Baseline Analysis of a Conventional and Virtual Reality Lifelog Retrieval System

Aaron Duane and Cathal Gurrin

Insight Centre for Data Analytics
Dublin City University, Ireland
aaron.duane@insight-centre.org
cathal.gurrin@insight-centre.org

Abstract. Continuous media capture via a wearable devices 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 multimodal corpora include visual and other sensor data and are enriched by content analysis, to generate as extensive a record of an individual's life experience. 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. In this work, we suggest that the increase in immersion and spatial dimensions provided by virtual reality could provide significant benefits to users when compared to more conventional access methodologies. Hence, we motivate virtual reality as a viable method of exploring multimedia archives (specifically lifelogs) by performing a baseline comparative analysis using a novel application prototype built for the HTC Vive and a conventional prototype built for a standard personal computer.

Keywords: lifelog · virtual reality · information retrieval.

1 Introduction

A natural by-product of progress in computing technology has resulted in the progression from desktop-based computing to mobile computing and (more recently) pervasive and immersive computing. Every individual with a smartphone or other wearable sensors can generate large and continuous archives of multimodal data, while at the same time, new access methodologies are becoming popular, such as augmented and virtual reality. In this position paper, we explore how feasible it is to employ new access methodologies (virtual reality) to provide a novel access mechanism for personal multimodal data and we hypothesise that a well designed virtual reality interactive retrieval system can be as effective as a conventional retrieval system. We do this by developing prototype retrieval systems for archives of continuous multimodal data gathered by individuals, otherwise known as lifelogs [12], which are passively captured, automatic and continuous collections of data pertaining to a person's life or life

experience. Lifelogs can vary in their comprehensiveness from highly selective [1][14] to indiscriminate total capture [8][18].

Total capture lifelogging typically produces vast image corpora which are often further enriched by a new generation of multimodal content analysis such as visual concept detection [2][4] and event segmentation [13]. Hence such archives pose significant retrieval challenges for the community. Developing applications which enable lifelog retrieval remains an active area of research and there has been notable work designing such systems on a myriad of access mechanisms such as personal computers, tablets and even smartphones [20][17]. However, to date, there has been very little research into developing such applications for less conventional platforms such as virtual reality, which has seen a notable resurgence in recent history. It is our objective in this work to motivate others in the community to consider novel access mechanisms operating over challenging datasets by providing an initial evaluation of how successful a novel virtual realty access mechanism can be used to support interactive access to lifelog archives.

2   Motivation

Though virtual reality has yet to fully mature as a platform, and is not yet as ubiquitous as other access mechanisms, there is evidence to suggest it may effectively support lifelog retrieval. For example, some researchers believe the most valuable aspect of virtual reality 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 plot informing" [19]. 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 [3] and more recent research has suggested that immersion can greatly improve user recall [15]. For many, the underlying belief motivating most virtual reality research is that it will lead to more natural and effective human-computer interfaces and there have already been promising results in several key application domains [16].

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 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. In this paper we describe such an analysis, intended to show that a virtual reality-specific interface has strong potential to effectively support lifelog retrieval. This analysis consisted of two prototype lifelog retrieval systems, one designed for virtual reality, specifically the HTC Vive, and one designed for

a standard personal computer. The primary goal was to determine a baseline design for a virtual reality application which incorporated conventions found in state-of-the-art lifelog retrieval applications and evaluate its effectiveness in comparison to a conventional alternative which contained the same features and targeted the same dataset. The research described in this paper is part of a larger body of work [6] [7] aimed at motivating virtual reality as a viable candidate for the development of lifelog interaction systems.

3 Methodology

3.1 Dataset

The primary data chosen for this analysis was the NTCIR-13 test collection [10] first released for the lifelog workshop at NTCIR-13 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 content, referred to as visual concepts, but which we refer to as *lifelog concepts* (e.g. desk, chair, phone, etc.). 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 [9] 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 analysis.

3.2 Evaluation

Our baseline analysis was conducted primarily in the form of a user study consisting of 16 participants of varying technical background. Each participant was

asked to perform a total of 16 known-item search tasks, where half were performed on the virtual reality prototype and half were performed on the conventional prototype. In this context, a known-item search task refers to a scenario where the participant is provided with a specific description of an item, for example an image, and are then asked to retrieve it using the designated lifelog system. This style of task was chosen as it is the most commonly used method to evaluate lifelog retrieval [11] and does not necessitate the user being the owner of the lifelog dataset. Further details regarding our precise experiment configuration and the topics used for our known-item search tasks used are outlined in a parallel work [5] which is outside the scope of this paper to describe.

The users were evaluated on the speed of retrieval, the number of incorrect submissions and the number of times they failed to retrieve a correct item within the allotted time (180 seconds). Furthermore, each user was asked to fill out a user feedback questionnaire after each experiment and also participated in an informal interview regarding their experience. The decision to record this type of feedback was chosen because, while it would have been preferable to recruit several hundred participants of varying experience to produce more robust quantitative data for this analysis, time constraints and project costs made this unfeasible. As a result, an emphasis on qualitative data became a primary focus, utilising open-ended questions to garner user feedback, which could then better inform the quantitative results.

4   System Overview

In this section we will provide an overview of both prototype systems in the context of their two main interactive elements; the user interface and the data visualisation. The design and approaches outlined in this section were the result of a series of iterative user studies [6] [7] which are outside the scope of this paper to be fully described in this overview.

4.1   User Interface

There were two primary components to the user interface on both retrieval prototypes; one for concept selection and one for temporal selection. This can be seen clearly in Figure 1 where we can see a user has generated a query with four concepts. The concepts are arranged alphabetically and filtered by letter whereas, to select a time, the user can choose from the seven days of the week or the twenty-four hours in the day. These two components are identical between the conventional and virtual reality prototypes, the only difference being how each is interacted with. On the personal computer, the interface occupies the entire screen and the user can select interface elements with their mouse. However, within virtual reality, this interface floats in front of the user within the virtual environment. Its position can be adjusted at any time and the user interacts with the interface via a beam emanating from each controller (see Figure 2).

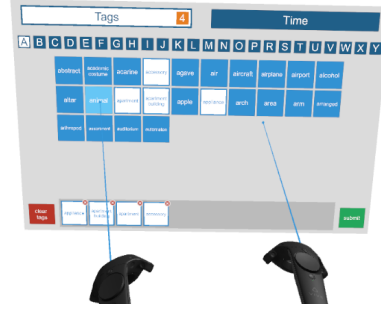**Fig. 1.** Interface interaction on conventional lifelog retrieval system

Fig. 2. Interface interaction on virtual reality lifelog retrieval system

4.2 Data Visualisation

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 3 we can observe a set of results which have been returned after the user submits a query, where each horizontal line of images represents a summary of an event, and each event is ranked from top to bottom based on the user's query. In contrast, within virtual reality (see Figure 4) we can see these summaries are aligned in a 3x3 grid and are instead ranked from left to right. To explore all the images within an event summary, the necessary interactions are very similar to how each system interacts with its respective user interface. On the conventional prototype, the user simply hovers their mouse over a relevant event and selects it, whereas on the virtual reality prototype, the user simply points their controller at a relevant event, guided by a beam. and selects it.

5 Experiment Results

5.1 Technical Background

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.

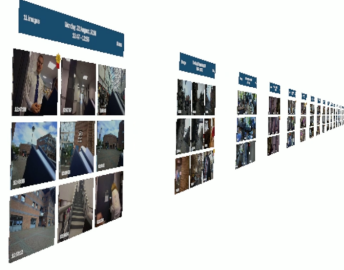**Fig. 3.** Data visualisation on conventional lifelog retrieval system

Fig. 4. Data visualisation on virtual reality lifelog retrieval system

In Figure 5 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 we can observe that almost every user had 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.
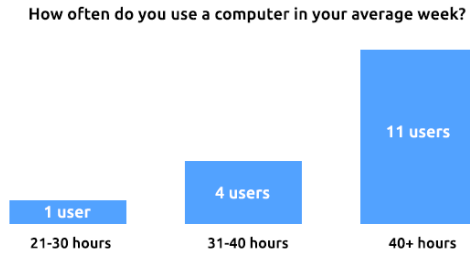

Fig. 5. User weekly exposure to computers

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 computer 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 study. For example,
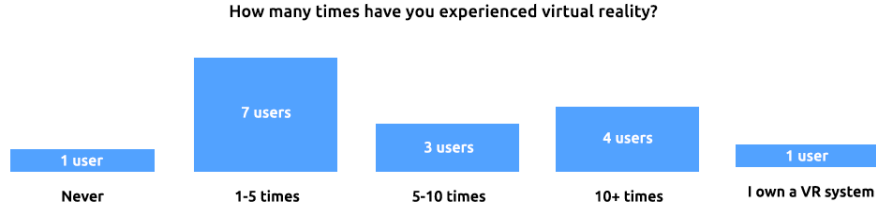
How many times have you experienced virtual reality?

Fig. 6. User previous experience with virtual reality

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.

5.2 User Performance

After all the users had participated in the experiment, each known-item search task, which were referred to as topics, 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 7 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 light 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.

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 explanation is that the topic required users to retrieve an image where the lifelogger entered 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.
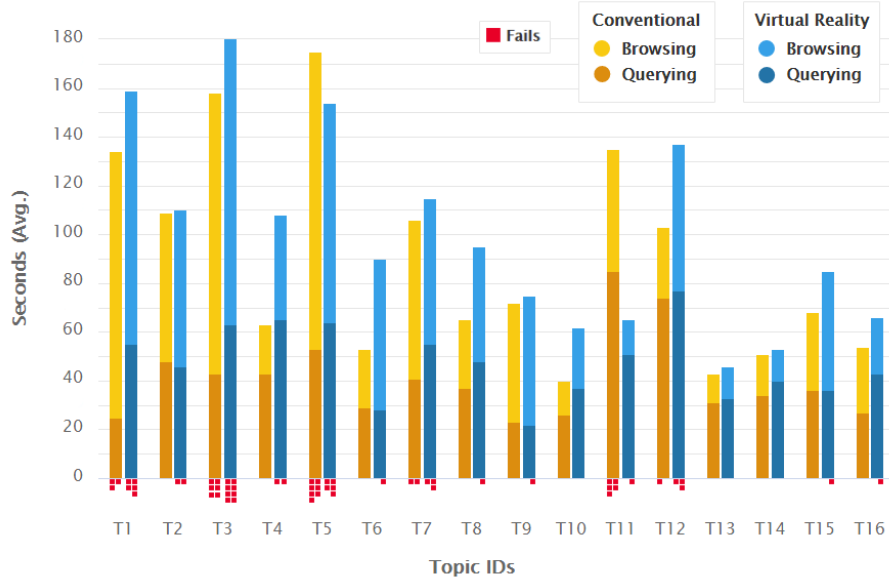
Fig. 7. Average seconds taken querying and browsing per topic for each prototype (with total failures below axis)
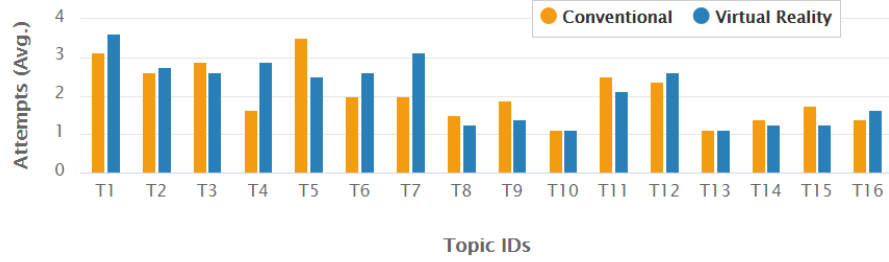

Fig. 8. Average retrieval attempts per topic for each prototype

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 and resulted in comparably short browsing times. The source of this trend likely relates to behaviour which was

observed during testing, where users were reluctant to adjust their query after they had committed to a specific set of lifelog concepts. Only after browsing a significant number of results did they concede their query was not retrieving what they intended and considered other options.

In Figure 8 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 discussed reluctance users had to attempting successive queries might be compounded more on the virtual reality system than on the conventional system.

5.3   User Feedback

Upon completion of their tasks, each participant was asked to fill out a questionnaire detailing their experiences during the experiment. The questionnaires 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 9 and Figure 10. In addition to an informal interview regarding their general experience with each prototype system, the questionnaire also contained three open questions to ensure no relevant feedback was overlooked. These questions asked users what they liked about the system, what they disliked about the system, and what they suggest 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).

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,
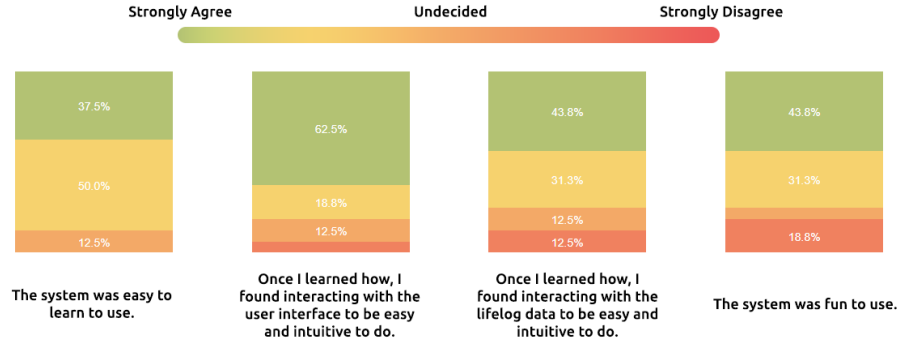
Fig. 9. User feedback for conventional retrieval prototype

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.
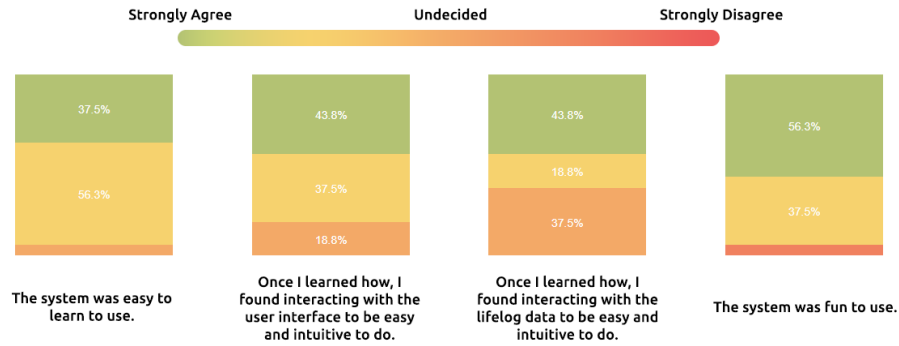
Fig. 10. User feedback for virtual reality retrieval prototype

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. Interestingly, 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.

5.4   Conclusion

In this paper we have presented a baseline analysis intended to show that a virtual reality-specific interface can effectively support interactive retrieval from a large personal multimodal data archive. This was achieved by presenting two prototype retrieval systems, one designed for the novel virtual reality platform and one designed for a conventional computing platform. Since there has been very little research into developing such retrieval applications for virtual reality, the findings of this paper suggest that novel platforms can indeed provide comparative retrieval performance, and potentially support more convenient and attractive retrieval applications. Whilst the conventional prototype proved better in terms of performance, the benefits were not significant. 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.

We acknowledge that improvements could be made to the conventional prototype through further design iterations, but this is also true for our virtual reality system. By comparing these two systems, we have determined that a primary obstacle for users when transitioning from conventional media to virtual reality to perform multimedia retrieval is the interactions necessary to formulate the retrieval query itself. Despite this obstacle, 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 a virtual reality platform. However, it should also be acknowledged that this lack of experience may directly relate to the user's enjoyment of the platform, 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. We believe that this position paper can motivate others in the community to consider novel access mechanisms operating over challenging datasets by providing an initial evaluation of how successful a novel virtual realty access mechanism can be used to support interactive access to lifelog archives.

References

1. Barrett, M.A., Humblet, O., Hiatt, R.A., Adler, N.E.: Big Data and Disease Prevention: From Quantified Self to Quantified Communities. Big Data **1**(3), 168–175 (2013)
2. Byrne, D., Doherty, A.R., Snoek, C.G., Jones, G.J., Smeaton, A.F.: Everyday concept detection in visual lifelogs: Validation, relationships and trends. Multimedia Tools and Applications **49**(1), 119–144 (2010)
3. Dale, E.: Audio-Visual Methods in Teaching (3rd Ed.). Dryden Press (1969)
4. Doherty, A.R., Moulin, C.J.a., Smeaton, A.F.: Automatically assisting human memory: A SenseCam browser. Memory **19**(7), 785–795 (2011)
5. Duane, A.: Visual Access to Lifelog Data in a Virtual Environment. Ph.D. thesis, Insight Centre for Data Analytics, Dublin City University, Ireland (2019)

6. Duane, A., Gurrin, C.: Lifelog exploration prototype in virtual reality. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). vol. 10705 LNCS, pp. 377–380 (2018)
7. Duane, A., Huerst, W.: Virtual Reality Lifelog Explorer. In: Proceedings of the 2018 ACM Workshop on The Lifelog Search Challenge (2018)
8. Gemmell, J., Bell, G., Lueder, R.: MyLifeBits: A Personal Database for Everything. CACM **49**(1), 88 (2006)
9. Gurrin, C., Joho, H., Hopfgartner, F., Zhou, L., Albatal, R.: Overview of NTCIR-12 Lifelog Task. In: Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies. pp. 354–360 (2016)
10. Gurrin, C., Joho, H., Hopfgartner, F., Zhou, L., Gupta, R., Albatal, R., Dang-Nguyen, D.T.: Overview of NTCIR-13 Lifelog-2 Task. In: Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies. pp. 6–11 (2017)
11. Gurrin, C., Schoeffmann, K., Joho, H., Leibetseder, A., Zhou, L., Duane, A., Dang-Nguyen, D.T., Riegler, M., Piras, L., Tran, M.T., Lokoc, J., Hurst, W.: Comparing Approaches to Interactive Lifelog Search at the Lifelog Search Challenge (LSC2018). ITE Transactions on Media Technology and Applications **7**(2), 46–59 (2019)
12. Gurrin, C., Smeaton, A.F., Doherty, A.R.: LifeLogging: Personal Big Data. Foundations and Trends in Information Retrieval **8**(1), 1–125 (2014)
13. Hodges, S., Williams, L., Berry, E., Izadi, S., Srinivasan, J., Bulter, A., Smyth, G., Kapur, N., Wood, K.: SenseCam: a retrospective memory aid. UbiComp 2006: Ubiquitous Computing **4206**, 177–193 (2006)
14. Hoy, M.B.: Personal Activity Trackers and the Quantified Self. Medical Reference Services Quarterly **35**(1), 94–100 (1 2016)
15. Krokos, E., Plaisant, C., Varshney, A.: Virtual memory palaces: immersion aids recall. Virtual Reality **23**(1), 1–15 (5 2018)
16. Mine, M.R., Brooks, J., Sequin, C.: Moving Objects in Space : Exploiting Proprioception In Virtual-Environment Interaction. 24th Annual Conference on Computer Graphics and Interactive Techniques pp. 19–26 (1997)
17. Qiu, Z., Gurrin, C., Smeaton, A.F.: Evaluating access mechanisms for multimodal representations of lifelogs. In: The 22nd International Conference on Multimedia Modeling. Lecture Notes in Computer Science (2016)
18. Sellen, A., Whittaker, S.: Beyond total capture. Communications of the ACM **53**(5), 70 (5 2010)
19. Slater, M., Wilbur, S.: A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. Presence: Teleoperators and Virtual Environments **6**(6), 603–616 (12 1997)
20. Yang, Y., Lee, H., Gurrin, C.: Visualizing lifelog data for different interaction platforms. CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13 p. 1785 (2013)