects, a contact-based approach is likely more appropriate. For example, if we were to implement a text-based concept filter, we would enable users to invoke a virtual keyboard which could be interacted with using the contact-based methodology. Once they had successfully located a relevant concept, the keyboard would disappear and the user would return to the distance-based interaction methodology.

With respect to issues regarding concept accuracy and clarity of meaning, it is important to note that these are issues which are not unique to virtual reality, but rather are symptomatic of the modern techniques supporting lifelog retrieval across all platforms. However, these issues still must be addressed in order to produce an effective retrieval system and different access mechanisms might address these issues in different ways, so it is worth acknowledging it at this time. Furthermore, though there is little a lifelog system can do to improve the accuracy of its underlying dataset, it should certainly assist in clarifying the nature of this data. For example, in the situation where users misinterpret the definition of a concept like 'house', a solution might be to categorise the concepts under certain headings, such as 'exterior' or 'interior', to better contextualise the concept during selection. This is not an exhaustive list of every insight determined by this user study, but should outline the primary influences which will be used to inform our design methodology and subsequent evaluations.

4.6 Conclusions

In this chapter we have begun our investigation into developing effective methods of generating lifelog retrieval queries in virtual reality. This consisted of evaluating a series of potential user interaction methodologies based on a combination of existing literature and recurring conventions in lifelog applications. From the perspective of user performance, we conclude that there was no clearly superior interaction methodology, though we acknowledge a larger scale study might produce more firm statistical data. From the perspective of user feedback, we conclude that, though the distance-based interaction methodology was more positively received, it is likely that as our prototype continues to evolve, a hybrid interaction methodology would be most appropriate as it would better address both low and high precision interactions. As we have already stated, our objective was to utilise the insights garnered from this study to inform our subsequent design decisions, with the intent that our comparative analysis with a conventional analogue will validate these decisions. It is for this reason that, though the subject matter covered in this chapter explicitly contributes to our second research question, it cannot be fully addressed until later in this work following our comparative analysis with a conventional lifelog system analogue. In our next chapter we will begin to investigate methods of visualising the lifelog data in a manner that supports lifelog retrieval, which in turn will contribute to our third and final research question.

Chapter 5

Visualising Lifelog Data in a Virtual Environment

Our third research question focuses on the visualisation of lifelog data in virtual reality and how it might compare to a conventional alternative. Physically occupying a virtual environment affords users three spatial dimensions to interact with digital information and the potential methodologies to support this are practically infinite. However, in the context of lifelog retrieval, the salient factors are how quickly and easily specific digital information can be retrieved. To support this, users must be able to visually digest lifelog data as efficiently as possible. In this chapter we motivate an event-based visualisation strategy and evaluate a selection of summarisation techniques to establish a baseline for their effectiveness in supporting lifelog retrieval in virtual reality. Similar to our previous study, because we were not yet evaluating against a conventional lifelog system analogue, our objective was to utilise the insights garnered from this work to inform our subsequent design decisions, with the intent that our comparative analysis would directly address the third and final research question and validate our overall hypothesis.

5.1 Visualising Continuous Image Streams

Like the majority of lifelog datasets aiming for total capture, the primary focus of the NTCIR-13 test collection is on continuous streams of images captured from the perspective of the lifelogger. Any one of these thousands of images, when retrieved, could represent a potential cue to promote autobiographical memory. Within the academic community there are currently three active areas of research in the context of lifelog retrieval and continuous image streams. These can be easily identified based on the lifelog tasks presented at the more recent lifelog focused conferences, such as NTCIR-13 (Gurrin, Joho, Hopfgartner, Zhou, Gupta, et al., 2017). One area of focus is on the annotation of lifelog image data, referred to as lifelogging concepts, which is typically achieved by an automatic visual analysis. Another area of focus is on establishing insights in the lifelog data, such as by identifying activities, or contextualising the data using the lifelogger's multimedia and biometric data. The final active area of focus is on event segmentation, or the grouping of continuous image streams into semantic categories based on their visual content (Doherty and Smeaton, 2008; Lee et al., 2008; Doherty, Gurrin, and Smeaton, 2009).

We have already established that for the scope of research carried out in this work, we did not incorporate the multimedia or biometric data provided by the NTCIR-13 test collection. As our focus was on the test collection's images and associated metadata, namely the temporal, concept and event data, we determined an event-based visualisation strategy incorporating the concept and temporal data would serve as a suitable baseline to support lifelog retrieval in virtual reality. This decision was further influenced by the fact that event-based visualisation strategies have been implemented on a number of conventional lifelog applications in the past to varying success (Lee et al., 2008; Doherty, Moulin, and Smeaton, 2011), but their effectiveness in a virtual environment has yet to be evaluated. It is likely that the increase in accessible dimensions, highly immersive quality and ability to coexist in the same space as the visualised data could have a notable impact on supporting lifelog retrieval.

5.1.1 Ranking Events

Before we can discuss our event-based visualisation approaches, we must first describe the underlying ranking algorithm which determines the relevance of our data in relation to the user's query. Since a user query only consists of two pieces of metadata, a temporal facet and a concept facet, and the temporal facet can be easily filtered from any set of results, the primary metadata we must use to determine an event's rank is the concept score of its constituent images. When ranking individual images, a concept score can be calculated very simply; if the image contains all of the queried concepts, it is perceived as having the maximum possible concept score, and if the image contains none of the queried concepts, it is perceived as having a concept score of zero. Therefore, in a situation where a user submits a query containing three concepts, we can rank images containing all three at the top of our list, followed by images containing two, before finally the remaining results which only contain one of the submitted concepts. The largest drawback to this approach is that, due to the continuous nature of capture, the image results are often cluttered with hundreds, sometimes thousands, of nearly identical images, and if a group of those images are not related to the topic the user is trying to retrieve, they immediately become an obstacle to successful retrieval. This is where event segmentation can be very helpful, as it can reduce the impact of these obstacles. However, ranking events based on concept score is not as simple. It is important to note that the ranking algorithm described in this thesis is not a reflection of the state-of-the-art as determining the most effective ranking algorithm for lifelog events was not within the scope of research carried out for this work.

We determined for our purposes that there were two primary factors which should impact an event's rank with respect to the user's concept query. First, and most importantly, how often the user's submitted concepts appear together in the same images in an event. This is important because, while an event is still potentially relevant when it contains numerous images labelled with the individual submitted concepts, an event where those concepts all appear on the same images should always be considered more potentially relevant as every concept the user is querying exists in one image. The second factor impacting rank is how many of the

Rank	Event Ranking Breakdown
1^{st}	3 queried concepts appear together on the same image(s) in the event
2^{nd}	3 queried concepts appear, but only 2 of them appear on the same image(s) in the event
3^{rd}	3 queried concepts appear, but not together on the same image(s) in the event
4^{th}	2 queried concepts appear together on the same image(s) in the event
5^{th}	2 queried concepts appear, but not together on the same image(s) in the event
6^{th}	1 queried concept appears on an image in the event

Table 5.1: Breakdown of event ranking when querying 3 concepts

queried concepts appear in the event irrespective of whether or not they exist on the same images. This is important because if an event contains a small number of images labelled with every submitted concept, where the concepts never appear together on the same image, it should always be ranked higher than an event which contains a large number of images labelled with only some of the submitted concepts. A practical example of this ranking algorithm is summarised in Table 5.1 where we can observe the breakdown of a query where the user has submitted three concepts. In the very rare situation that concept scoring is identical between two events, the events are ranked temporally from earliest to latest.

The formula used to calculate an event's concept score can be seen in Equation 5.1 where S is the concept score, x is the number of concepts queried by the user and n is the total number of images in the event. R_i is the number of images where all but i queried concepts where found. For example, R_0 is the number of images where all queried concepts were found, R_1 is the number images where all but 1 queried objects were found, and so on.

$$S = \sum_{i=0}^{x-1} \frac{R_i}{(n+1)^i}$$

Figure 5.1: Formula to calculate an event's concept score

In Equation 5.2 we can see an example score equation with an event containing 9 images and a user query containing 3 concepts. It is important to observe how each subsequent iteration of R_i is progressively down-weighted to ensure that an event containing 1 image tagged with n concepts is always weighted above an event where every image is tagged with $n-1$ concepts.

$$\begin{aligned} S &= \frac{R_0}{(9+1)^0} + \frac{R_1}{(9+1)^1} + \frac{R_2}{(9+1)^2} \\ &= R_0 + \frac{R_1}{10} + \frac{R_2}{100} \end{aligned}$$

Figure 5.2: Example score equation with 9 images and 3 queried concepts

In Equation 5.3 we can see a full calculation of the above example where all the queried concepts were found in 2 images, 2 queried concepts were found in none of the images and 1 queried concept was found in 7 images, therefore $R_0 = 2$, $R_1 = 0$ and $R_2 = 7$.

$$S = R_0 + \frac{R_1}{10} + \frac{R_2}{100}$$

$$= 2 + \frac{0}{10} + \frac{7}{100}$$

$$= 2 + 0 + 0.07$$

$$= 2.07$$

Figure 5.3: Example of one event's concept score calculation

5.1.2 Visualising Events

Now that we have established a method of ranking events via their concept score, we can begin to discuss methods to visualise these events. However, since the goal of successful lifelog retrieval is not an event itself, but rather the images it contains, it is necessary to somehow expose these images to the user. As we have already established, displaying every image in an event will quickly inhibit retrieval rather than assist it. A simple solution to this is to summarise an event's content using a subset of its images which can then be further explored if the user considers the event to be relevant. However, then we must determine which images best convey the content and potential relevancy of an event, and how many should be exposed before it becomes inhibitive.

In total, the NTCIR-13 test collection contained 1865 unique events. These events were not annotated with a label or activity and did not conform to a specific length of time or number of images. Only 10% of the events in the dataset contained more than 100 images, with the majority containing less than 30 images. In Figure 5.4 we can observe a complete frequency distribution of event sizes. With respect to determining an appropriate event summary length, we originally suspected that larger events (containing 100+ images) might require a larger summary, but as large events were rarer, and often contained a significant number of visually identical images, affording it a larger summary most often only served to increase visual clutter and distract from other potentially relevant events.

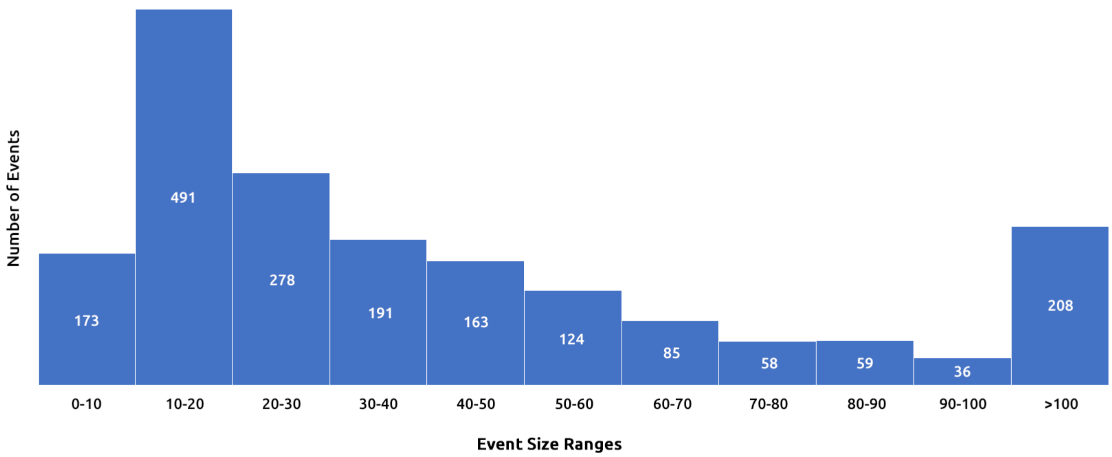


Figure 5.4: Frequency distribution of event sizes

Through informal testing, it quickly became apparent that above a certain threshold, increasing the number of images in an event's summary only served to increase the time necessary to visually digest it without notable benefits to discerning its potential relevancy. This threshold was determined to be as little as three or four images for some events whereas others benefited from as many as six or seven. Continuing to increase this number resulted in a rapid decrease in efficiency, where users spent more time having to examine events or even began to ignore images completely. Though there is potential for future work in this area, it was decided that a nine image summary would adequately convey the relevancy of the majority of events in the dataset without notably impacting efficiency and also could be conveniently arranged in a 3x3 grid for ease of viewing (see Figure 5.5).

Figure 5.5: Events ranked visually within the virtual environment

The design of this horizontal grid arrangement was based on a number of criteria. First, unlike conventional applications, the data needed to scroll horizontally rather than vertically. This is because positioning virtual elements below a user can convey a sense of being high off the ground which causes severe vertigo for some individuals. Second, the size of the virtual images needed to be large enough to retain their clarity when positioned within the user's reference frame but small enough so as not to dominate the user's visual range. Third, once an appropriate size of image was determined, we needed to establish the maximum number which could be displayed vertically without demanding excessive head movement from the user. We determined this number to be three, as using less than three images poorly utilised the available virtual space and using more than three images demanded notably more head movement.

Now that we have established an appropriate size for the event summaries, we must determine a method of selecting appropriate images to effectively convey the event's content and potential relevancy. We refer to the images in our summary as event keyframes and, though there has been several approaches to keyframe selection within video retrieval (Cooper and Foote, 2005; Rogério dos Santos Alves; Alex Soares de Souza, 2014), there has been little work on keyframe selection within the lifelogging domain. For example, Lee et al. (2008) and Doherty, Byrne, et al. (2008) describe a selection of keyframe selection techniques, but their scope is restricted to extracting a single keyframe to represent an event. For our work we determined to evaluate three approaches to keyframe selection, two primary selection methodologies and one methodology which did not visualise events at all to serve as a control. The objective was to investigate the most effective summarisation method to support lifelog retrieval and directly contribute to answering our third research question.

Temporal Event Summary

We referred to the first keyframe selection methodology as the temporal event summary. This is because the selection process is characterised entirely by the temporal arrangement of images present in the event. The goal was to provide an unbiased summary of the event by selecting nine equally spaced images ordered from earliest to latest where three of the images were always the first, last and centermost image occurring in the event. This approach is comparable to the computationally inexpensive keyframe selection method utilised in video retrieval (Smeaton and Browne, 2006) where simply choosing the middle frame of any video clip proved effective in representing the full video. Since we were not restricted to selecting one keyframe, we also included the first and last images as these images signalled the beginning and end of the visually similar features which resulted in the event's segmentation and therefore could serve in assisting to contextualise the event's content.

For example, we can observe in Figure 5.6 that the beginning and end of an event where the lifelogger is sitting down to work will often capture images of the lifelogger having just entered and about to leave the office. These images could serve as valuable keyframe images for users in contextualising the event or its potential relevancy. To further assist this, each event is accompanied by an overhead tooltip conveying how many images are in the event, how long the event occurred for and the date and time it took place.

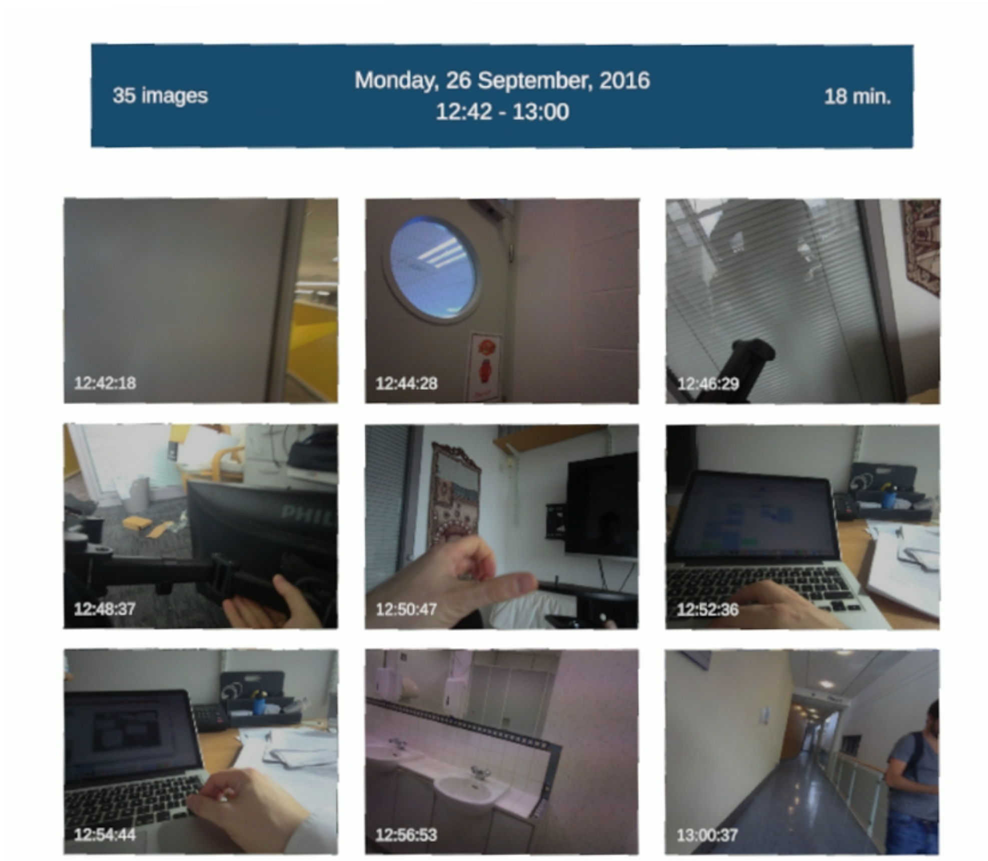


Figure 5.6: Summary of an event using temporal method

Ranked Event Summary

We referred to the second method of selecting keyframe images as the ranked event summary. This is because, although the images in the summary are still arranged temporally from earliest to latest, the keyframe selection is dictated by the concept score of the images with respect to the user's query. The goal was to provide a summary of the event with a bias toward images with a high concept score. Specifically, this approach compiles a list of every image in the event ranked by the number of the queried concepts each image contains. If there are nine or more images in this ranked list, the top nine images are chosen for the summary of the event. However, when there are less than nine images, the system reverts to an adapted version of the previous event summary where it selects keyframe images equally spaced around the ranked images which have already been selected.

Figure 5.7: Summary of an event using ranked method

In Figure 5.7 we can observe how both the temporal and ranked event summaries are visually very similar but, in addition to the overhead tooltip, the ranked summary also highlights specific images using an icon in the top right corner resembling a star and red ribbon. This icon indicates that the corresponding image has been labelled with at least one of the queried concepts; the specific number being indicated within the star section of the icon itself. This functionality is based on previous work emphasising specific event keyframes to users (Doherty, Moulin, and Smeaton, 2011) though this was more often accomplished via increasing the size of the keyframe image rather than using an icon. However, as we had determined a very specific scale of image for rendering in virtual reality, adjusting the size of some images was not appropriate for our system.

Control (No Events)

Whilst the benefits of utilising events to prevent visual clutter and reduce the impact of repetitive data are notable, we should not assume that it is the best way to support lifelog retrieval. Any obscuring of data, even when trying to reduce clutter or repetition, must be considered carefully or we risk introducing unintended biases that could inhibit retrieval in a manner that outweighs the potential benefits. For this reason, in addition to the temporal and ranked event summaries, we also implemented an approach which does not incorporate any event visualisation. This is intended to serve as a control during our evaluation and is identical to the visualisation method utilised in our user interaction study where each image is individually ranked based on how many queried concepts it is labelled with and there is no grouping or obscuring of visually similar images (see Figure 5.8).

Figure 5.8: Control visualisation with no event visualisation

5.2 User Interaction

Similar to our previous chapter's primary focus being on user interaction, this chapter's focus is deliberately on data visualisation and presentation. However, in focusing on this we have introduced new features to our lifelog retrieval system which require corresponding new user interactions. For example, with respect to the event summary, we now require a mechanism by which users can explore an event in its entirety. To achieve this, along with supporting other possible feature implementations, we established the concept of a contextual interface that would only be exposed when necessary. This type of contextual interface is common in event-based lifelog applications (Lee et al., 2008; Doherty, Moulin, and Smeaton,

2011) where it is typically achieved simply by hovering over relevant images.

An initial design overlaid the contextual interface elements on target images within the virtual space but, after informal testing, this was determined to be unsuitable as it became clear the overlaid elements were obscuring the data the user was trying to examine and negatively impacting retrieval. One solution to this was to redesign the interface elements to be smaller and less obtrusive, however, small icons and text rendered within the virtual environment needed to be within a specific range to remain clear and legible to users. This is because the resolution afforded by the virtual reality hardware is low enough that aliasing (jagged edges) can occur after a short distance. Since users were capable of moving around the virtual space and adjusting their position with respect to the data, we could not rely on them always being within the appropriate range to render interface elements clearly.

To address this, the contextual interface was redesigned so that it would appear alongside the user's active controller, which meant that it would always appear within the appropriate range. In Figure 5.9 we can observe a user pointing at a specific image with the contextual interface rendered slightly above the controller and providing two options based on the current target of interest. This style of interaction was chosen based on evidence suggesting users prefer object-action sequences over action-object sequences as it requires less mental effort (McMahan and Bowman, 2007). To navigate between these options the user uses their thumb on the controller's touchpad to highlight and make their selection. The context menu currently provides three possible functions depending on what image is being

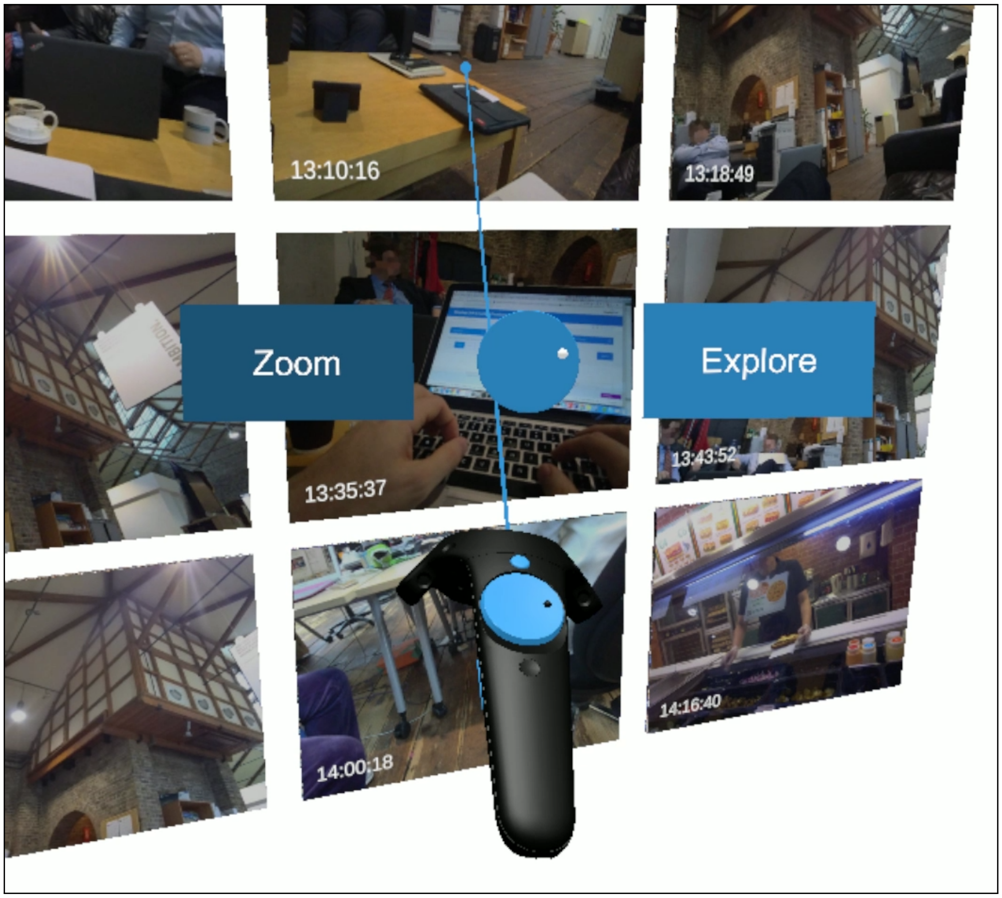


Figure 5.9: Context menu appearing above active controller with two available options

targeted. The first and most important function is 'Explore', which enables the user to explore all the images in an event, or in the case of our visualisation without events, exploring all the images captured thirty minutes before and after the target image. These explored images are presented in a line in front of the previously rendered results (see Figure 5.10) and can be navigated or scrolled through in an identical fashion by gripping the controller and performing a throwing gesture in the chosen direction.

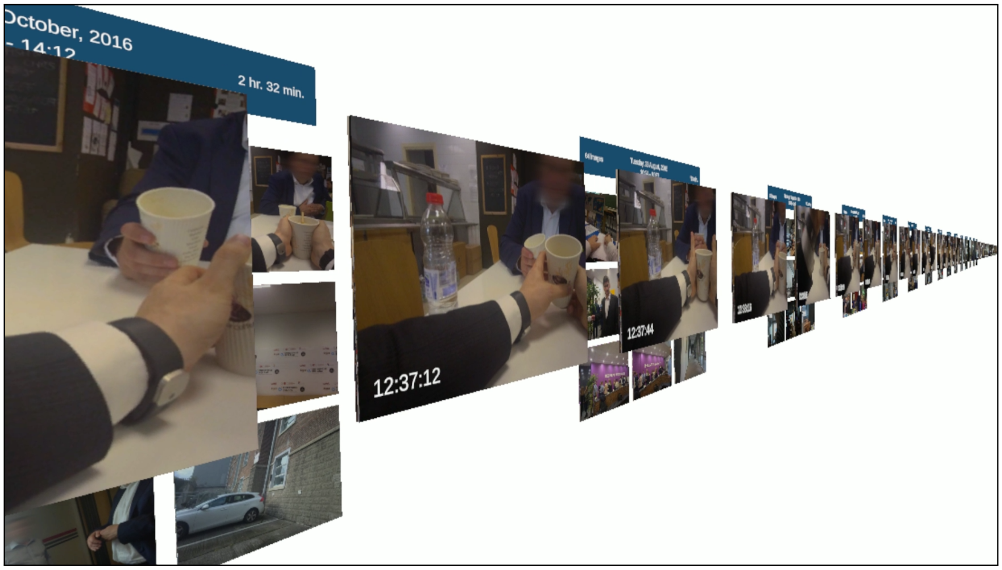


Figure 5.10: Explored images positioned in front of ranked results

The remaining two functions provided by the contextual interface are 'Zoom' and 'Search Tags' which are intended as secondary features provided to improve the quality of life for some users. The 'Zoom' option significantly increases the scale of a target image to make it easier to examine. This is helpful in rare situations were the image is particularly detailed or contains a concept which may only occupy a small portion of the image. The 'Search Tags' option copies all of the concepts the target image is labelled with and reloads the main menu with those concepts prepared for submission. This is helpful when an image contains numerous relevant concepts and the user wants to quickly submit a related query.

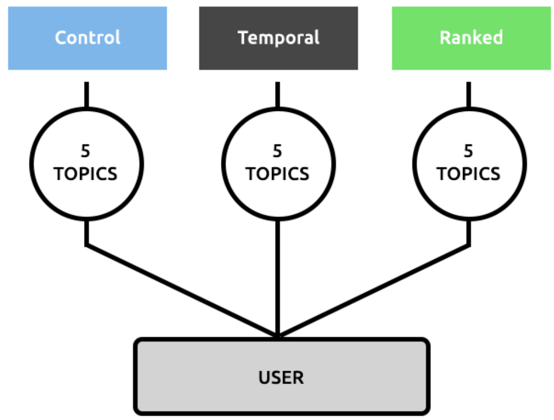


Figure 5.11: Example configuration for one participant in the user study

5.3 Data Visualisation Study

Now that we have determined two potential visualisation methods using events (ranked event summary and temporal event summary), and a control method (using no events), it is necessary to evaluate their effectiveness in relation to finding relevant content. As in our previous chapter, this was achieved via a user study involving known-item search tasks. For this experiment we recruited 18 volunteer participants, of which about half were volunteers from our previous user study. This meant we could not use the same topics provided by the NTCIR-13 lifelog semantic access task (LSAT) as many of the participants would have prior knowledge of how to quickly retrieve these topics. To address this, it would be necessary to request additional topics from the lifelogger who generated the NTCIR-13 test collection. Fortunately, we had access to this individual and they were able to provide 15 new known-item search tasks for use in our experiment. Following the methodology described in our previous chapters, the total number of topics was divided evenly by the total number of variants to produce 3 topic groups containing 5 topics each (see Figure 5.11). These groups are shown in Tables 5.2, 5.3 and 5.4 where we can observe the topic ID, title and accompanying descriptions which were provided to users before beginning each retrieval task.

ID	Title	Description
1	Waiting for a Train	Find a moment when the lifelogger was waiting at a train station
2	Looking at a Mirror	Find a moment where the lifelogger was looking at themself in a mirror (lifelogger must be visible in mirror)
3	Packing a Suitcase	Find a moment when the lifelogger was packing a suitcase
4	Gold Living Room	Find a moment when the lifelogger was sitting in a gold coloured living room
5	Picking Berries	Find a moment when the lifelogger was picking berries on a Sunday afternoon (berries must be visible in hand)

Table 5.2: Data Visualisation Study - Topic Group A

As before, each user performed their group of five topics using a different visualisation method and in a different arrangement to address the learning bias across all users, utilising the Latin square technique described in chapter 3. Each time a user switched to a different visualisation technique, they received a thorough walkthrough describing how it worked as well as being required to complete a practice known-item search task using a test topic. We maintained the 180 seconds time limit to submit an image matching the criteria described by the topic and users still had an opportunity to clarify their understanding of the topic

ID	Title	Description
6	Whiteboard Session	Find a moment when the lifelogger was writing on a whiteboard (marker must be touching whiteboard)
7	Asian Restaurant	Find a moment when the lifelogger was eating food in a restaurant using chopsticks
8	Aeroplane	Find a moment when the lifelogger was walking across an airport apron (apron is where aeroplanes are parked in the airport)
9	Inside a Church	Find a moment when the lifelogger was inside a church
10	Morning Coffee	Find a moment when the lifelogger was having coffee on a Saturday morning

Table 5.3: Data Visualisation Study - Topic Group B

description before they began. When the user identified an image they wished to submit, they were instructed to notify the examiner who would inform the user if their submission was correct. There was no penalty for submitting an incorrect image and the user would continue to submit until time ran out. Finally, it is important to note that for the purposes of this experiment every user performed their tasks using the same interaction methodology, specifically the 'billboard' style of interaction, as overall feedback from the previous study suggested it was the most preferred approach, especially for novice users.

5.3.1 User Performance

As per our previous experiment, the length of time taken to complete each known-item search task was recorded along with the number of times the user could not

ID	Title	Description
11	Fruit Bowl	Find a moment when the lifelogger was looking at a bowl of fruit
12	Writing on Paper	Find a moment when the lifelogger was writing on a piece of paper (pen/pencil must be touching paper)
13	Eating Eggs	Find a moment when the lifelogger was eating eggs
14	Using a Ticket Machine	Find a moment when the lifelogger was using a ticket machine to get a train ticket
15	Antique Clock	Find a moment when the lifelogger was looking at an antique clock early on a weekday

Table 5.4: Data Visualisation Study - Topic Group C

complete the task in the allotted time frame. After all the users had participated in the experiment, each topic had been searched for a total of 18 times, 6 times on each of the 3 visualisation methods. In Figure 5.12 we can observe a bar chart displaying the average time taken to locate a relevant image using each of the visualisation approaches. The 15 topics are labelled on the horizontal axis and the average time taken in seconds is labelled on the vertical axis. The three visualisation methods are represented by three coloured bars for each topic and there is an indication in red beneath each topic of how many times a user exceeded the 180 second time limit using a visualisation method and failed the task. It should be acknowledged that the topics created for this study were notably more difficult than the previous topics released for the NTCIR-13 retrieval task. This is likely due to the inherent difficulty generating unique topics which are not overly specific and are challenging but also not impossible to retrieve. The total average seconds over all topics with error margins is shown in Figure 5.13.

Despite some minor exceptions, it is immediately evident that the ranked summary approach served as the most efficient visualisation approach. Even in the three most difficult topics T4, T9 and T14, where the majority of users exceeded the time limit for retrieval, the ranked summary enabled users to complete the task in the shortest amount of time. The primary issue users faced with these more difficult topics was the availability of relevant concepts. For example, T9 described a situation where the lifelogger was inside a church, but there were no obvious concept such as 'church' available in the metadata, so users spent a lot of time trying to improvise with words like 'cross' and 'statue', which were also not very helpful. Yet even in these situations where concept availability was not ideal, the ranked summary approach resulted in the least number of failures and the shortest amount of time to complete the retrieval tasks. This suggests that effective visualisation can have a notable impact in situations where the lifelog metadata is perceived by the user to be poor or even unavailable.

However, upon observing the performance of the temporal summary approach, there is also evidence to suggest that an ineffective visualisation can compound the difficulties users encounter when metadata is perceived to be poor. For example, several topics performed even worse using the temporal summary than the control approach which used no event visualisation at all. This is likely because the temporal summary's focus on more accurately contextualising the content of events did not effectively translate to reflecting an event's concept relevancy, or the likelihood of it containing concept relevant images. This resulted in users

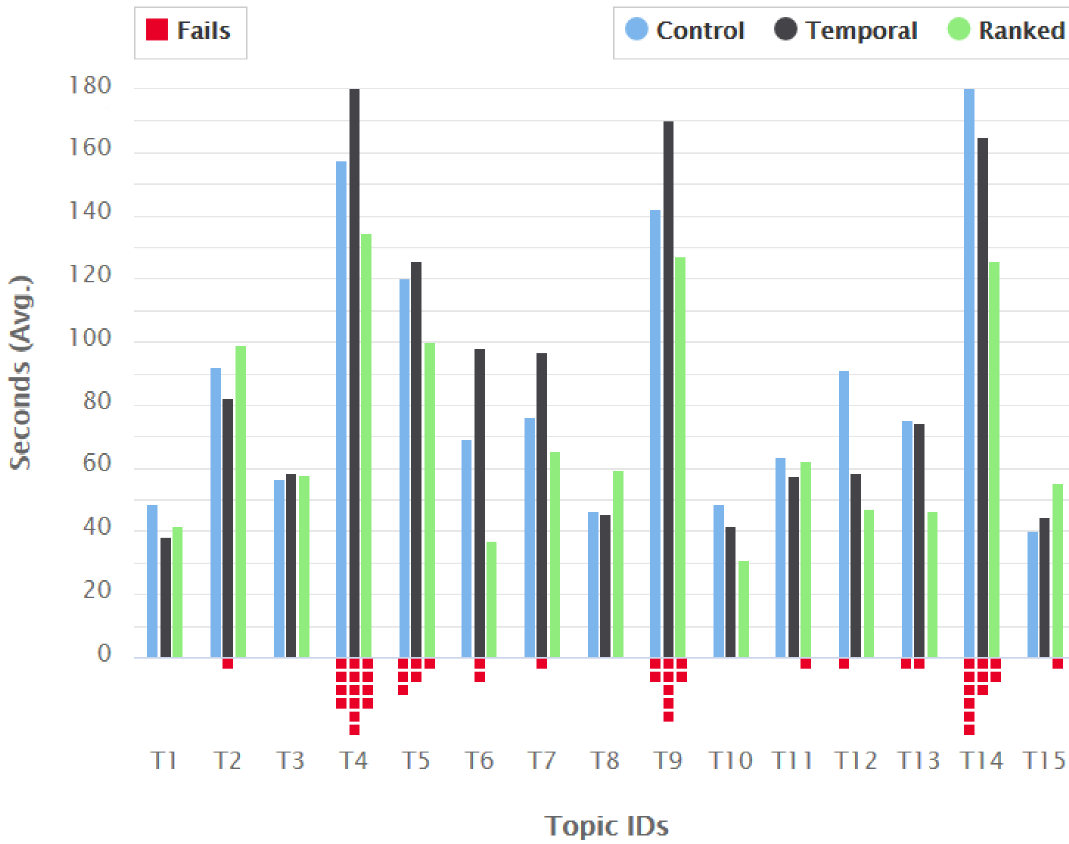


Figure 5.12: Average seconds taken per topic for each visualisation variant (with total failures below axis)

often overlooking valuable events because the summary did not appear to be relevant at a glance. This negatively impacted user performance even worse than the obstacles encountered by repetitive data using the control visualisation approach. It appears that, though the temporal event summary could be effective for other lifelogging criteria such as recollection or reminiscence, where event context could be more valuable, its performance in this study suggests that its usage in lifelog retrieval is poorly optimised.

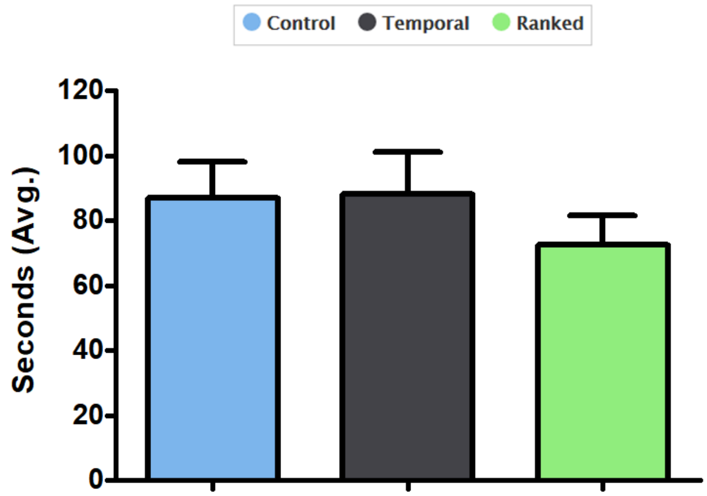


Figure 5.13: Total average seconds for each data visualisation approach with error margins

5.3.2 User Feedback

Similar to our previous experiment, each participant was asked to provide feedback on their experience. However, knowledge gained from our previous study led us to make adjustments in how this feedback was gathered. In our previous experiment users were asked to fill out a questionnaire after each interaction with the test system. This resulted in significantly long experiment times, which stretched the already generous allocation of time volunteers had set aside for experimentation. To address this, the questionnaire for this study was consolidated into a single iteration which would be provided at the very end of the user's experiment. To further increase efficiency, a selection of questions originally reserved for the informal interview were moved to the questionnaire. These questions asked users what they liked about the system, what they disliked about the system, and finally what they suggested might improve the system. Finally, the questionnaire also contained a selection of usability statements which the users needed to state their level of agreement with on a five-point Likert scale, the answers to which are visualised in Figure 5.14. The full questionnaire can be viewed in Appendix A.2.

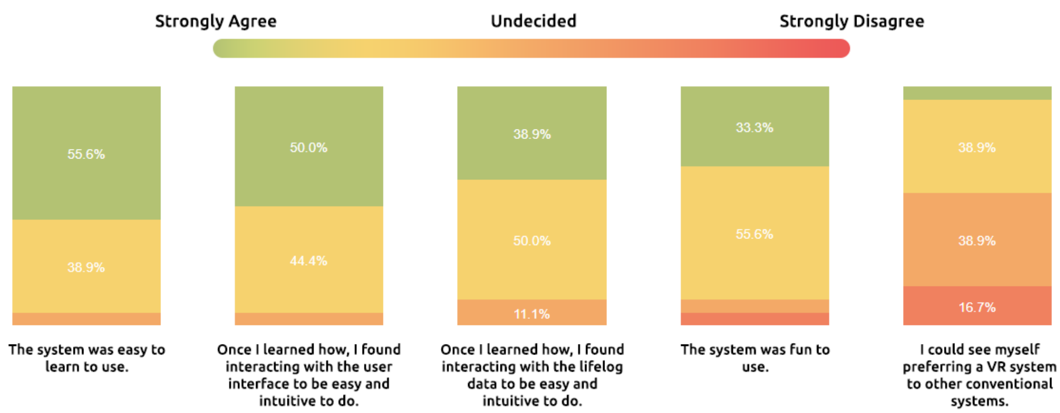


Figure 5.14: User feedback across all three visualisation methods

Gathering user feedback specifically regarding the differences in visualisation approaches proved difficult. From a user's perspective the ranked and temporal event summaries visually appeared very similar except for the presence of star icons indicating an individual image contained a queried concept. However, when asked their order of preference regarding the three visualisation approaches (see Figure 5.15), the users seemed to unanimously choose the ranked method, then the temporal method, followed by the control with no event visualisation. These preferences seemed to be based entirely on the user's perceived success in retrieval rather than any other aspect, such as a poorly understood feature or discomfort with a particular interaction. Similar to the previous study, the majority of feedback from users focused on the availability and accuracy of concepts. This is likely because the concept metadata was the primary data used to generate queries and, when pressed to evaluate the system, this was an obvious critique for users to identify.

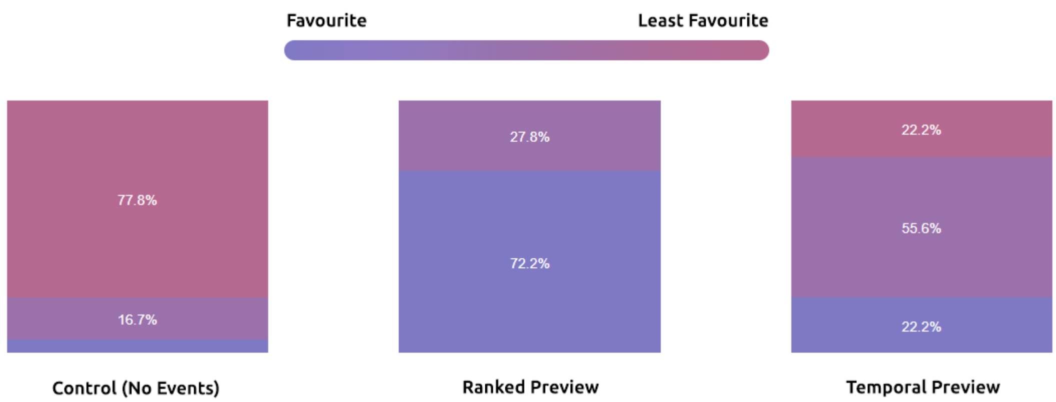


Figure 5.15: Order of preference for each visualisation type

Much of the positive sentiment regarding the system's visualisation focused on the ease by which users could quickly browse and navigate the image data. Many stated that they felt highly immersed in the system and this positively impacted their experience digesting visual information. Users perceived that they would more accurately and efficiently review a large number of images in virtual reality better than they would on a traditional screen, though whether this is actually true remains unclear. With respect to negative sentiment, aside from feedback related to concept accuracy, a significant number of users stated they had difficulties interacting with the context menu. The requirement to aim and hold the trigger while simultaneously using your thumb to choose a contextual option proved unintuitive for many users, especially for those with less experience in virtual reality.

5.3.3 Analysis

As we complete our second user study, our understanding of the issues encountered by a lifelog retrieval system in virtual reality come more into focus. The recurring issues with the accuracy of concepts and the ambiguity associated with their definition is a noted obstacle to successful retrieval. The introduction of visual indicators (stars and ribbons) of a concept score in the ranked event summary was a direct attempt to address some of these issues, and the favourable performance of this visualisation approach suggests there remains opportunity to further clarify concept data without strictly relying on improvements to the data itself. For example, a more thorough breakdown of the event's metadata visualised alongside its summary might better convey its potential relevancy and content. This breakdown might include specifically how many images in the event are labelled with the queried concepts and how many times those specific concepts actually overlapped on individual images. Exposing this information along with improvements to the virtual user interface such as concept categorisation could offset any inaccuracy or confusion associated with the lifelog metadata.

Unlike our previous study, where we determined a hybrid of the experiment variants is likely to be most appropriate, it is clear that for the purposes of lifelog retrieval in this context the ranked event summary is an effective method of visualising events. This disparity in effectiveness may be due to the nature of the known-item search tasks utilised in lifelog retrieval evaluation. For example, in video retrieval, the goal is not an individual frame, but the entire video itself and therefore a broad understanding of the video's content is useful. In this context, we suspect our temporal strategy which utilises the first, last and centermost images could be very effective. In contrast, most lifelog retrieval is based on a goal of an individual image rather than an entire event. In this context, even though conveying a broad understanding of the event can be useful, it is more useful to try and include potentially relevant images in the summary.

We also determined from user feedback that the design of our virtual context menu needs further evaluation in order to be properly utilised by users. Future work could compare a selection of approaches to determine the most intuitive and effective method of achieving this. This could be as simple as reducing the number of simultaneous interactions necessary to navigate the interface, but could require a complete redesign depending on additional feedback. As we have already mentioned in our previous chapter, the relatively small scale nature of this user study, both in scope and number of participants, makes it unwise to make too many assumptions without further iterative studies, however, it should be sufficient in addressing our research question and overall hypothesis.

5.4 Conclusions

In this chapter we have begun our formal investigation into visualising lifelog data in a manner that effectively supports lifelog retrieval in virtual reality. The primary methodology supporting the visualisation of our lifelog data relied on event segmentation, a common lifelogging convention when interacting with continuous image streams. With this as our foundation, we evaluated a series of visualisation techniques for presenting this data to users within a virtual environment. We concluded that the methodology we refer to as the ranked event summary proved the most effective approach to the visualisation of events in terms of user performance and user feedback. This investigation, along with the groundwork conducted in our previous chapter, has established the preliminary research necessary to develop our baseline virtual reality lifelog retrieval prototype and which we will now directly compare to a conventional retrieval analogue to properly address our research questions and validate our proposed hypothesis. This will occur in our next chapter, where we will discuss the development of this conventional analogue and outline the final user study which took place as part of this work.

Chapter 6

Conventional and Virtual Reality Baseline Comparison

In our previous two chapters we have divided our focus between two primary aspects of our virtual reality lifelog retrieval prototype, namely the virtual user interface and the visualisation of the target lifelog data. Future work would continue to refine this prototype through further iterative design but the scope of this research was to establish a baseline virtual reality lifelog retrieval system for evaluation. In this chapter we will integrate the knowledge gained from our previous user studies and perform a comparative analysis of our virtual reality prototype with a conventional analogue. This is intended to directly address our three research questions and contribute to the overall validation of our hypothesis.

6.1 Conventional Analogue for Lifelog Retrieval

It should be emphasised that accurately comparing two systems designed for fundamentally different technologies is inherently difficult. As was outlined in our lifelog design framework, an effective lifelog system must recognise the intrinsic relationship between the lifelog data, the corresponding criteria, and the technology being targeted. Even if two systems contain the same data and target the same criteria, such as lifelog retrieval, if they are designed for two different technologies, they remain fundamentally different and will address separate use cases. The contrast between these use cases can vary depending on whether the dissimilarities are minor (e.g. tablet and smartphone) or major (e.g. desktop and virtual reality). However, this does not mean that comparing a virtual reality and a conventional technology can not be insightful. Such a comparison enables us to investigate the potential applications of lifelog retrieval in virtual reality and how it compares to applications provided by more established alternatives.

However, determining an adequate candidate to represent the primary use cases provided by conventional lifelog systems is not a simple task. State-of-the-art systems targeting lifelog retrieval are not plentiful and many recently published systems (Schoeffmann et al., 2018; Lokoč, Souček, and Kovalčik, 2018) are often just generic multimedia information retrieval systems which have been repurposed for lifelog interaction. As a consequence, it was necessary to develop our own conventional lifelog retrieval system to serve as a baseline for comparison. The benefit of this was that we could rapidly prototype a conventional system based on the same back-end architecture used by our virtual reality prototype, which afforded us a lot more flexibility during our user study than if we had to rely on a third-party system. It is important to emphasis that for this comparative analysis, both systems utilised the same event ranking algorithm, described in our previous chapter, and visualised events using the ranked summary methodology we established in our previous chapter.

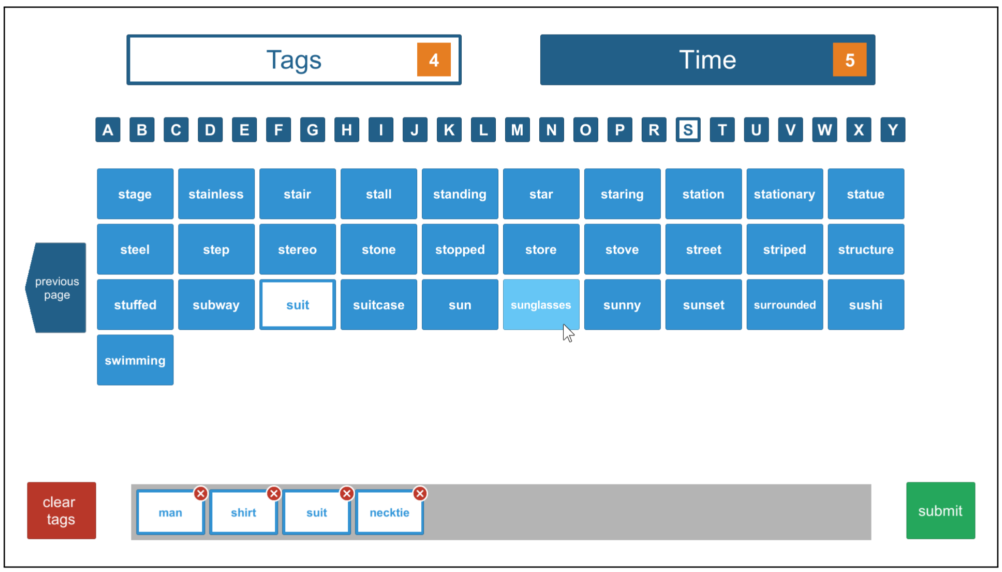


Figure 6.1: Concept selection on conventional lifelog retrieval system

The most common technology used for lifelog retrieval is undoubtedly a personal computer, most often a desktop or laptop. This is likely because of the device's ubiquity and the familiarity most users have in its operation. As a result we targeted this platform, specifically a two-dimensional screen paired with a mouse and keyboard, as the basis for our conventional analogue. As previously mentioned, the system shared the same back-end architecture, and therefore provided the same set of features, as our virtual reality lifelog prototype. The primary difference, aside from the obvious increase in accessible spatial dimensions, is how the user interacted with these features. In Figure 6.1 we can see that the overall design of the primary interface remains unchanged, maintaining the same layout and colours, and now occupying the entirety of the visual space when active. Instead of pointing a controller, the user simply points their mouse and can interact with interface elements by left-clicking.

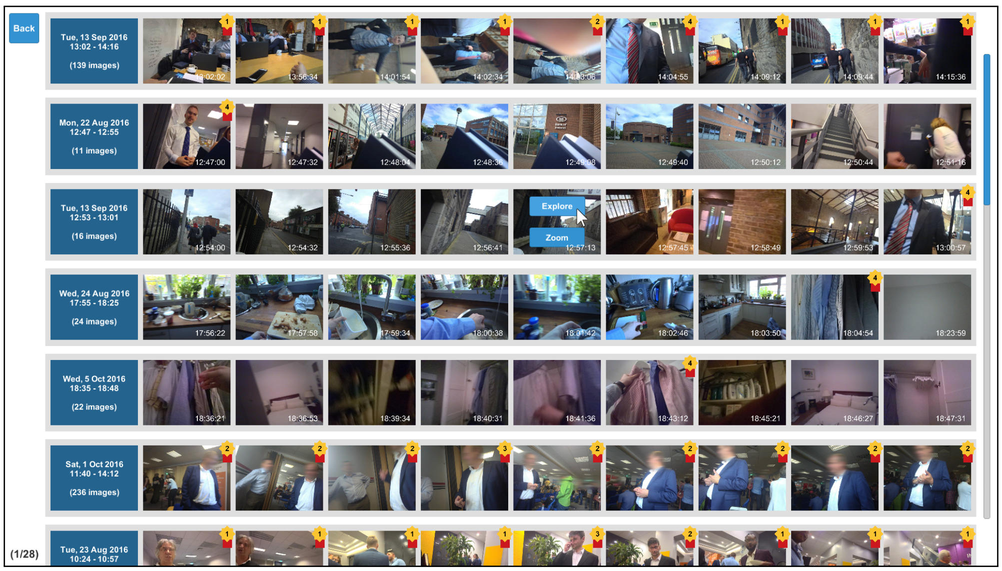


Figure 6.2: Event summaries on conventional lifelog retrieval system

The major differences in transitioning from virtual reality to a more conventional medium is in the visualisation and presentation of the lifelog data itself. In a virtual environment the data must be arranged with careful consideration regarding the user's position, as they occupy the same space as the data being explored. This means navigating data from the traditional perspective of top to bottom is poorly suited as it forces the user to crane their neck and can even lead to vertigo when data is presented too far below them. This was addressed in our virtual reality prototype by having the user scroll through data horizontally rather than vertically. However, as we move back to traditional media, we must revise this methodology in favour of common practice, which also has implications regarding the specific arrangement of the data as it is being navigated. In Figure 6.2 we can observe a set of results which have been returned after the user submits a query. Note that each line of images represents an event summary, identical to the ones visualised in virtual reality, but instead of being aligned in a 3x3 grid, they are aligned in a horizontal line to better exploit the visual space afforded by computer screens.

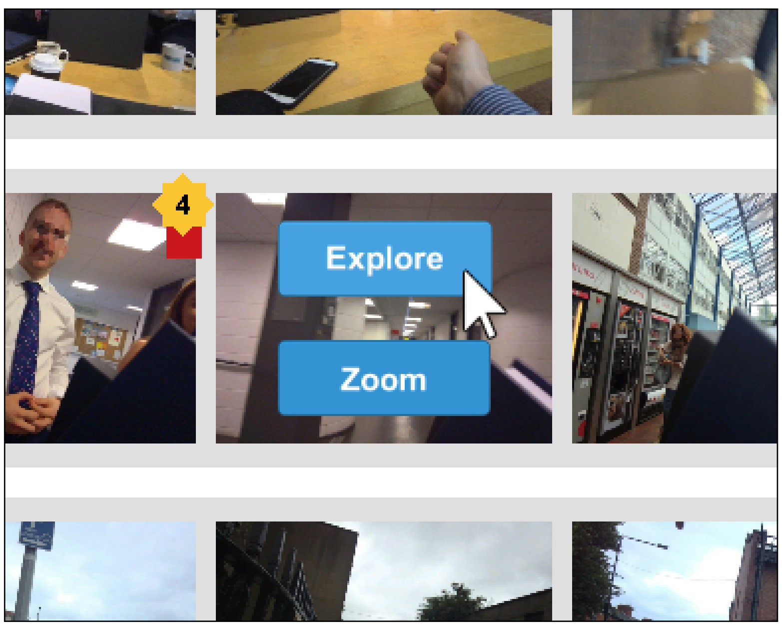


Figure 6.3: Context menu on conventional lifelog retrieval system

The introduction of a contextual menu to support interactions with the lifelog data inside a virtual environment dictated many design considerations to ensure clarity and ease of use, which remained unintuitive for some users. If we recall in our previous chapter, a number of users stated they found the interactions necessary to fully utilise the contextual interface cumbersome. In our conventional medium this process is greatly simplified, where users only have to hover their cursor over a relevant interface element and a set of corresponding options are exposed. This can be observed in Figure 6.3 where we see a user hovering their cursor over an image to reveal two contextual options to zoom or explore. These options provide identical features to our virtual reality prototype but once again the presentation of the data must adapt to the conventional nature of our media. When a user explores an event, instead of the images being arranged in a single row, they are arranged in a grid, ordered temporally from top-left to bottom-right. This can be seen in Figure 6.4 where all the images of an explored event are presented on the screen. To preserve context when exploring a large event, the application will automatically scroll through the images in an event until it reaches the image the user had clicked on from the event summary. This image is highlighted with a blue border and from there the user can scroll up to see past images, or continue scrolling down to see future images.

The remaining features of our lifelog retrieval prototype, 'zoom' and 'search tags', work in much the same way they do in virtual reality. In Figure 6.5 we can observe the zoomed state of an individual image alongside its respective metadata. Though this function serves the same purpose as it does in virtual reality, the feature itself becomes a lot more important on conventional media. Unlike in a

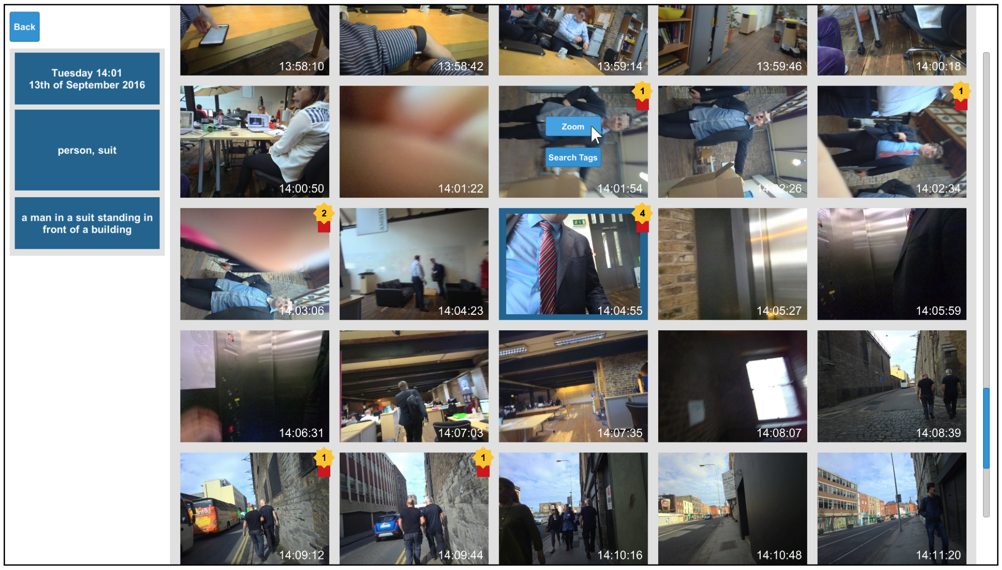


Figure 6.4: Event exploration on conventional lifelog retrieval system

virtual environment, where to get a closer look a user can literally step forward and examine an object at a shorter distance, getting closer to a two-dimensional screen most often does not afford enhanced clarity or resolution.

6.2 Virtual Reality Lifelog Retrieval

As we have established it was outside the scope of this work to continuously iterate, there were no major adjustments made to the virtual reality prototype from that described in our previous chapters. The majority of feedback from previous user studies focused on the accuracy of lifelog concepts, in terms of labelling accuracy (if the concept actually related to the image) but also in terms of definition accuracy (if the concept meant what the user thought it meant). Attending to these issues is not a straightforward task, and it is likely that improvements to the

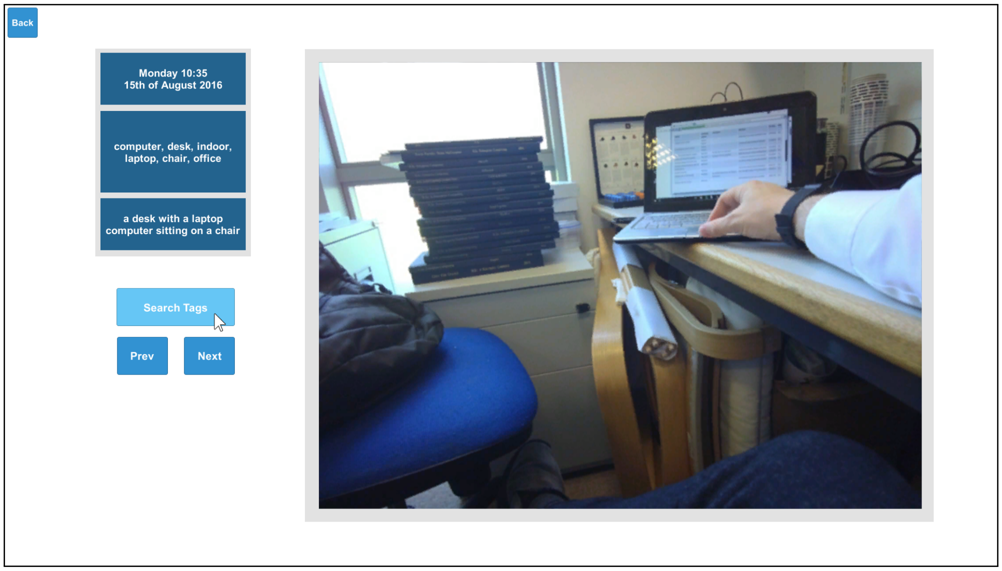


Figure 6.5: Image zoom on conventional lifelog retrieval system

underlying analytics which produce the lifelog concepts would provide the most benefit, but this has no impact on our research. However, as we have mentioned in previous chapters, this does not mean that enhancements to the user-facing lifelog retrieval system could not reduce the severity of this issue. One of the methods proposed to achieve this was the categorisation of concepts under semantic headings such as 'exterior' and 'interior' or 'food' and 'drink'. However, implementing this type of categorisation is also not straightforward and would require its own evaluation to determine what headings successfully conveyed their content, and how granular these categories should be. If this was done incorrectly, then the selection of concepts could become even more unintuitive. As a result of this, for the purposes of the final comparative analysis, the core design of the virtual user interface remained unchanged.

However, it was still necessary to determine which of the previous user study system variants would persist during the experiment. Though we still maintained from our conclusion in chapter four that a hybrid approach was likely most appropriate, the necessity to perform precise interactions such as typing was yet to be introduced to the system, therefore the 'billboard' style of interaction remained the primary methodology used to interact with the virtual interface. However, it should be noted that this methodology was slightly modified to enable users to adjust the position of the interface within the environment to their preference. With respect to the visualisation of event summaries, the unanimous preference of users for the ranked event summary made it a clear choice. It would also have been preferable to expand on the methodology used to convey an event's concepts or develop a new strategy to better reflect an event's content and relevance, but this was reserved for future work. Although it should be noted that the addition of these features would provide equal benefit to both the virtual and conventional prototypes so the scope of identifying design baselines and investigating our hypothesis should not be affected by the exclusion of these features.

6.3 Comparative Study

Our comparative analysis of a virtual and conventional lifelog retrieval system maintains the same experiment setup as our previous two user studies with some minor adjustments. For this experiment 16 volunteer participants were recruited and 16 known-item search topics were gathered for retrieval. As was the case with our previous study, approximately half of the volunteers were past participants so this meant we once again had to request new topics from the lifelogger who had generated the NTCIR-13 dataset. This resulted in a further increase to topic difficulty as the lifelogger found it difficult to generate topic descriptions that were sufficiently different from previous topics while still being possible to retrieve. Another necessary adjustment for this experiment was the configuration methodology which was utilised to account for the learning bias across all users during the study. If we followed our previous method, we would divide our 16 known-item search topics by our 2 system variants to produce 2 groups of 8 topics and then arrange these deliberately to ensure each group of topics was performed first and last an even number of times, with topics inside each group arranged in a random order. However, the number of topics in these two groups was determined to be too large and, despite the randomisation of the topic order within these groups, the number of participants would be insufficient to prevent the introduction of a learning bias.

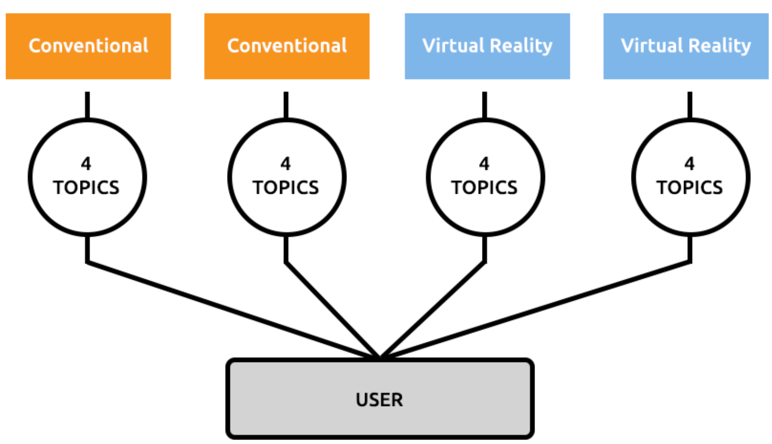


Figure 6.6: Example configuration for one participant in the user study

To address this issue, we modified our previous configuration methodology by dividing our total number of topics by twice the number of system variants. This produced 4 groups of 4 topics (see Figure 6.6) instead of 2 groups of 8 which provided us with more control in arranging the topics to prevent the introduction of a notable bias across all users. This configuration can be clearly seen in Figure 6.7 where we can observe each of the four topic groups coloured-coded and labelled A, B, C and D.

USER	CONVENTIONAL	CONVENTIONAL	VIRTUAL REALITY	VIRTUAL REALITY
1	A	B	C	D
2	B	A	D	C
3	C	D	A	B
4	D	C	B	A
5	C	A	D	B
6	B	C	A	D
7	D	B	C	A
8	A	D	B	C

USER	VIRTUAL REALITY	VIRTUAL REALITY	CONVENTIONAL	CONVENTIONAL
9	A	B	C	D
10	B	A	D	C
11	C	D	A	B
12	D	C	B	A
13	C	A	D	B
14	B	C	A	D
15	D	B	C	A
16	A	D	B	C

Figure 6.7: Modified user configuration for comparative analysis user study

At the top of the user configuration tables we can see that the first 8 users attempted to retrieve their topics using the conventional prototype for the first two topic groups, and then using the virtual reality prototype for the remaining topic groups, whereas the last 8 users did this in reverse. We can also see that each topic group was arranged so the four topics it contained were encountered evenly throughout out all the users during the experiment. This prevented any single topic from appearing unevenly at the start or end of the users' retrieval tasks, which would have resulted in additional potential for bias. The list of topics contained in each topic group are outlined in Tables 6.1, 6.2, 6.3 and 6.4

where we can observe the topic ID, title and accompanying descriptions which were provided to users before beginning each retrieval task.

ID	Title	Description
1	Microwave Timer	Find the moment where the lifelogger is setting the timer on a microwave
2	Garden Shears	Find a moment where the lifelogger is using garden shears
3	Restaurant Photograph	Find a moment where the lifelogger is taking a photo of a person and their food in a restaurant
4	Striped Shirt	Find a moment where the lifelogger is holding up a red striped shirt at home after work

Table 6.1: Comparative Analysis Study - Topic Group A

ID	Title	Description
5	Refrigerator	Find a moment when the lifelogger is taking milk from the refrigerator
6	Guitars	Find a moment where the lifelogger is looking at multiple guitars in their home
7	House Number	Find a moment where the lifelogger enters a home where the house number is clearly visible
8	Reading a Menu	Find a moment where the lifelogger is looking at a menu on a week day

Table 6.2: Comparative Analysis Study - Topic Group B

6.3.1 Technical Background

Before participating in the experiment, each user was asked a set of background questions to gauge their technical experience with computers as well as any ex-

ID	Title	Description
9	Coffee Machine	Find a moment where the lifelogger is operating a coffee making machine
10	Breakfast and Television	Find a moment where the lifelogger is having breakfast while watching television
11	50th Birthday Party	Find the moment where the lifelogger is attending a 50th birthday party
12	Self-Service Checkout	Find the moment where the lifelogger was using a self-service checkout at a grocery store on a Thursday evening

Table 6.3: Comparative Analysis Study - Topic Group C

perience they had using a virtual reality platform. The full questionnaire can be viewed in Appendix A.4. While these same questions were also asked in our previous user studies to identify any individuals who had especially poor technical or language skills, here the answers to these questions are especially important in properly contextualising our comparative analysis. It was clear before recruiting the participants that many of them would have significant experience using computers and comparatively less experience using virtual reality, and this is reflected in the nature of the questions that were asked. In Figure 6.8 we can observe that the participants were asked how often they use a computer in their average week and almost all of the participants stated they used computers for over 30 hours on average, with the majority of them stating it was over 40 hours. In contrast, with respect to virtual reality, participants were not asked about their average time, but rather the total number of times they had used a virtual reality platform in their entire lifetime. In Figure 6.9 we can observe that almost every user had

ID	Title	Description
13	Fresh Herbs	Find a moment where the lifelogger is looking at fresh herbs in their home
14	Football in Bed	Find a moment where the lifelogger is watching football on a bed
15	Driving for Groceries	Find a moment where the lifelogger was driving their car to a grocery store
16	Walking in the Rain	Find a moment where the lifelogger is walking outside on a wet weekend morning

Table 6.4: Comparative Analysis Study - Topic Group D

some previous experience with virtual reality but it was very limited. Only four users stated they had used a virtual reality platform more than ten times in their life, and only one user stated they actually owned a virtual reality headset.

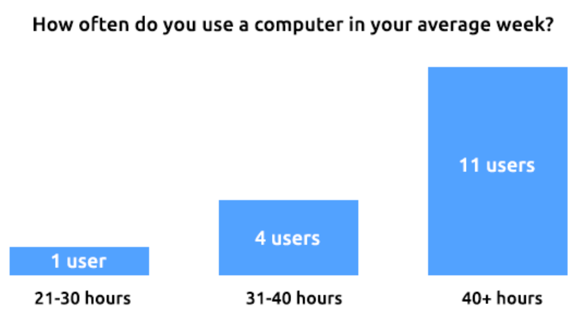


Figure 6.8: User weekly exposure to computers

It is clear from these results that the majority of volunteers recruited for this study used computers very regularly throughout their week. Though it may have been valuable to perform a more thorough analysis of their technical abilities, the primary goal was to establish their baseline experience performing general com-

Figure 6.9: User previous experience with virtual reality

puter tasks and familiarity with digital user interfaces. The stark contrast in experience using computers over virtual reality was expected and likely contributed to both positive and negative outcomes during the user study. For example, the fact that every user had at least some technical experience meant that transitioning onto a virtual reality platform wasn't completely unfamiliar. However, the fact that so many of the users had so little experience with virtual reality, and so much experience with computers, means there will be an unavoidable bias when trying to use each system effectively. However, since this bias is in favour of the conventional platform, we can assume that as people become more familiar with virtual reality platforms, the potential effectiveness of our baseline prototype can only improve.

6.3.2 User Performance

In our previous user studies, the length of time taken to complete each known-item search task was recorded, and while this was sufficient to infer the relative performance of the experiment variants, the recorded results did not capture the ratio of querying time versus browsing time. Note that we define querying time as any time spent selecting concepts or times using the prototype's primary interface, and browsing time as any time spent navigating the visualised lifelog data. Based on the knowledge gained from this previous work, we determined it may be beneficial for our comparative analysis to record this ratio, as the nature of different technologies can impact the parts of a system in different ways. For example, interactions on a personal computer might afford faster querying, but interactions in virtual reality might afford faster browsing of the data. In addition to this, we also recorded the average number of queries submitted by the users on each system for each topic (which we refer to as retrieval attempts), as well as the total number of times the users exceeded the 180 second time limit (which we refer to as a retrieval failure). After all the users had participated in the experiment, each topic had been searched for a total of 16 times, 8 times on each of the 2 retrieval prototypes.

We have visualised these results in Figure 6.10 where we can observe a bar chart displaying the average seconds taken querying and browsing per topic for each prototype. The 16 topics are labelled on the horizontal axis and the average time in seconds is labelled on the vertical axis. The lifelog retrieval prototypes are represented by two coloured bars, each shaded in dark and bright to represent querying and browsing respectively. The red squares beneath the chart's horizontal axis indicate the total number of retrieval failures per topic which occurred using each prototype. The total average seconds over all topics with error margins is shown in Figure 6.13.

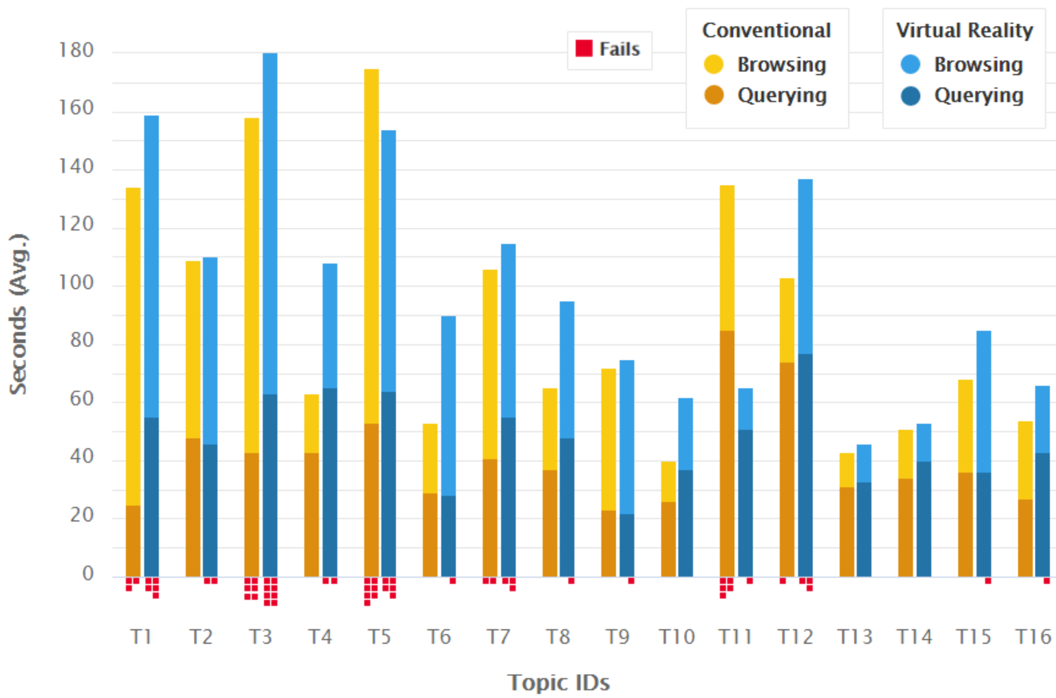


Figure 6.10: Stacked average seconds taken querying and browsing per topic for each prototype (with total failures below axis)

With a few exceptions, it is clear that the conventional prototype proved most effective in terms of total retrieval time, though in several instances its effectiveness over virtual reality was negligible. Furthermore, though we can see a notable contrast in retrieval time in favour of the conventional prototype across a number of topics, such as T4 and T6, the largest contrast across all the topics was in favour of virtual reality on T11. This topic also resulted in the largest contrast of failed retrievals, with the virtual reality system failing only once and the conventional system failing a total of five times. It is difficult to speculate on the reason for this outlier, but one theory is that the topic described the lifelogger entering

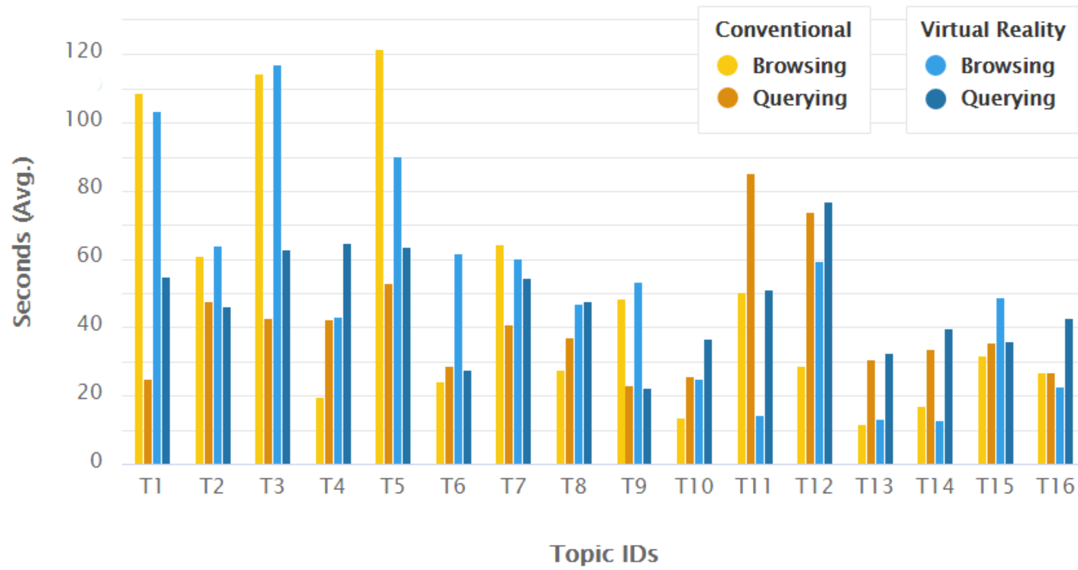


Figure 6.11: Unstacked average seconds taken querying and browsing per topic for each prototype

a home where the house number is clearly visible, and the virtual reality system afforded a better visualisation of the data to detect this relatively small detail in the images. Similar to our previous two studies, a small selection of topics proved exceedingly difficult for users irrespective of what prototype they were using. This is clearly evident on T1, T3 and T5 which resulted in the highest average retrieval times and the most retrieval failures, which were comparable on both prototypes.

With respect to the ratio of querying time versus browsing time, evidence suggests that on average more difficult topics resulted in longer browsing times and easier topics resulted in shorter browsing times. This is most notable in topics T1, T3 and T5 which many users found difficult and resulted in comparably long browsing times, and in topics T13, T14 and T16 which many users found easy

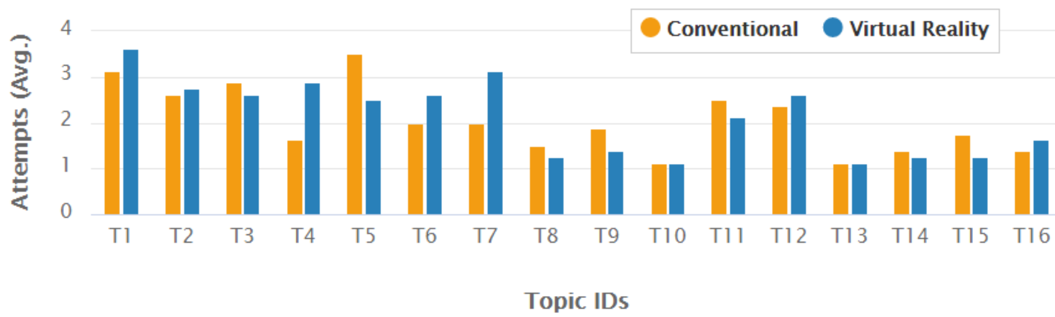


Figure 6.12: Average retrieval attempts per topic for each prototype

and resulted in comparably short browsing times. This can be seen more clearly in the unstacked version of the bar chart in Figure 6.11. The source of this trend likely pertains to the users' opposition to adjusting their query after they have committed to a specific set of lifelog concepts. Only after browsing a significant number of results, do they concede their query is not retrieving what they intended and consider other options.

In Figure 6.12 we can observe the average number of times each user submitted a query for each topic, which we defined as retrieval attempts. As one might expect, there is a strong correlation between the number of retrieval attempts and the length of time taken to complete a retrieval task, however this is not always the case. For example, we can see in T8, T9 and T15 that even though virtual reality resulted in longer average retrieval times, it also resulted in fewer average retrieval attempts. These outliers suggest that the previously stated aversion users have to attempting successive queries might be compounded more on the virtual reality system than on the conventional system.

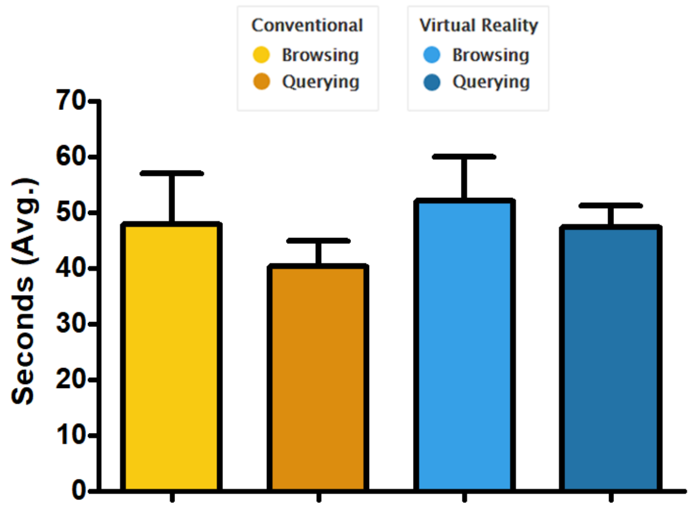


Figure 6.13: Total average seconds for querying and browsing on each prototype with error margins

6.3.3 User Feedback

Upon completion of their tasks, each participant was asked to fill out a questionnaire detailing their experiences during the experiment. The questionnaire contained usability statements which the users needed to state their level of agreement with on a five-point Likert scale; the answers to which are visualised in Figure 6.14 and Figure 6.15. In addition to an informal interview regarding their general experience with each prototype system, the questionnaire also contained three open questions to ensure no important feedback was omitted. These questions asked users what they liked about the system, what they disliked about the system, and finally what they suggested might improve the system.

General sentiment regarding the prototype systems was mixed but mostly positive. Both systems where perceived to be intuitive, easy to use, and even fun, by most participants. However it is interesting to note that, despite its improved performance, feedback toward the conventional prototype was slightly more negative than its virtual counterpart. When asked their agreement with respect to our usability statements, almost no user responded negatively in relation to virtual reality, whereas several users responded negatively in relation to the conventional system, with as many as 20% disagreeing that it was fun to use. This is likely related to the novelty factor associated with interacting with something unfamiliar (virtual reality) versus something very familiar (a desktop computer).

Conventional Prototype

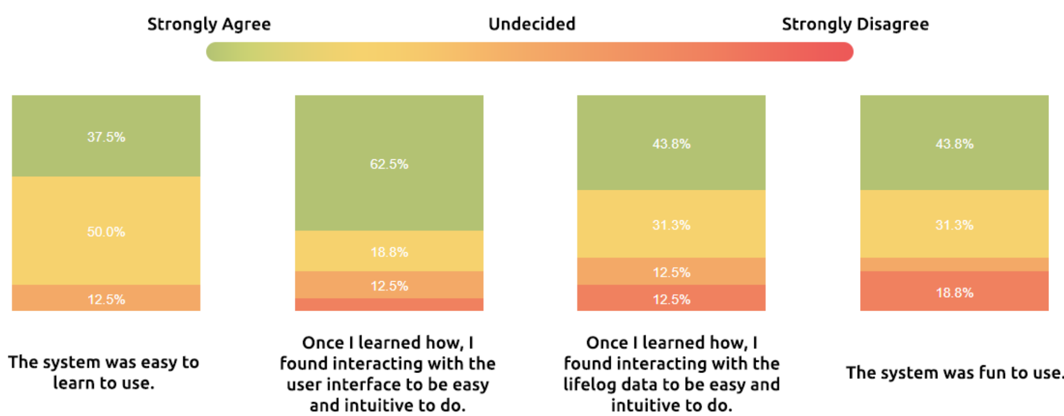


Figure 6.14: User feedback for conventional retrieval prototype

Positive feedback regarding the conventional prototype centred around its familiar user interface and style of interaction. The majority of users required very little instruction prior to interacting with the system and began discovering features naturally without being prompted. There was notably mixed feedback regarding the exploration of lifelog data on the conventional system, as some users felt it utilised its visual space very efficiently whereas others felt it was too cluttered and inhibited retrieval. Related sentiment was also evident during testing, where some users seemed to initially respond well to interacting with the data on the conventional system, but then determined that the presentation of the data was causing them to overlook important information. For example, despite their awareness that highly ranked data is presented at the top of the list, the users' instinct to immediately use the scroll wheel to move down the page often resulted in them scrolling past relevant results.

Virtual Reality Prototype

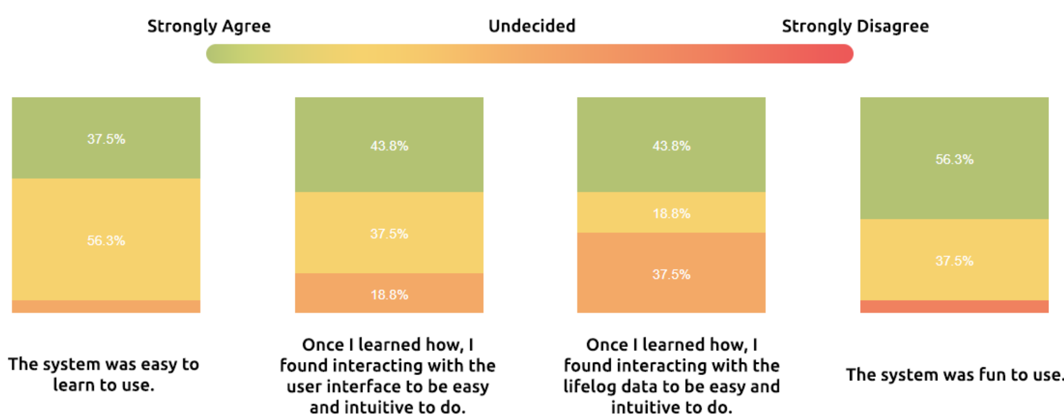


Figure 6.15: User feedback for virtual reality retrieval prototype

Positive feedback regarding the virtual reality prototype centred around the immersive experience and the novelty of exploring a virtual environment. Compared to the conventional system, it took noticeably more time for users to become comfortable with their surroundings and it required far more prompting and instruction to discover all of the virtual reality system's features. However, despite the steeper learning curve, many users stated they adapted quickly once they grew accustomed to the environment and enjoyed having more direct control of the data during search (i.e. pulling and pushing the data, or tossing it left and right as they searched). Some users also stated that the visualisation of data at various depths felt more natural than the 'forward and back nature' of conventional interfaces and provided better context while supporting the consumption of more information at once. The only explicitly negative sentiment regarding the virtual reality system was in relation to the context menu which enabled users to explore and zoom on the retrieved lifelog data. Similar to our previous experiment, many users stated it required too many synchronised interactions and therefore was not intuitive.

General

Most of the general feedback regarding the prototype systems was unchanged from our previous studies, such as the request for new concepts and improved accuracy on existing ones. However, it should be noted that this sentiment was much less prevalent during this user study and there is evidence to suggest that further additions to the user interface could reduce this even further. For example, the inclusion of a feature to search for concepts via an open input instead of exclusively from a list of words, which was another piece of feedback provided by a number of users. Finally, after each participant had finished their retrieval tasks on both prototypes, they were asked which system they preferred overall. Surprisingly, despite their increased familiarity with the platform alongside its slightly improved performance, only six out of the sixteen participants stated they preferred the conventional prototype overall, with the remaining ten stating they preferred virtual reality.

6.3.4 Analysis

Within the scope of this comparative analysis, there appears to be no outright superior approach for lifelog retrieval. Whilst the conventional prototype proved better in terms of performance, the benefits were not notable. Furthermore, the conventional system had an additional advantage in that every user had notable experience interacting with similar conventional systems, and far less experience performing interactions in virtual reality. This is not to suggest that the conventional system developed for this study is wholly representative of what is possible with conventional media. We have acknowledged that improvements could be made to the conventional prototype through further, and larger, iterative studies, but this is also true for our virtual reality system. With this study we have examined some of the most common conventions implemented in the field of lifelog retrieval and presented them in contrast to novel methods proposed for virtual reality to evaluate their baseline effectiveness.

By comparing these two systems, we have determined that a primary obstacle for users when transitioning from conventional media to virtual reality to perform lifelog retrieval is the interactions necessary to formulate the retrieval query itself. Though proprioceptive interactions such as picking up a virtual ball and throwing it are highly intuitive, even when it must be performed with an unfamiliar input device, interacting with a virtual interface is often less intuitive, as the user must identify how their input device connects to the interface's content. This issue can be mitigated by mimicking proprioceptive interactions like pointing or throwing, but the inability to physically connect with virtual objects remains a limiting factor. This same obstacle extends to interacting with the retrieved lifelog data within the virtual environment, as was identified by users when they described the prototype virtual context menu as unintuitive. However, despite these issues, exploring the retrieved lifelog data within the virtual environment proved to be a very positive, and highly enjoyable, experience for many users. This did not directly result in reduced retrieval time but it is likely this relates to users lack of experience with the technology. However, it should also be acknowledged that this lack of experience may directly relate to the user's enjoyment of the technology, as it is possible that the pleasurable experience of interacting with lifelog data in a virtual environment could diminish as users become more familiar with the technology and the novelty is less prevalent.

6.4 Comparative Analysis with State-of-the-Art

Though we have completed our analysis and can begin to directly address our research questions regarding the comparison of virtual reality and conventional lifelog retrieval applications, we must first acknowledge another evaluation of our virtual reality prototype which occurred during the course of this research. In chapter two we introduced a number of state-of-the-art lifelog retrieval systems which were published in accordance with their participation in the Lifelog Search Challenge (LSC) (Gurrin, Schoeffmann, et al., 2019). Also participating in this challenge was a version of our virtual reality lifelog retrieval system which was adjusted to target a subset of the NTCIR-13 test collection which served as the competition's target datatset. Though our own user studies and comparative analysis have been useful in our investigation of virtual reality as means to support lifelog retrieval, the insights gained from comparing our prototype to five different state-of-the-art conventional lifelog retrieval systems are especially relevant. Though it would have been preferable to participate in a number of these such challenges during the investigation of our hypothesis, unfortunately this was the first time such an event had been organised, and it did not occur a second time until after this research was concluded.

As we described in chapter two, the structure of the LSC was based on the participation of both expert and novice users in a real-time competitive lifelog retrieval challenge, where the experts were typically the system developers themselves and the novices were audience members recruited from the conference who were not familiar with any internal details of the competing systems. There were 18 retrieval tasks in total, where the expert users performed 6 tasks, and the novice users performed 12 tasks on each of the participating systems. The competition organisers made the decision to have more novice tasks because they

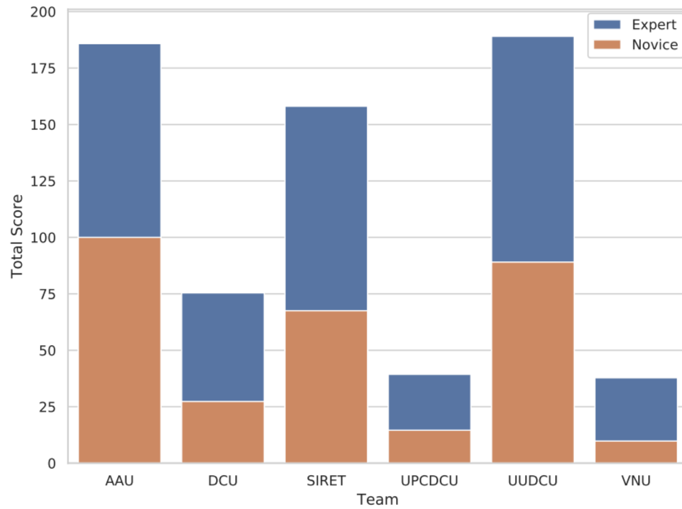


Figure 6.16: Overall (normalised) score for both novice and expert users in the Lifelog Search Challenge

stated it better reflected the goal of the LSC which was to promote research into user-friendly lifelog systems. A system's overall performance in the LSC was determined by a ranking formula based on the speed of retrieval and the number of incorrect submissions for each task, and was a composite of the results from both the expert and novice users (Gurrin, Schoeffmann, et al., 2019). In Figure 6.16 we can observe that the best performing system in the LSC was our virtual reality prototype (Duane and Huerst, 2018), referred to as UU-DCU, with the liveExplore system (Schoeffmann et al., 2018), referred to as AAU, ranking second and the SIRET system (Lokoč, Souček, and Kovalčik, 2018) ranking third. All of the participating systems' correct submission times for each task can be observed in Figure 6.17.

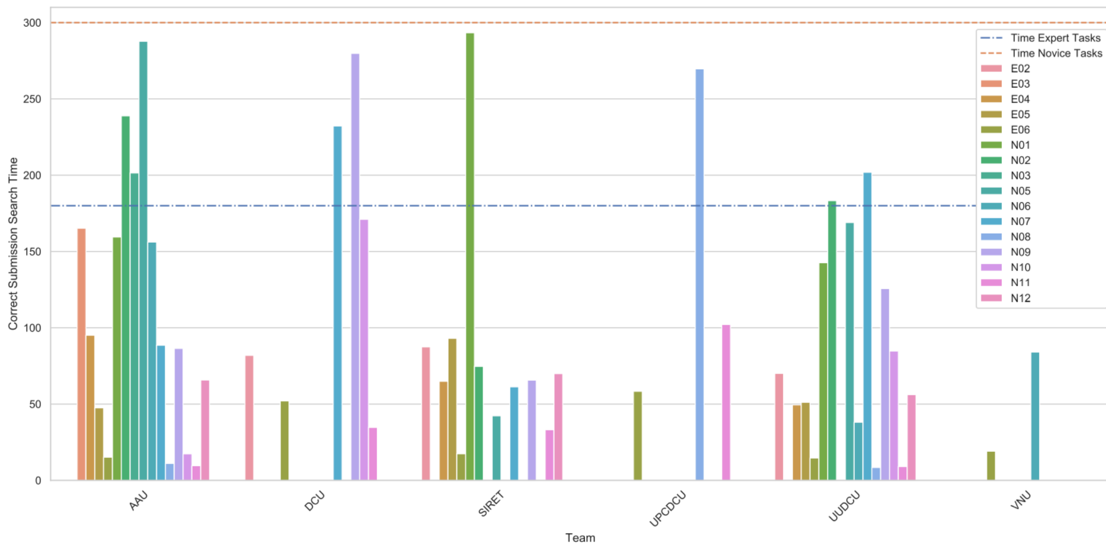


Figure 6.17: Correct submission times for both novice and expert users in the Lifelog Search Challenge

Among the six participating teams at the LSC, the most common prototype features consisted of facet filtering, novel ranked list visualisations and event/scene organisation. As these systems represent the current state-of-the-art, we can surmise that these are some of the most effective methods of supporting lifelog retrieval. In addition to this, the lifeXplore (AAU) system and the SIRET system, which ranked second and third, contained even more features such as non-textual/faceted querying mechanisms and enhanced visual analytics. In this context enhanced visual analytics refers to any content analysis other than what was already provided by the competition dataset. Yet it is clear that simply including these features does not guarantee an effective lifelog retrieval system. As we can observe in Figure 6.18, the majority of the LSC systems contained features which were not available on our virtual reality prototype, yet our system (UU-DCU) performed better overall. Furthermore, the other two top performing systems were based on pre-existing retrieval systems which had undergone significantly more development. This may suggest virtual reality provides a notably better platform for lifelog retrieval compared to the conventional state-of-the-art, but it is also likely that this reflects the inherent shortcoming of state-of-the-art lifelog retrieval systems which more often focus on improving non-interactive aspects of lifelog retrieval and do not prioritise enhancing the relationship between the user and the retrieval system itself. We believe this reinforces the necessity for lifelog application designers to utilise a design framework such as what was introduced previously in this work.

Feature	AAU	DCU	UPC-DCU	UU-DCU	VNUHCM-US	VIRET
Facet Filters	Y	Y	Y	Y	Y	Y
Event/Scene Organisation	Y	Y	N	N	Y	Y
Visual Clustering	Y	N	N	N	Y	N
Novel Ranked List Visualisation	Y	N	Y	Y	N	Y
Enhanced Visual Analytics	Y	N	N	N	Y	Y
Integration of Biometric Data	Y	Y	N	N	Y	N
Non-textual/faceted Querying Mechanism	Y	N	N	N	Y	Y
Based on Existing Video Search Tool	Y	N	N	N	N	Y

Figure 6.18: Summary of the features used by all six systems participating in LSC 2018)

The organisers of the LSC concluded that a good baseline approach for interactive lifelog retrieval was not yet clearly defined, but it appeared as if a well-tested interactive system, placing significant emphasis on the visual element of lifelog data was a good starting point, and that indexing lifelog data will require the development of multimodal lifelog-specific tool-kits to enhance performance beyond a baseline level. From the perspective of the research carried out in this thesis, the result of this challenge suggests that virtual reality has significant potential in terms of its ability to support interactive lifelog retrieval in comparison to conventional media. The organisers of the competition acknowledged that some of the limitations of the LSC included varying technical abilities between participating novice users and the likelihood of expert users having acquired a learning bias from interacting with the target dataset while developing their systems. This occurs because the developers invariably observe a large number of the dataset's images while working on their prototypes, and this background knowledge can prove beneficial when retrieving a known-item search task during competition. However, despite these limitations, the structure of the LSC has enabled us to evaluate our virtual reality lifelog retrieval prototype in comparison to the state-of-the-art and we have shown that it provides comparative overall retrieval performance which directly contributes to the validity of our hypothesis and the baselines which we are establishing in this work.

6.5 Conclusions

In this chapter we have outlined our third major user study which was a comparative analysis of our virtual reality prototype and a conventional analogue. This also included an additional analysis in the form of our prototype's participation in an international competition with five other conventional lifelog retrieval systems. Though the results of our comparative analysis did not reflect virtual reality as an outright superior method of supporting lifelog retrieval, its superior performance compared to state-of-the-art conventional alternatives and its overall preference amongst experiment participants suggest it has significant potential for future work. This is further compounded by the prototype's performance in the LSC, where it ranked first place alongside state-of-the-art conventional lifelog applications. Now that we have outlined all of the insights garnered from the evaluations performed throughout this research, in our final chapter we can finally examine our three research questions and determine the validity of our proposed hypothesis.

Chapter 7

Conclusions

In this thesis we proposed the hypothesis that interactive lifelog retrieval in virtual reality could be as effective as a conventional system in terms of speed and usability while being more immersive, intuitive, and providing increased user satisfaction. To validate this, we determined three primary research questions which we intended to answer through a series of evaluations, primarily in the form of iterative user studies. In this final chapter we directly address these research questions and determine if our hypothesis is upheld.

7.1 Summary

For our first research question, we asked what was an appropriate design framework that integrates lifelog principles, personal data, and the wide variety of available technologies, to support the development and evaluation of interactive lifelog systems. This question was proposed because in order to promote the development of novel interactive lifelog systems, such a framework would be very useful, yet to-date no such framework exists. In chapter three we introduced the Lifelog Design Framework which served as the foundation for the development of our virtual reality prototype. Based on the feedback from users and overall performance of this prototype during research and experimentation, especially when it was compared to conventional state-of-the-art alternatives, we have determined that the Lifelog Design Framework is an appropriate framework to support the development and evaluation of lifelog interaction systems. With its support, we were able to develop a first-generation lifelog prototype on an emerging platform that competed with, and even surpassed, state-of-the-art alternatives. However, we note that since our primary evaluation consisted of only one dataset targeting lifelog retrieval on a virtual reality platform, we cannot confirm its appropriateness outside of these confines without further study, though we believe that it will hold true. Despite this, we hope the intended broader utility of the Lifelog Design Framework will make it more appealing for future researchers.

For our second research question, we asked could a virtual reality system support the generation of lifelog retrieval queries as effectively as a conventional system. This question was asked because we determined a primary component of interactive lifelog retrieval was the ease and effectiveness of generating suitable queries based on the information need. As virtual reality is currently a far less ubiquitous medium for the average user, it was necessary to discern how it might impact the generation of these queries compared to the conventional alternatives. Though we established through our evaluation that virtual reality was capable of generating queries as effectively as a conventional system on a number of occasions, in many instances the conventional analogue seemed to prove more effective. However, this is likely related to the inexperience many of the users stated they had with respect to virtual reality applications. Furthermore, as this is a prototype system built on a first-generation virtual reality platform, it is likely that further advancement in virtual reality technology as a whole along with wider adoption and increased familiarity amongst users will do even more to mitigate this issue.

For our third research question, we asked could a virtual reality system visualise lifelog data in a manner that supports lifelog retrieval as effectively as a conventional system. This question was asked because we determined one of the primary benefits of moving to virtual reality was the ability to visualise data in a highly immersive environment with six degrees of freedom. However, it was uncertain how valuable immersion and a virtual space would be in terms of supporting lifelog retrieval and if it would be worth the transition from conventional media. Through our evaluation we have determined that there is significant potential for supporting lifelog retrieval in virtual reality, with many users throughout this study stating it was their preferred method of exploring the visualised information. Despite the fact the system's overall performance in our direct comparative analysis was slightly inferior to our conventional analogue, its performance in the LSC suggests that it has potential benefits in supporting lifelog retrieval, even to the point of performing better than any of the other state-of-the-art systems. Furthermore, the style of the LSC better adjusts to the learning bias which is unavoidable when testing users who have little experience with virtual reality and significant experience with conventional media, and better reflects the type of novice user whom these systems are ultimately designed for.

Now that we have addressed our three primary research questions, we can discuss the validity of our overarching hypothesis. When we consider the limitations of a first-generation prototype developed for a dramatically new mode of access, it is clear from our evaluation that there is significant potential for virtual reality to support lifelog retrieval in a method at least equally as effective as a conventional alternative as shown by our experimentation. Though the overall performance of our prototype was not superior to our conventional analogue, neither was it sufficiently inferior. Furthermore, as was previously stated, its performance in the LSC better reflects the potential benefit to lifelog retrieval over conventional media. These benefits are further compounded by the positive feedback provided by users and the clear overall preference toward a virtual reality lifelog retrieval system. Given the limitations of this research, which are described in the following section, we consider our proposed hypothesis to be upheld.

7.2 Limitations

Throughout this work we have attempted to highlight any limitations to our evaluation and research where relevant. As we have already discussed, the primary limitations have centred around the configuration of our user studies. As it was not feasible to perform a user study with several hundred users of varying age, sex and technical backgrounds, our user studies must be considered limited in their scope. This is especially true when we consider the technical expertise of the participants in our comparative analysis. Almost every participant stated they operated a computer on average over thirty hours per week, whereas expertise related to virtual reality was so infrequent it had to be measured in individual occurrences. From the perspective of interactive lifelog retrieval, an LSC-style evaluation, with expert and novices users, and multiple competing state-of-the-art applications, is a far more ideal evaluation methodology. Unfortunately such a style of evaluation does not occur frequently and relies on the independent participation of other researchers.

It is also important to acknowledge that the test collection targeted in this work focuses on one lifelogger and multiple users and therefore we cannot say with certainty that our findings will be equally valid when transferred to a real-world application where there is only one user, and they are also the lifelogger. However, we also note that it is standard practice within the research community to evaluate lifelog retrieval using this experimentation method, so despite this limitation, we have still engaged in evaluations using best practice within the lifelog domain.

On the subject of our back-end infrastructure, we must also note that the underlying retrieval engine developed for this work was not representative of the latest techniques supporting ah-hoc information retrieval, however, we do not recognise this as a major limitation as all the comparative systems utilised the same engine and, from participating in the LSC in 2018, we have proven that this engine was shown to perform at least to the standard of other state-of-the-art systems. There is likely significant scope for its enhancement but it is not clear if this would have had any impact on the experiment results.

Another limitation of this work is the emphasis on retrieval focused lifelog applications. Though this is not unusual within the lifelogging community, as retrieval is the most feasible lifelog criteria to evaluate and is implicitly linked to other lifelog application benefits such as recollection and reminiscence, it still remains only one of the five core lifelogging criteria. This issue of an overemphasis on retrieval is also why a primary contribution of this work was a design framework aimed at promoting the development of interactive lifelog systems for all lifelogging criteria, though the scope of this research made it unfeasible to evaluate the framework to its fullest extent. If lifelog applications are to gain broader traction, both inside and outside of academia, there needs to be broader appeal outside of information retrieval and a strong focus on user-friendly interfaces.

On this topic, we must also acknowledge a limitation concerning the hardware component of this study. Though virtual reality is rapidly maturing, to date hardware manufacturers have not established a ubiquitous standard across available devices. For example, some virtual reality platforms must be connected to personal computers, whereas others remain standalone at the cost of performance. Some provide wireless input devices, whereas others provide no input devices and only a heads-up display. As it is difficult to predict the long-term future of virtual reality hardware, including the potential for its convergence with augmented reality, the research outlined in this work has had to focus on establishing baseline design considerations for future virtual reality lifelog applications.

7.3 Future Work

The research carried out in this work aimed to develop a wholly novel lifelog retrieval prototype for virtual reality and set a baseline for the development of future such applications. As a result, there is inherent potential for future work in this area, especially with the addition of more diverse lifelogging datasets and more sophisticated metadata. In this section we will outline potential areas of study which we determine are natural extensions of the research carried out in this work.

7.3.1 Virtual Interface Design to Support Lifelog Applications

If virtual reality lifelog applications are to progress, they must support the generation of lifelog queries as effectively as current conventional alternatives. Though applications based on simulating real-world use cases such as driving a car or flying a plane have obvious benefits when interacted with in virtual reality, the same is not yet true for more traditional application use cases. Future research could devise more effective methods of addressing this in virtual reality. For example, feedback from users throughout this study suggested the addition of a free-text search option would be invaluable in supporting the generation of their lifelog retrieval queries. On conventional media interacting with such a feature would be a relatively simple process, but in virtual reality it would require notably more work in terms of its design and implementation. In the absence of a physical keyboard, a virtual alternative would need to be designed and evaluated, based on aspects such as its layout, shape, size and tactile feedback. It should be acknowledged that these challenges for virtual reality platforms are not unique to lifelog applications, but for virtual reality to evolve as a viable method of supporting lifelog applications, these challenges must be addressed in the context of the lifelog design criteria defined in this work.

Furthermore, as we have yet to observe virtual reality systems becoming as ubiquitous as personal computers or smartphones, it is difficult to predict what interactions will become standard within virtual interface design. Though we can maintain the traditional desktop metaphor which has existed on personal computers for decades consisting of windows, files, folders and tabs, it is also possible a completely different approach will become commonplace in virtual reality design. For example, other feedback from participants throughout this study have suggested the introduction of voice commands as a replacement for traditional interactive methods. Future research could evaluate the feasibility of this in the context of supporting lifelog applications. The introduction of such interactions would not invalidate the baseline methods of interaction established in this work as the pervasiveness and application of voice interactions has yet to be fully realised and many modern users rely on more conventional interactions as a fallback when voice interactions are not appropriate.

7.3.2 Lifelog Data Visualisation in Virtual Reality

The most promising feedback provided by users throughout this study has focused on the visualisation of the lifelog data in a virtual environment. In this work we have identified that inhabiting the same space as the lifelog data you are investigating is preferable for many users and provides a level of immersion that is impossible with conventional alternatives. Our research has introduced a set of visualisation methodologies based on conventional lifelog applications but the three-dimensional nature of virtual reality has enormous potential for future work which builds on these underlying methodologies. From the perspective of lifelog retrieval and continuous image capture, there are numerous other techniques which could be implemented and evaluated. For example, the introduction of a wrap-around style memory wall to take better advantage of user's peripheral vision or the investigation of new methods to highlight keyframe images in event summaries. This could also include an exploration of a wide variety of novel visualisation approaches which better utilise the additional spatial dimensions afforded by virtual reality platforms.

Other future work in this area could focus on lifelogging criteria outside of lifelog retrieval. For example, in the context of reminiscence, lifeloggers could provide and explore their own datasets from the perspective of reliving memories or experiences they had forgotten and evaluate if their exploration in a virtual environment proves more effective than conventional alternatives. Such a system might include visualisations which do not rely on a user's information need, but instead focus on providing events which the system has determined to be noteworthy or interesting. For example, events such as birthdays, weddings or holidays, which could be exposed to users seamlessly within the virtual environment. This type of visualisation technique might resemble a virtual landscape of personal memories which the user explores at their own pace and discretion. Such interactions could be supplemented with 360 degree images and videos which, when rendered within virtual reality, could recreate personal experiences in a manner so immersive it is almost indistinguishable from the original experience.

Finally, there remains a body of work surrounding the precise effectiveness of the algorithm used to rank events described in chapter 5 of this research, as it was not formally evaluated against the state-of-the-art. Though it affected all experiment variables equally, and therefore should not have introduced a bias, future work could better isolate the algorithm's effectiveness alongside more traditional ranking techniques to determine if there is further room for improvement in the support of lifelog retrieval.

7.3.3 Further Validation of Lifelog Design Framework

There is significant potential for future research in the further evaluation of the Lifelog Design Framework introduced in this work. Though we have determined its appropriateness in the development of a novel virtual reality lifelog retrieval application, its usage on other technology, such as personal computers or smartphones, and other lifelogging criteria, such as reflection or recollection, would greatly contribute to its further validation and refinement. However, this relies on the continued release of lifelogging test collections, preferably of greater size and diversity, so that more novel lifelog applications can be designed and evaluated.

7.4 Publication List

In this section we provide a list of each publication produced throughout this work. Though we recognise each entry as a valuable contribution to the domain of lifelogging, only a subset of these contributions are directly related to the research described in this thesis. For the most relevant further reading in relation to this work, we recommend some of the more recent papers listed below, specifically 'Comparing Approaches to Interactive Lifelog Search at the Lifelog Search Challenge (LSC2018)' and 'Virtual Reality Lifelog Explorer: Lifelog Search Challenge at ACM ICMR 2018'.

- C. Gurrin, K. Schoeffmann, H. Joho, A. Leibetseder, L. Zhou and **A. Duane**, D. Dang-Nguyen, M. Riegler, L. Piras, M. Tran, J. Lokoč, W. Hürst, 'Comparing Approaches to Interactive Lifelog Search at the Lifelog Search Challenge (LSC2018)' (2019), ITE Transactions on Media Technology and Applications, Volume: 7, Issue: 2, p46-59

- **A. Duane**, C. Gurrin, 'User Interaction for Visual Lifelog Retrieval in a Virtual Environment' (2019), Proceedings of the 2019 International Conference on Multimedia Modelling, p239-250

- **A. Duane**, C. Gurrin, W. Huerst, 'Virtual Reality Lifelog Explorer: Lifelog Search Challenge at ACM ICMR 2018' (2018), Proceedings of the 2018 ACM Workshop on The Lifelog Search Challenge, p20-23

- W. Hürst, K. Ouwehand, M. Mengerink, **A. Duane**, C. Gurrin, 'Geospatial Access to Lifelogging Photos in Virtual Reality' (2018), Proceedings of the 2018 ACM Workshop on The Lifelog Search Challenge, p20-23

- **A. Duane**, C. Gurrin, 'Lifelog Exploration Prototype in Virtual reality' (2018), Proceedings of the 2018 International Conference on Multimedia Modelling, p377-380

- L. Zhou, **A. Duane**, D. Nguyen, D. Tien, C. Gurrin, 'DCU at the NTCIR-13 Lifelog-2 Task' (2017), Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies

- L. N. Signal, J. Stanley, M. Smith, M. B. Barr, T. J. Chambers, J. Zhou, **A. Duane**, C. Gurrin, A. F. Smeaton, C. McKerchar, A. L. Pearson, J. Hoek, G. L. S. Jenkin and C. Ni Mhurchu 'Kids' Cam: An Objective Methodology to Study the World in which Children live' (2017), American Journal of Preventive Medicine, Volume: 53, Issue: 3, p89-95

- L. N. Signal, J. Stanley, M. Smith, M. B. Barr, T. J. Chambers, J. Zhou, **A. Duane**, C. Gurrin, A. F. Smeaton, C. McKerchar, A. L. Pearson, J. Hoek, G. L. S. Jenkin and C. Ni Mhurchu 'Children's Everyday Exposure to Food Marketing: An Objective Analysis using Wearable Cameras' (2017), The International Journal of Behavioural Nutrition and Physical Activity, Volume: 14, Issue: 1, p137

- **A. Duane**, C. Gurrin, 'Pilot Study to Investigate Feasibility of Visual Lifelog Exploration in Virtual Reality' (2017), Proceedings of the 2nd Workshop on Lifelogging Tools and Applications, p29-32

- **A. Duane**, J. Zhou, S. Little, C. Gurrin, A. Smeaton 'An Annotation System for Egocentric Image Media' (2017), Proceedings of the 2017 International Conference on Multimedia Modelling, p442-445

- **A. Duane**, R. Gupta, L. Zhou, C. Gurrin, 'Visual Insights from Personal Lifelogs' (2016), Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies, p386-389

- J. Zhou, **A. Duane**, R. Albatal, C. Gurrin, D. Johansen, 'Wearable Cameras for Real-Time Activity Annotation' (2015), Proceedings of the 2015 International Conference on Multimedia Modelling, p319-322

7.5 Conclusion

The research addressed in this work has focused on the confluence of interactive information retrieval within virtual reality and visual access to lifelog datasets based on total capture to provide an early exploration into virtual reality's viability in supporting lifelog applications. These studies informed the design of a prototype lifelog retrieval system for virtual reality which is intended as a design baseline for future such applications developed in this field of research and should be further supported by the introduction of a novel design framework intended specifically to assist in the reproduction of effective lifelog applications. These contributions enabled us to perform an in-depth analysis within this domain, where we proposed three research questions which, upon investigation, assisted in evaluating the validity of our hypothesis which we consider to be upheld.

Bibliography

Alsina, A., Giró, X., and Gurrin, C. (2018). “An Interactive Lifelog Search Engine for LSC2018”. In: arXiv: 1404.7828.

Aris, A. et al. (2004). “Exploiting location and time for photo search and storytelling in MyLifeBits”. In: *Citeseer* September, pp. 1–8.

Bahrainian, S. A. and Crestani, F. (2017a). “Are Conversation Logs Useful Sources for Generating Memory Cues for Recalling Past Memories?” In: 8, pages.

Bahrainian, S. A. and Crestani, F. (2018). “Augmentation of Human Memory: Anticipating Topics that Continue in the Next Meeting”. In: *ACM Reference*, pp. 150–159.

Bahrainian, S. A. and Crestani, F. (2017b). “Towards the Next Generation of Personal Assistants”. In: vol. 9, pp. 169–176.

Barrett, M. A. et al. (2013). “Big Data and Disease Prevention: From Quantified Self to Quantified Communities”. In: *Big Data* 1.3, pp. 168–175.

Bell, G., Chen, R., and Rege, S. (1972). "Effect of Technology on Near Term Computer Structures". In: *Computer* 5.2, pp. 29–38.

Bell, G. and Gemmell, J. (2009). *Total Recall: How the E-Memory Revolution Will Change Everything.*

Boff, K. R., Kaufman, L., and Thomas, J. P. (1986). *Handbook of perception and human performance*. New York, John Wiley and Sons.

Bowman, D. A., Kruijff, E., et al. (2004). *3D User Interfaces: Theory and Practice*, pp. 117–118.

Bowman, D. A., McMahan, R. P., and Ragan, E. D. (2012). "Questioning naturalism in 3D user interfaces". In: *Communications of the ACM* 55.9, p. 78.

Brooks, F. P. and P., F. (1987). "Walkthrough—a dynamic graphics system for simulating virtual buildings". In: *Proceedings of the 1986 workshop on Interactive 3D graphics - SI3D '86*. New York, New York, USA: ACM Press, pp. 9–21.

Bryson, S. and Levit, C. (1992). "The Virtual Windtunnel: A High-Performance Virtual Reality Application". In: *IEEE Computer Graphics and Applications* 12.4, pp. 25–34.

Burgoyne, N., Graham, C. R., and Sudweeks, R. (2010). "The Validation of an Instrument Measuring TPACK". In: *Proceedings of Society for Information Technology and Teacher Education International Conference 2010*. Ed. by D.

Gibson and B. Dodge. San Diego, CA, USA: Association for the Advancement of Computing in Education (AACE), pp. 3787–3794.

Bush, V. (1945). “As We May Think”. In: *The Atlantic*.

Butterworth, J. et al. (1992). “3DM: a three dimensional modeler using a head-mounted display”. In: *Proceedings of the 1992 symposium on Interactive 3D graphics - SI3D '92*. New York, New York, USA: ACM Press, pp. 135–138.

Caprani, N. et al. (2010). “Designing a Touch-screen Sensecam Browser to Support an Aging Population”. In: *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, pp. 4291–4296.

Chang, C.-C. et al. (2019). “An Interactive Approach to Integrating External Textual Knowledge for Multimodal Lifelog Retrieval”. In: *Proceedings of the ACM Workshop on Lifelog Search Challenge - LSC '19*, pp. 41–44.

Conway, M. (1990). *Autobiographical Memory*. Milton Keynes, UK: Open University Press.

Cooper, M. and Foote, J. (2005). “DISCRIMINATIVE TECHNIQUES FOR KEYFRAME SELECTION Matthew Cooper and Jonathan Foote Palo Alto , CA 94304”. In: *ICME 2005 - IEEE International Conference on Multimedia and Expo*.

Czerwinski, M. et al. (2006). “Digital memories in an era of ubiquitous computing and abundant storage”. In: *Communications of the ACM* 49.1, p. 44.

Dale, E. (1969). *Audio-Visual Methods in Teaching (3rd Ed.)* Dryden Press, pp. 12, 13.

Davies, C. and Harrison, J. (1996). "Osmose: towards broadening the aesthetics of virtual reality". In: *ACM SIGGRAPH Computer Graphics* 30.4, pp. 25–28.

Dodge, M. and Kitchin, R. (2007). "'Outlines of a world coming into existence': Pervasive computing and the ethics of forgetting". In: *Environment and Planning B: Planning and Design* 34.3, pp. 431–445.

Doherty, A. R., Byrne, D., et al. (2008). "Investigating keyframe selection methods in the novel domain of passively captured visual lifelogs". In: p. 259.

Doherty, A. R., Caprani, N., et al. (2011). "Passively recognising human activities through lifelogging". In: *Computers in Human Behavior*. Vol. 27. 5. Pergamon, pp. 1948–1958.

Doherty, A. R., Moulin, C. J. a., and Smeaton, A. F. (2011). "Automatically assisting human memory: A SenseCam browser". In: *Memory* 19.7, pp. 785–795.

Doherty, A. R., Pauly-Takacs, K., et al. (2012). "Experiences of Aiding Autobiographical Memory Using the SenseCam". In: *Human–Computer Interaction* 27.1-2, pp. 151–174.

Doherty, A. R. and Smeaton, A. F. (2008). "Automatically segmenting lifelog data into events". In: *WIAMIS 2008 - Proceedings of the 9th International Workshop on Image Analysis for Multimedia Interactive Services*, pp. 20–23.

Doherty, A. R., Smeaton, A. F., et al. (2007). "Multimodal segmentation of lifelog data". In: *RIAO 2007 LargeScale Semantic Access to Content Text Image Video and Sound* 2007, pp. 21–38.

Doherty, A. R., Gurrin, C., and Smeaton, A. F. (2009). "An investigation into event decay from large personal media archives". In: *Proceedings of the 1st ACM international workshop on Events in multimedia EiMM 09*, p. 49.

Duane, A. and Huerst, W. (2018). "Virtual Reality Lifelog Explorer". In:

Garon, M. et al. (2017). "Real-Time High Resolution 3D Data on the HoloLens". In: *Adjunct Proceedings of the 2016 IEEE International Symposium on Mixed and Augmented Reality, ISMAR-Adjunct 2016*, pp. 189–191.

Gemmell, J., Aris, A., and Lueder, R. (2005). "Telling stories with MyLifeBits". In: *IEEE International Conference on Multimedia and Expo, ICME 2005.* Vol. 2005, pp. 1536–1539.

Gemmell, J., Bell, G., and Lueder, R. (2006). "MyLifeBits: A Personal Database for Everything". In: *CACM* 49.1, p. 88.

Gemmell, J., Bell, G., and Lueder, R. (2002). "MyLifeBits: Fulfilling the Memex Vision". In: *Proceedings of the tenth ...* Pp. 235–238.

Graham, C. R. (2011). “Theoretical considerations for understanding technological pedagogical content knowledge (TPACK)”. In: *Computers and Education* 57.3, pp. 1953–1960.

Gregory, R. L. (1997). *Third edition: revised and updated Eye and the psychology of seeing.*

Gurrin, C., Joho, H., Hopfgartner, F., Zhou, L., and Albatal, R. (2016). “Overview of NTCIR-12 Lifelog Task”. In: pp. 354–360.

Gurrin, C., Joho, H., Hopfgartner, F., Zhou, L., Gupta, R., et al. (2017). “Overview of NTCIR-13 Lifelog-2 Task”. In: *Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies*, pp. 6–11.

Gurrin, C., Lee, H., et al. (2010). “Browsing Large Personal Multimedia Archives in a Lean-back Environment”. In: *MMM 2010: Advances in Multimedia Modeling*, pp. 98–109.

Gurrin, C., Schoeffmann, K., et al. (2019). “Comparing Approaches to Interactive Lifelog Search at the Lifelog Search Challenge (LSC2018)”. In: *ITE Transactions on Media Technology and Applications* 7.2, pp. 46–59.

Gurrin, C., Smeaton, A. F., and Doherty, A. R. (2014). “LifeLogging: Personal Big Data”. In: *Foundations and Trends in Information Retrieval* 8.1, pp. 1–125.

Gurrin, C., Smeaton, A. F., Qiu, Z., et al. (2013). *Exploring the Technical Challenges of Large-scale Lifelogging.*

Heilig, M. (1960). *Stereoscopic-television apparatus for individual use, US Patent 2955156.*

Hodges, S. et al. (2006). "SenseCam: a retrospective memory aid". In: *UbiComp 2006: Ubiquitous Computing* 4206, pp. 177–193.

Hoy, M. B. (2016). "Personal Activity Trackers and the Quantified Self". In: *Medical Reference Services Quarterly* 35.1, pp. 94–100.

Hunter, I. W. et al. (1993). "A Teleoperated Microsurgical Robot and Associated Virtual Environment for Eye Surgery". In: *Presence: Teleoperators and Virtual Environments* 2.4, pp. 265–280.

Hutchins, E. L., Hollan, J. D., and Norman, D. A. (1985). "Direct Manipulation Interfaces". In: 1, pp. 311–338.

Jerald, J. (2015). *The VR Book: Human-Centered Design for Virtual Reality.* Morgan & Claypool.

Kalantari, M. and Rauschnabel, P. (2018). "Exploring the Early Adopters of Augmented Reality Smart Glasses: The Case of Microsoft HoloLens". In: *Augmented Reality and Virtual Reality: Empowering Human, Place and Business.* Ed. by T. Jung and M. C. tom Dieck. Cham: Springer International Publishing, pp. 229–245.

Kelly, D. (2009). “Methods for Evaluating Interactive Information Retrieval Systems with Users”. In: *Foundations and Trends R in Information Retrieval* 3.2, pp. 1–224.

Khan, O. S. et al. (2019). “Exquisitor at the Lifelog Search Challenge 2019”. In: *Proceedings of the ACM Workshop on Lifelog Search Challenge - LSC ’19*, pp. 7–11.

Koehler, M. and Mishra, P. (2009). “What is technological pedagogical content knowledge (TPACK)?” In: *Contemporary issues in technology and teacher education* 9.1, pp. 60–70.

Krokos, E., Plaisant, C., and Varshney, A. (2018). “Virtual memory palaces: immersion aids recall”. In: *Virtual Reality* 23.1, pp. 1–15.

Le, T.-K. et al. (2019). “LifeSeeker”. In: *Proceedings of the ACM Workshop on Lifelog Search Challenge - LSC ’19*, pp. 37–40.

Le, N.-K. et al. (2019). “Smart Lifelog Retrieval System with Habit-based Concepts and Moment Visualization”. In: *Proceedings of the ACM Workshop on Lifelog Search Challenge - LSC ’19*, pp. 1–6.

Lee, H. et al. (2008). “Constructing a SenseCam visual diary as a media process”. In: *Multimedia Systems* 14.6, pp. 341–349.

Leibetseder, A. et al. (2019). “lifeXplore at the Lifelog Search Challenge 2019”. In: *Proceedings of the ACM Workshop on Lifelog Search Challenge - LSC ’19.*

Lokoč, J., Bailer, W., et al. (2018). “On Influential Trends in Interactive Video Retrieval: Video Browser Showdown 2015–2017”. In: *IEEE Transactions on Multimedia* 20.12, pp. 3361–3376.

Lokoč, J., Souček, T., Čech, P., et al. (2019). “Enhanced VIRET Tool for Lifelog Data”. In: *Proceedings of the ACM Workshop on Lifelog Search Challenge - LSC '19*, pp. 25–26.

Lokoč, J., Souček, T., and Kovalčik, G. (2018). “Using an Interactive Video Retrieval Tool for LifeLog Data”. In: pp. 15–19.

Macedonia, M. R. et al. (1995). “Exploiting reality with multicast groups”. In: *IEEE Computer Graphics and Applications* 15.5, pp. 38–45.

Mann, S. (2004). “Continuous lifelong capture of personal experience with EyeTap”. In: *Proceedings of the the 1st ACM workshop on Continuous archival and retrieval of personal experiences - CARPE'04*, pp. 1–21.

Mann, S. (1997). “Wearable computing: A first step toward personal imaging”. In: *Computer* 30.2, pp. 25–32.

Mann, S. et al. (2012). “Realtime HDR (High Dynamic Range) Video for Eyetap Wearable Computers, FPGA-Based Seeing Aids, and Glasseyes (Eyetaps)”. In: *25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE).*

Mapes, D. P. and Moshell, J. M. (1995). “A Two-Handed Interface for Object Manipulation in Virtual Environments”. In: *Presence: Teleoperators and Virtual Environments* 4.4, pp. 403–416.

McMahan, R. P. and Bowman, D. A. (2007). “An empirical comparison of task sequences for immersive virtual environments”. In: *IEEE Symposium on 3D User Interfaces 2007 - Proceedings, 3DUI 2007*, pp. 25–32.

Meyer, J. et al. (2014). “Beyond quantified self”. In: *Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems - CHI EA ’14*, pp. 95–98.

Mine, M. R., Brooks, J., and Sequin, C. (1997). “Moving Objects in Space : Exploiting Proprioception In Virtual-Environment Interaction”. In: *24th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 19–26.

Nguyen Van Khan, I. et al. (2019). “A Two-Level Lifelog Search Engine at the LSC 2019”. In: *Proceedings of the ACM Workshop on Lifelog Search Challenge - LSC ’19*, pp. 19–23.

Nguyen, S. P. et al. (2019). “VieLens, An Interactive Search engine for LSC2019”. In: *Proceedings of the ACM Workshop on Lifelog Search Challenge - LSC ’19*, pp. 33–35. arXiv: 1810.04805.

Norman, D. (2013). *The Design of Everyday Things: Revised and Expanded Edition.*

North, C., Shneiderman, B., and Plaisant, C. (1996). "User Controlled Overviews of an Image Library: A Case Study of the Visible Human". In: *DL'96: Proceedings of the 1st ACM International Conference on Digital Libraries*, pp. 74–82.

Pausch, R., Proffitt, D., and Williams, G. (1997). "Quantifying immersion in virtual reality". In: *Proceedings of the 24th annual conference on Computer graphics and interactive techniques - SIGGRAPH '97*, pp. 13–18.

Qiu, Z., Gurrin, C., and Smeaton, A. F. (2016). "Evaluating access mechanisms for multimodal representations of lifelogs". In: *The 22nd International Conference on Multimedia Modeling. Lecture Notes in Computer Science.*

Rogério dos Santos Alves; Alex Soares de Souza, et all (2014). "Video Retrieval using Global Features in Keyframes". In: *Igarss 2014* 3.1, pp. 1–5. arXiv: `arXiv:1011.1669v3`.

Rossetto, L. et al. (2019). "Retrieval of Structured and Unstructured Data with vitrivr". In: *Proceedings of the ACM Workshop on Lifelog Search Challenge - LSC '19*, pp. 27–31.

Rothbaum, B. O. et al. (2000). "A controlled study of virtual reality exposure therapy for the fear of flying". In: *Journal of Consulting and Clinical Psychology* 68.6, pp. 1020–1026.

Schmidt, D. A. et al. (2009). "Technological pedagogical content knowledge (Track): The development and validation of an assessment instrument for preservice teachers". In: *Journal of Research on Technology in Education* 42.2, pp. 123–149.

Schoeffmann, K. et al. (2018). "The diveXplore System at the Video Browser Showdown 2018". In: *CoRR* abs/1804.01863. arXiv: 1804.01863.

Sellen, A., Fogg, A., et al. (2007). "Do Life-Logging Technologies Support Memory for the Past? An Experimental Study Using SenseCam". In: *CHI '07 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 81–90.

Sellen, A. and Whittaker, S. (2010). "Beyond total capture". In: *Communications of the ACM* 53.5, p. 70.

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

Smeaton, A. F. and Browne, P. (2006). "A usage study of retrieval modalities for video shot retrieval". In: *Information Processing and Management* 42.5, pp. 1330–1344.

Taylor, R. M. et al. (1993). "The nanomanipulator". In: *Proceedings of the 20th annual conference on Computer graphics and interactive techniques - SIGGRAPH '93*, pp. 127–134.

Wang, L., Guanghui, W., and Alexander, C. A. (2015). “Big Data and Visualization: Methods, Challenges and Technology Progress”. In: *Digital Technologies* 1.1, pp. 33–38.

Whittaker, S., Bergman, O., and Clough, P. (2010). “Easy on that trigger dad: a study of long term family photo retrieval”. In: *Personal and Ubiquitous Computing* 14.1, pp. 31–43.

Wilding, E. (2004). “The Handbook of memory Disorders”. In: *Brain* 127.7, pp. 1681–1682.

Wilson, J. R. (1997). “Virtual environments and ergonomics: Needs and opportunities”. In: *Ergonomics*. Vol. 40. 10. Taylor & Francis Group, pp. 1057–1077.

Witmer, B. G. et al. (1996). “Virtual spaces and real world places: Transfer of route knowledge”. In: *International Journal of Human Computer Studies* 45.4, pp. 413–428.

Yang, Y., Lee, H., and Gurrin, C. (2013). “Visualizing lifelog data for different interaction platforms”. In: *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*, p. 1785.

Zhou, L. et al. (2018). “LIFER: An Interactive Lifelog Retrieval System”. In:

Appendices

Appendix A

User Feedback Questionnaires

A.1 User Interaction

How much do you agree with the following statements?

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
This interface was easy to use.	○	○	○	○	○
I could have used it without instructions.	○	○	○	○	○
I learned to use it quickly.	○	○	○	○	○
It was easy to learn to use.	○	○	○	○	○
It was fun to use.	○	○	○	○	○
I am satisfied with it.	○	○	○	○	○

Could this interface be improved?

(0/4000)

A.2 Data Visualisation

How much do you agree with the following statements?

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
The system was easy to learn to use.	○	○	○	○	○
Once I learned how, I found interacting with the user interface to be easy and intuitive to do.	○	○	○	○	○
Once I learned how, I found interacting with the lifelog data to be easy and intuitive to do.	○	○	○	○	○
The system was fun to use.	○	○	○	○	○
I could see myself preferring a virtual reality system to other conventional systems (e.g. mouse, keyboard, touchscreen).	○	○	○	○	○

What did you like most about the overall system?

(0/4000)

What did you dislike about the overall system?

(0/4000)

What suggestions or ideas do you have for improving the overall system?

(0/4000)

A.3 Comparative Analysis

How much do you agree with the following statements?

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
The system was easy to learn to use.	○	○	○	○	○
Once I learned how, I found interacting with the user interface to be easy and intuitive to do.	○	○	○	○	○
Once I learned how, I found interacting with the lifelog data to be easy and intuitive to do.	○	○	○	○	○
The system was fun to use.	○	○	○	○	○

What did you like most about the overall system?

(0/4000)

What did you dislike about the overall system?

(0/4000)

What suggestions or ideas do you have for improving the overall system?

(0/4000)

A.4 Technical Background

What age group are you in?

- ◯ 12 – 17
- ◯ 18 – 24
- ◯ 25 – 34
- ◯ 35 – 44
- ◯ 45 – 54
- ◯ 55 – 64
- ◯ 65+

How fluent are you in the English language?

- ◯ Native speaker
- ◯ Very fluent
- ◯ Somewhat fluent
- ◯ Partially fluent

How often do you use a computer in your average week?

- ◯ Less than 1 hour
- ◯ 1-10 hours
- ◯ 11-20 hours
- ◯ 21-30 hours
- ◯ 31-40 hours
- ◯ 40+ hours

How many times have you experienced virtual reality?

- ◯ Never
- ◯ 1-5 times
- ◯ 5-10 times
- ◯ 10+ times
- ◯ I own a virtual reality system

Appendix B

Known-Item Search Topic Relevance Judgements

B.1 Data Visualisation Topics

ID	Title	Relevance Judgement
1	Waiting for a Train	Any image where the lifelogger is beside a train or train tracks
2	Looking at a Mirror	Any image where the lifelogger is clearly visible in a mirror
3	Packing a Suitcase	Any image where an open suitcase and folded clothes are clearly visible
4	Gold Living Room	Any image captured in a living room with clearly visible gold-coloured walls
5	Picking Berries	Any image where the lifelogger's hand is clearly visible holding berries outdoors on a Sunday afternoon

ID	Title	Relevance Judgement
6	Whiteboard Session	Any image where the lifelogger is holding a marker up and is physically touching a whiteboard
7	Asian Restaurant	Any image where the lifelogger is in a restaurant and chopsticks are clearly visible
8	Aeroplane	Any image where the lifelogger is walking across an airport apron (where aeroplanes are parked in the airport)
9	Inside a Church	Any image where the lifelogger is indoors and an alter, pews or stained glass windows are clearly visible
10	Morning Coffee	Any image of the lifelogger holding a coffee cup on a Saturday morning

ID	Title	Relevance Judgement
11	Fruit Bowl	Any image where a bowl of fruit is clearly visible
12	Writing on Paper	Any image where the lifelogger is writing on a piece of paper (pen/pencil must be touching paper)
13	Eating Eggs	Any image of a plate in front of the lifelogger containing eggs
14	Using a Ticket Machine	Any image where the lifelogger is beside a ticket machine in a train station which is clearly visible
15	Antique Clock	Any image where an antique clock is clearly visible on a weekday

B.2 Comparative Analysis Topics

ID	Title	Relevance Judgement
1	Microwave Timer	Any image where the lifelogger's hand is in contact with the buttons on a microwave (not an oven)
2	Garden Shears	Any image where garden shears are clearly visible
3	Restaurant Photograph	Any image in a restaurant where the lifelogger is taking a photo of a woman using their phone
4	Striped Shirt	Any image where the lifelogger is looking at a red striped shirt in their home after work

ID	Title	Relevance Judgement
5	Refrigerator	Any image containing an open refrigerator and the lifelogger is touching a milk carton
6	Guitars	Any image containing more than one guitar in the lifelogger's home
7	House Number	Any image of a house where the number of the house is clearly visible at the entrance
8	Reading a Menu	Any image where the lifelogger is in a restaurant and a menu is clearly visible in front of them

ID	Title	Relevance Judgement
9	Coffee Machine	Any image where the lifelogger is interacting with a coffee machine (not a kettle or teapot)
10	Breakfast and Television	Any image of the lifelogger where food and a television are clearly visible in the morning time
11	50th Birthday Party	Any image of a birthday cake and '50th' is clearly visible
12	Self-Service Checkout	Any image where the lifelogger is observing a self-service checkout machine at a grocery store on a Thursday evening

ID	Title	Relevance Judgement
13	Fresh Herbs	Any image where green potted herbs are visible in the lifelogger's kitchen
14	Football in Bed	Any image where the lifelogger is clearly visible on a bed watching football (soccer) on an electronic device
15	Driving for Groceries	Any image of the lifelogger driving a car in a single day before arriving at a grocery store
16	Walking in the Rain	Any image of the lifelogger on a weekend morning where rain or an umbrella being used is clearly visible