Lifelog Access in VR via Filtering

Master Thesis Game and Media Technology (GMT)

Hidde Veer

November 8, 2021

Student Number: 5721156
Supervisor and first examiner: Wolfgang Hürst
Second examiner: Remco Veltkamp

Graduate School of Natural Sciences
Utrecht University
The Netherlands

PREFACE

This Msc thesis, made by Hidde Veer and supervised by Wolfgang Hürst & Remco Veltkamp is presented in the form of a scientific paper with appendices. The main paper presents the research in a complete and concise manner, whereas the appendices provide more context and details about the paper itself, and the decisions made.

The appendices include a full literature study with a list of references (A + B), a summary of technical implementation details (C), a list of tasks used for the Introduction phase of the experiment (D), a list of queries used for the Retrieval phase of the experiment (E), a summary and elaboration on the used metrics for evaluation (F), a more extensive presentation and analysis of the results (G), and the forms used during the experiment in PDF form (H).

CONTENTS

I Introduction and Related Work 4
I-A Lifelogging . . . 4
I-B Lifelog Retrieval . . . 4
I-C Virtual Reality for Lifelog Access . . . 5
I-D Geospatial Filtering . . . 5
I-E Temporal Filtering . . . 5
I-F Conceptual Filtering . . . 5

II Implementation 5
II-A Map Interface . . . 5
II-A1 Markers . . . 6
II-A2 Heatmap . . . 6
II-B *DateTime* Interface . . . 6
II-B1 *Buttons* . . . 6
II-B2 *C&C* . . . 6
II-C *Tag* Interface . . . 6
II-C1 *TagList* . . . 6
II-C2 *Hierarchy* . . . 7
II-D ImageView . . . 7

III Study Goal and Research Questions 7

IV User Study 7
IV-A Procedure . . . 7
IV-B Introduction Phase . . . 7
IV-C Retrieval Phase . . . 8
IV-D Evaluation Phase . . . 8

V Data Set 8

VI Results 8

VII Discussion 9
VII-A Map Interface . . . 9
VII-A1 Markers vs. Heatmap . . . 9
VII-A2 Controls . . . 9
VII-A3 Filtering . . . 9
VII-B DateTime Interface . . . 9
VII-B1 Buttons vs. C&C . . . 9
VII-C Tag Interface . . . 9
VII-C1 Controls . . . 9
VII-C2 Tag Selection . . . 10
VII-C3 Readability . . . 10

VIII Conclusion 10

IX Future Work 10

References 10

Appendix A: Literature Study 12
A-A Introduction . . . 12
A-B The totality of a user's experience . . . 12
A-C Multimodal Data Capturing . . . 13
A-D Multimedia Archives . . . 13
A-E Lifelogging Applications . . . 14
A-F Data Set . . . 14
A-G Event Segmentation . . . 15

A-H Image Classification . . . 15
A-I Filtering Interface . . . 16
A-J Temporal Filtering . . . 18
A-K Conceptual Filtering . . . 18
A-L Image Visualization . . . 18
A-M Conclusion . . . 18

Appendix B: References 19

Appendix C: Implementation Details 20

Appendix D: Tasks 21
D-A Map Interface . . . 21
D-B Temporal Interface . . . 21
D-C Tag Interface . . . 21
D-D Image List . . . 21

Appendix E: Queries 21
E-A Set 1 . . . 21
E-A1 Q1.1 . . . 21
E-A2 Q1.2 . . . 21
E-A3 Q1.3 . . . 21
E-A4 Q1.4 . . . 21
E-A5 Q1.5 . . . 21
E-B Set 2 . . . 21
E-B1 Q2.1 . . . 21
E-B2 Q2.2 . . . 21
E-B3 Q2.3 . . . 22
E-B4 Q2.4 . . . 22
E-B5 Q2.5 . . . 22

Appendix F: Evaluation 22
F-A Demographics . . . 22
F-B Performance . . . 23

Appendix G: Results 23
G-A Demographics . . . 23
G-B Time per query & Attempts made . . . 24
G-C Time per interface . . . 25
G-D SUS Average . . . 25

Appendix H: Forms 25

Lifelog Access in VR via Filtering

Hidde Veer
Graduate School of Natural Sciences (GSNS)
Utrecht University
Utrecht, The Netherlands
h.s.veer@students.uu.nl

***Abstract*—Archives of lifelog data are generally difficult to access and explore due to the large amount of data that they contain. Especially for visual data, that is, lifelog images, we suggest that the larger and more immersive screens of virtual reality headsets may provide a good way for efficient and pleasant lifelog exploration. Yet, standard interaction designs used for lifelog data on regular screens may not translate well into virtual reality. We propose a visualization-based approach where characteristics such as location, time, and type of content are visualized via maps, calendars, and tags, respectively. We expect that such a vision-based interface is easy to handle and allows people to explore the database and find information by filtering out images that fulfill certain criteria. Different implementations of these visualizations are evaluated in a comparative study testing both general exploration tasks and more targeted search tasks. Our results prove the feasibility of this idea, but also illustrate the relevance of how certain visualizations are implemented, and related functionality is provided. For example, while the map-based visualization was suffering from different shortcomings, participants were very positive about the tag-based part of the interface that allowed them to solve the required tasks efficiently. Our results provide a proof of concept for such a filtering- and visualization-based approach for lifelog access in virtual reality. The identified issues offer concrete suggestions for further development and necessary modifications.**

***Index Terms*—Lifelogging, Lifelog Retrieval, Virtual Reality, Filtering Interface**

I. Introduction and Related Work

We evaluate how virtual reality (VR) can be used to provide efficient access to lifelog data, in particular images captured at regular time intervals with body-worn cameras. Our work is motivated by the assumption that the bigger and fully immersive screen of a head-worn VR display enables a good and effective interaction experience, but knowledge about good interaction design for this context, i.e., lifelog access in VR, is lacking. In the following, we therefore start by introducing the lifelogging context (Section I-A) and common methods for interaction with and access to lifelog data (Section I-B), before introducing VR and its characteristics that are relevant in this context. We argue that in the relation of lifelog access in VR, filtering may be superior or at least an important complement to common querying approaches. Sections I-D to I-F therefore address related work with respect to our scenario.

A. Lifelogging

A *lifelog*, in its broadest sense, is a personal record of one's daily life. Those who do the process of recording their life are commonly referred to as *lifeloggers* and the process of doing so is *lifelogging*. Gurrin, Smeaton and Doherty [1] define lifelogging as "*the process of passively gathering, processing, and reflecting on life experience data collected by a variety of sensors, and is carried out by an individual, the lifelogger*".

A definition that better highlights the large scope of lifelogging was proposed by Dodge and Kitchin, who defined it as "*a form of pervasive computing, consisting of a unified digital record of the totality of an individual's experiences, captured multi-modally through digital sensors and stored permanently as a personal multimedia archive*" [2]. We can extract three key aspects of lifelogging, namely the recording of the totality of an individual's experience (i), capturing it through a wide variety of digital sensors (ii), and storing it in a personal multimedia archive (iii). This study focuses on the third aspect, specifically how to access such an archive and interact with it.

B. Lifelog Retrieval

Lifelog Retrieval is the process of retrieving specific lifelog data from an archive. In practice, this data is often in the form of images that were captured automatically in regular time intervals, such as every 1-3 minutes, with small, body-worn cameras. Attributes such as the GPS location and visual properties of the photo are used as classification tools to help discern within a large data set. Numerous applications have been created that allow users easy access to lifelog data. These systems generally aim to optimize one of the following aspects: User experience for exploration, or query efficiency for retrieval. The latter type of systems utilize various techniques ranging from interface design to backhand systems that found their origin in multimedia/video retrieval, and face off each year in the Lifelog Search Challenge(LSC) [3]. In this challenge, each system is faced with a set of retrieval tasks they must resolve as quickly and efficiently as possible. The LSC is therefore a good comparison of state-of-the-art retrieval systems. Yet, it is focused on search time and thus efficiency, but less on user experience and, for example, discovery of unknown data.

In this research, we present and test a system for lifelog retrieval in VR through three main filtering dimensions (Geospatial, Temporal & Conceptual), that aims to find a balance between efficiency and user experience. Thus, while allowing some typical search tasks of the LSC, our evaluation of the system does not solely focus on performance but general usability and user satisfaction as well.

C. Virtual Reality for Lifelog Access

Head-mounted displays (HMDs) for VR can offer users an egocentric 360 degree view of a simulated environment. The larger field of view compared to desktop applications, combined with 3D visualizations and a more immersive experience may offer interesting opportunities for lifelog data exploration and retrieval. Duane [4] created the VRLE (Virtual Reality Life Explorer), a lifelog retrieval application that is primarily focused on efficiency. The interface of the VRLE is primarily inspired by standard desktop interfaces relying on mouse and keyboard input. While the VRLE has proven itself in previous LSCs, this approach may not be optimal for VR and better interaction designs for exploration and retrieval may exist. Ouwehand [5] and Van Abeelen [6] took a different approach, where they visualized the geospatial information contained in lifelog images in order to provide easy access to them. While their work was mostly motivated by providing a more exploratory experience, it can also be used to for retrieval if and only if the query is geospatially motivated (e.g., all photos from Central Park in NYC). This approach of visualizing metadata associated with photos and using it for filtering shows promise, but is restricted to the geospatial domain only. In the following, we therefore dive deeper into geospatial filtering before discussion alternative and complementary filtering approaches.

D. Geospatial Filtering

While Duane did not visually represent geospatial metadata, Ouwehand and Van Abeelen both used a flat floor-based map with pins. Images that are taken at approximately the same location are grouped into a single pin with an image thumbnail. Though this technique works well for exploratory purposes, it has limitations when it is used for retrieval. It is not very efficient to move large distances over the map quickly, for instance, moving from Ireland to China, and due to the nature of lifelog data (lifeloggers frequently visit the same place, e.g. home and work), high-density areas on the map can cause clutter. Motivated by these results, we decided to realize geospatial filtering with a 2D map located perpendicular to the user's view but use a more intelligent clustering algorithm.

E. Temporal Filtering

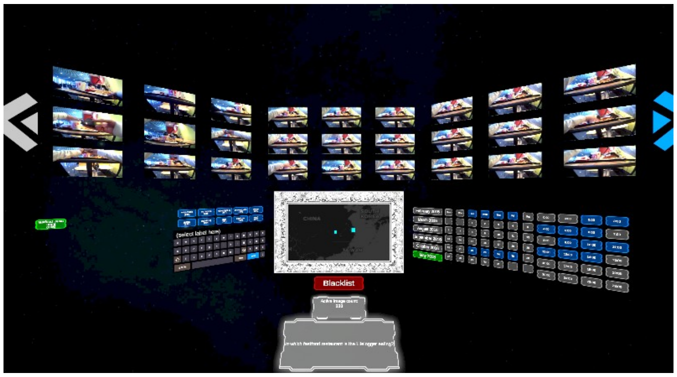
Temporal filtering allows images to be selected based on the date and time they were taken. Van Abeelen visualized temporal data by adding a third dimension to the previously created floor-based map, whereas Duane used a simple interface in which the date and time was selected using a button system. Our study aims to expand on Duane's temporal interface, maintaining its simplicity while better utilizing the possibilities of VR, for example, by better utilizing the user's peripheral view.

F. Conceptual Filtering

Each image is classified with visual concepts that describe its objects and attributes. These concepts translate raw image data in keywords and attributes that humans can understand. The extraction of visual concepts often relies on computer vision techniques that retrieve shapes and objects from images, as manual annotation is infeasible for large data sets. The VRLE allows its users to query for concepts (or "tags") by querying on their first letter, which can be inefficient as some letters are likely to contain more tags than others. Our system extends this by introducing a more detailed tag filtering interface that grants the user more options to find certain tags.

II. Implementation

We created a novel lifelog retrieval application with the focus on filtering techniques to find and examine images. This application consists of three filtering interfaces (*Map*, *DateTime & Tag*) and a separate interface to inspect images (*ImageView*). The application can be used in its entirety while seated. The controls are specified for the HTC Vive®[1], but may be adjusted to support other devices. An overview of the application can be seen in Figure 1



In which fastfood restaurant is the Lifelogger eating?

Fig. 1. An overview of the VR lifelog retrieval system, with the *Map* located in the center, with the *DateTime* and *Tag* interfaces on the right and left respectively. In the top half, the *ImageView* can be seen. The select/blacklist button is located underneath the *Map*, with two displays containing the number of filtered images and Introduction/Retrieval queries used in the experiment.

A. Map Interface

In the center of the screen in front of the user is the *Map* interface, which is a 2D map located in a frame perpendicular to the user, on which lifelog images are geospatially represented. These images are grouped and displayed as clusters on the map. The DBSCAN algorithm [7] is used for clustering, as it is able to cluster images in real-time based on their geospatial proximity to other images. This approach is advantageous as it is able to bridge small gaps between GPS coordinates. In addition, no knowledge about the number of clusters needs to be known beforehand, as is the case with other commonly used approaches such as k-means [8].

Users can move the map by using the left trackpad of the VR system's controllers, with the speed increasing on higher zoom levels. Clusters can be filtered out by moving them outside the frame; only images within the frame are active.

Users can also zoom in and out by respectively tapping the top and bottom parts of the right trackpad. Clusters become smaller and more detailed when zoomed in, and are updated dynamically based on the current zoom level. In this study, we

[1] https://www.vive.com/eu/product/vive/

implemented and tested two visualization techniques for image clustering, namely the *Markers* and *Heatmap* visualizations.

1) Markers: Markers visualization (Figure 2) represents clusters as blue markers placed on the map. The size of these markers is based on the number of images in the cluster. Markers offer high accuracy and visibility, but could become cluttered in high density areas the DBSCAN algorithm does not consider as a single cluster.

2) Heatmap: Heatmap visualization (Figure 2) represents clusters in a heatmap overlay on the map, with size and density both based on the number of images per cluster. Heatmaps offer a clear density-based overview of cluster locations, at the cost of reduced visibility of individual clusters.

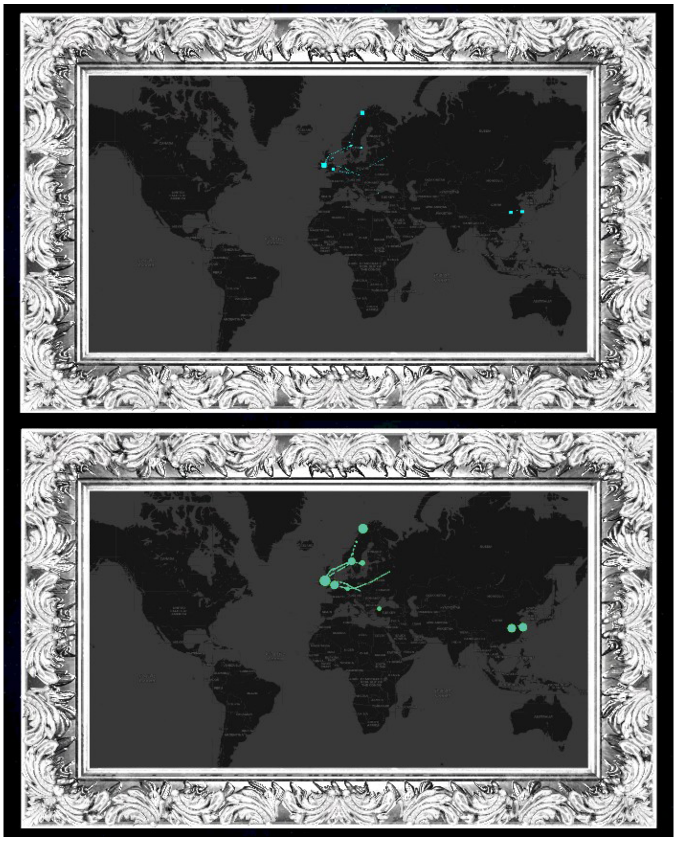


Fig. 2. Markers (top) and Heatmap (bottom) visualizations

B. *DateTime Interface*

On the right of the *Map* is the *DateTime* interface. This interface can be used to select or blacklist month/year combinations, days of the week (e.g. Friday), days and hours (e.g. 19:00-20:00). Selecting a filter will disable all images taken at a different point in time. Blacklisting does the opposite, only keeping images that are not taken at the given moment. Users interact with the interface using a virtual pointer attached to the right controller. Selecting and blacklisting is done by "clicking" the trigger button of their right controller on a date/time, clicking again will undo the action. Users may alternate between selecting and blacklisting by clicking a separate button right below the *Map*. In this study, we implemented two *DateTime* interfaces, namely the *Buttons* and the *C&C*.

1) Buttons: The *Buttons* interface (Figure 3) consists of a set of clickable buttons for selection and blacklisting, presenting a simple and straightforward filtering interface. Its strengths lie in its clarity and easy of use, but it is visually uninteresting and likely not familiar to users.

2) C&C: The *Calendar and Clock*, or *C&C* interface (Figure 3) uses a traditional calendar layout for the selection or blacklisting of month/year combinations, days of the week and days. Two clocks, AM and PM, are used for individual hours. These components are commonly used for desktop and web applications and should therefore be more familiar to users, but likely more complex to operate and interact with.

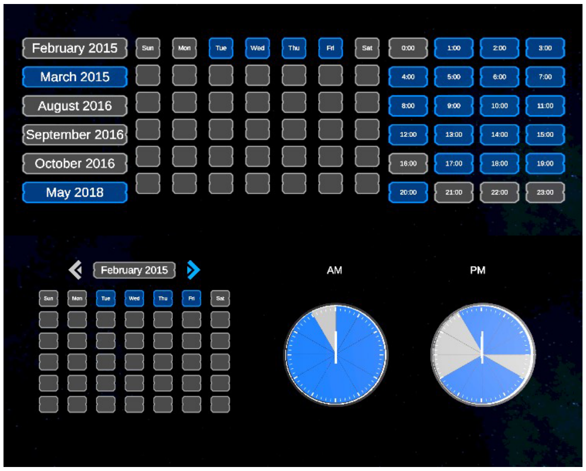


Fig. 3. Buttons (top) and C&C (bottom) visualizations

C. *Tag Interface*

On the left of the *Map* is the *Tag* interface. This interface is used to find, select and blacklist tags. Similar to the *DateTime* interface, users select, blacklist and deselect tags using the virtual pointer and trigger button of the right controller. We present two *Tag* interfaces, *TagList* and *Hierarchy*. The latter is an extension that builds on the former, which is why, also given the temporal constraints of this research, we only implemented and tested the first and suggest an evaluation of the second for future work pending a positive outcome of our evaluation of the *TagList*.

1) TagList: The *TagList* interface (Figure 4) primarily consists of a virtual keyboard supporting text based search. Up to 10 recommended tags are visible to the user based on the search query, ranking from a high to low number of occurrences. The *TagList* is precise and accurate in its search, and allows users to easily find specific tags. However, interaction with the virtual keyboard is expected to be less efficient compared to a physical keyboard.

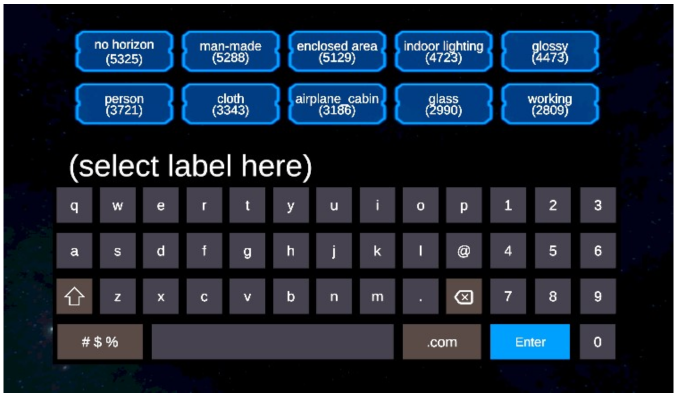


Fig. 4. *TagList* visualization

2) Hierarchy: A second interface, the *Hierarchy* is proposed but not implemented. Tags are placed in a search tree where leaf nodes are contained in the subset of the root node. Nodes can be selected by the user, and selecting a node will automatically select its subtree. An example of this can be seen in Figure 5. This root may or may not be a pre-existing tag with its own set of images. The data set used has no existing hierarchy, and we deemed manually constructing a tree containing the 532 existing tags of the *Hierarchy* interface too costly in time for this research.

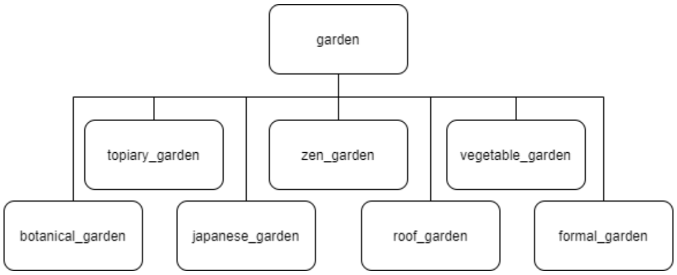


Fig. 5. Example hierarchical structure. Existing tags (leaves) are grouped under a more general, abstract tag (root).

D. *ImageView*

Located above the other interfaces, the *ImageView* becomes visible once the number of active images drops to 270 or less. It consists of three panels with a total of 27 images (one "page"), plus two buttons on the far left and right which can be clicked to scroll between pages.

Clicking an image enlarges it and hides all other UI elements, allowing the user to inspect a single image in detail. Hovering over an enlarged image with the pointer shows the image related metadata, such as its tag and location.

As the main focal point of this research lies within the filtering interfaces, the *ImageView* will not be a part of the main study, as it is solely used here to represent the filtering results. For this reason, we use a rather straightforward, standard approach of a grid-based representation of the images (Figure 1). Alternative representations of filtering results are a separate but equally interesting and relevant aspect to study in follow-up future research.

III. Study Goal and Research Questions

Our goal is to verify if filtering is an appropriate means to access and explore lifelog data in VR. Our study is a first step in verifying the feasibility and usefulness of this approach.Therefore, we focus on evaluating the implementations introduced in the last section as a proof of concept. If successful, we suggest a comparison to traditional, more query- instead of filtering-based approaches for future work.

1) Which of the implemented Geospatial filtering interfaces performs best in querying efficiency and user experience?
2) Which of the implemented Temporal filtering interfaces performs best in querying efficiency and user experience?
3) Which of the implemented Conceptual filtering interfaces performs best in querying efficiency and user experience?

To answer these research questions, we will be looking at retrieval performance and usability using the SUS [9] and qualitative feedback.

IV. User Study

As mentioned, we implemented two variations for the *Map* interface (*Markers* and *Heatmap*), two *DateTime* interfaces (*Buttons* and *C&C*) and one *Tag* interface (*TagList*). Each participant tested a unique combination of these variations (between-subjects).

In addition to some general filtering tasks used in the Introduction phase (Section IV-B), we used two sets of LSC-esque queries (*Set1* & *Set2*) that are more focused on accurate and efficient search in the Retrieval phase (Section IV-C). These queries have been selected and adjusted to compensate for the limitations of the system (e.g. no images without geospatial data) and expected inexperience of the participants with VR and/or Lifelogging. The filtering tasks, which mostly serve as a basis for the usability tests, are listed in Appendix D. The contents of each search query, which will contribute to the experienced usability, but also provide performance-related insight, can be seen in Appendix E.

The user study took place in a research lab at our university that provided a neutral space without interference or distraction. For this study the HTC Vive®HMD and controllers were used.

A. *Procedure*

Each experiment started with the participant filling in a consent and demographics form (Appendix H). We urged participants who actively suffer from motion or cyber sickness to not partake in the experiment. Participants were then seated in the center of a room, where they were asked to place the HTC Vive on their head. The HMD was adjusted to the preference of the participants and two controllers were handed to them. The main experiment was split in two phases, Introduction (Section IV-B) and Retrieval (Section IV-C). Afterwards, the participant was asked to fill in a usability survey and thanked for their time.

B. *Introduction Phase*

The Introduction phase, as the name suggests, served as an introduction to the system and each individual interface. Participants were sequentially introduced to the four interfaces (*Map*, *DateTime*, *Tag* and *ImageView*), one at a time. They were first shown a short guide to the interface in the form of a text panel, explaining the basic functionality and controls. Afterwards they were asked to solve a series of tasks related to the current interface. These tasks were questions that can be verbally answered with numbers (How many images (...)) or names (In what city (...)), or require participants to select a specific image. The exact tasks can be found in Appendix D. Participants had infinite tries for each task, and were allowed to communicate with the researcher when encountering difficulties using the interface.

C. Retrieval Phase

The Retrieval phase allowed participants to put the system to the text. Each participant attempted to solve a total of 5 queries, by submitting an image fulfilling the requirements of said query. Queries were presented in a manner similar to the LSC: Participants were faced with an initial query, which was extended every 30 seconds up to a maximum of 180 seconds. After a correct answer, participants moved on to the next query. Alternatively, not answering correctly in time or submitting the wrong image three times would also advance the query. During this game the participants' performance and behavior was measured through multiple metrics. The time and attempts taken per query was recorded, as well as the selected/blacklisted tags and dates/hours. Finally, the time spent per interface was recorded and normalized by the query time, by recording the users' view in VR. More details regarding the metrics can be found in Appendix F.

D. Evaluation Phase

After the experiment, the participant is asked to remove the HMD and answer a final evaluation form. Each interface is evaluated individually using the System Usability Scale (SUS) by Brooke [9]. The results are used to compare the individual interfaces themselves and against their variations. Afterwards, the participants are thanked for the participation, and sent on their way.

V. Data Set

We chose to use the LSC'20 [3] data set, consisting of a total amount of 191,439 wearable camera images at 1024 x 768 resolution (38.5GB). They were captured using the OMG Autographer and Narrative Clip wearable cameras, typically at a rate of 1-3 per minute during waking hours. In an effort to retain privacy, faces and most readible text has been blurred out. Accompanying the images is a collection of textual metadata, consisting of timestamps, physical activities, biometrics, and locations of the individual for every minute. Visual concepts have also been extracted, including bounding boxes of objects. Some preprocessing steps were executed to make the data set suitable for our applications. Physical activities and biometrics have been removed as they are not used, and entries without geospatial data (latitude and longitude) have been deleted.

VI. Results

A total of 16 participants took part in the study, with all but one participant (age 55-64) having an age of 18-24. One participant (*Set2*, *Heatmap*, *C&C*) was considered an outlier due to their difficulty handling VR controls and inability to properly attempt the retrieval tasks. One participant (*Set1*, *Heatmap*, *C&C*) had their screen recording damaged and thus has no data on the time spent per interface. Standard deviations for query performance were calculated by considering the average over all five queries per participant as a single data point. This is to avoid high standard deviations due to varying levels of difficulty per image. Statistical significance was calculated using the two-sample t-test with $\alpha = 0.05$ (95% Confidence Level).

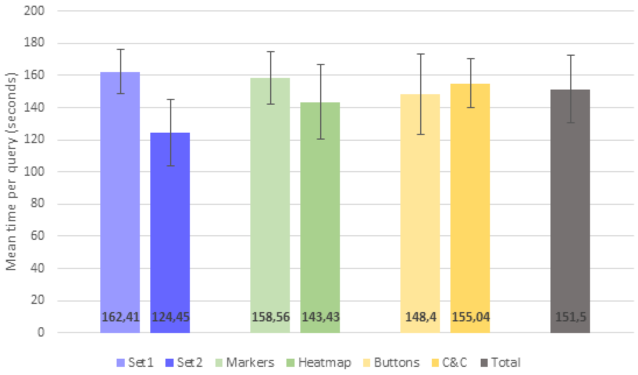


Fig. 6. Average time spent per query

Figure 6 shows the mean time spent per query per variation. If we order the queries from Set1 and Set2 by mean query time from low to high, we see that every query from *Set2* performs better than its counterpart in *Set1*. The queries from *Set1* were often more difficult than the ones from *Set2*, with Q1.4 having only been answered correctly once, and Q1.2 not having been answered correctly a single time.

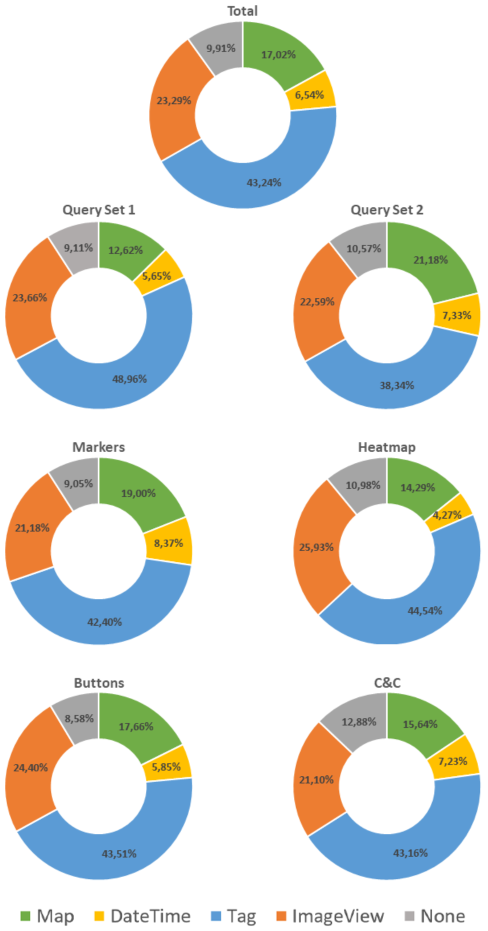


Fig. 7. Average time per interface over all queries, per query set and Map/DateTime variation.

Figure 7 shows the mean time spent on each interface per query. Time spent answering a retrieval query has been segmented in parts where the user actively focuses on one of the three filtering interfaces (*Map*, *DateTime*, *Tag*), the *ImageView* or is not focused on an individual interface (*None*), e.g. when reading the search query. On average, the *Tag* interface is most often used, followed by the *ImageView*, *Map* interface, *None*, and finally the *DateTime* interface.
The primary differences between *Set1* and *Set2* lie in the *Tag* and *Map* interfaces, with the former more prevalent in *Set1*, and the latter in *Set2*.

Table I shows the mean and standard deviation for the SUS (0-100) and its individual questions (1-5) per interface variation.

VII. Discussion

A. *Map Interface*

With an average SUS score of 61.09, the *Map* is rated worst of the three filtering interfaces. While participants praised the intuitiveness and familiarity of the interface, many also criticized the controls (Section VII-A2) and filtering methods (Section VII-A3).

1) Markers vs. Heatmap: With an average SUS score of 64.06 the *Heatmap* has received more positive reception than the *Markers* with a score of 58.13. Users also spent less time on average per query when using the former variation. No cause could be identified for this difference, as participants have made no remarks regarding clustering visualization used, during and after the experiment. We therefore hypothesize that the difference between variations is caused by the small number of participants (8 for *Markers*, 7 for *Heatmap*) and not a conclusive observation.

2) Controls: The control scheme was met with mostly negative reception from the participants. Panning was considered cumbersome and unresponsive due to the real-time processing of filtering parameters. A solution lies in further optimization of the system, or updating the parameters "on-demand" by introducing a dedicated button to update the filters.
Participants also requested more zoom levels and a legend to see the current zoom level. One user additionally requested the ability to see city names. Increasing the detail of the map requires improvements in its visibility, either with better hardware (Section VII-C3) or increasing the size of the map while in focus.

3) Filtering: Participants occasionally attempted to search for locations (e.g. Stockholm, Norway) using the *Tag* interface, before swapping to the *Map* once no matches were found. One participant expressed the wish to search by location name instead of the map. The current filtering method was also criticized: Participants felt the need to move the search target to the edge of the frame to separate it from other, nearby clusters (e.g. Ireland from the UK). Suggested improvements include the ability to select clusters with the pointer, and the ability to select/blacklist individual locations and countries. Adding location names as tags is another option, but will blur the line between the functionalities of the *Tag* and *Map* interfaces.

B. *DateTime Interface*

Users rate the *DateTime* interface marginally better than the *Map* (61.09), giving it an average SUS score of 63.44. Both variations were praised for their intuitiveness and clarity.

1) Buttons vs. C&C: With an average SUS score of 56.88, the *C&C* is the lowest rated interface variation, rated considerably worse than the *Buttons* (70). The average time per query of the *C&C* was also lower than that of its counterpart. It was mainly criticized for its clocks, which participants considered difficult to use. Participants occasionally used the wrong clock, or selected the wrong time when an exact hint was given (e.g. selecting 9:00-10:00 when the query states "around 8:50"). The inexperience of using the AM/PM system among the primarily European user base is the likely cause.

C. *Tag Interface*

With an average SUS score of 72.5, the *TagList* variation of the *Tag Interface* is the highest rated interface by the participants. It is praised for its intuitiveness and clarity, as well as its accuracy and autofill feature.

1) Controls: The primary critique was the difficulty of using the keyboard. Participants struggled to click on individual letters, often requiring more than one attempt to get it right. The causes lie in the small size of the keyboard, and lack of stability as participants unintentionally move their controllers due to the shaking of their hands.

TABLE I
AVERAGE SUS SCORE AND INDIVIDUAL QUESTION RATINGS FOR EACH FILTERING INTERFACE.

Interface	$\bar{x} \setminus \Sigma$	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	SUS Score
Markers	$\bar{x}$	3.25	4.25	3	4.13	2.63	4.13	2.88	2.75	3.25	3	58.13
	σ	1.16	0.71	1.2	0.83	0.52	1.13	1.64	0.71	1.04	1.2	12.94
Heatmap	$\bar{x}$	2.75	3.63	2.75	3.75	2.25	4.13	2.5	3.25	3	4.13	64.06
	σ	0.71	1.19	0.89	1.16	0.89	0.35	0.93	1.04	0.76	0.64	10.08
Buttons	$\bar{x}$	2.25	3.63	2.25	4	2.25	4.25	2.25	3.88	2.63	3.88	70
	σ	0.89	1.41	1.28	1.31	1.04	0.71	1.16	0.99	1.3	1.13	20.7
C&C	$\bar{x}$	2.5	3.38	3.25	3.5	2.5	3.11	2.75	3	2.125	2.88	56.88
	σ	1.2	1.06	0.89	1.07	0.93	1.55	0.89	1.07	0.83	1.55	17.31
TagList	$\bar{x}$	2.13	4.31	2.5	4.5	2.19	4.31	2.25	3.06	2.13	4	72.5
	σ	1.36	0.87	1.37	0.89	0.83	1.08	1	1.24	1.02	1.37	18.23

2) Tag Selection: Participants occasionally struggled to find a specific tag due to it not existing (Q1.2 is taken in a hotel lobby, but no such tag exists), or there being similar, applicable tags: Q1.3 shows a front yard, however up to seven garden-related tags exist (Figure 5), five of which have been selected at least once. Only one of the seven (vegetable_garden) was correct, despite not being fully accurate. One participant even selected the unrelated, similarly named "beer_garden" tag. These issues are caused by the automatic annotation of images from the LSC'21 data set, and can thus not easily be resolved. Manual annotation of the images is infeasible due to the size of the data set (191,432 images), and a different algorithm will likely carry its own flaws.

3) Readability: Participants mentioned the lack of readability of individual tags. One elderly participant (55-64) reported major difficulty reading to the point where they could not identify the tag. This is likely caused by the HMD display resolution: The HTC Vive ®has a resolution of 1080×1200 per eye, resulting in an often blurry view. Using a more state-of-the-art HMD like the Valve Index ®(1440×1600), Oculus Quest 2 ®(1832×1920) or HTC Vive Pro 2 ®(2488×2488) should circumvent this issue.

VIII. CONCLUSION

As Virtual Reality is becoming its own platform for lifelog retrieval, it is important to find interface designs that optimize query efficiency and user experience. This study investigated how filtering can be used as an alternative or complement to more traditional query-centered approaches. We argued that the latter are less suited for exploring data, especially in situations with vague or non-existing search goals, and that filtering might be better suited for VR where text and related standard input methods are often perceived as cumbersome and difficult. Our results established that the usage of a 2D map as a geospatial filtering tool, despite its familiarity and thus likely benefit, requires fine-tuned interaction and extensive functionality for it to be deemed functional and enjoyable. Feedback from the users indicated concrete aspects to address when making such a filtering tool useful and beneficial. Our major general observation is that when implementing click-based interfaces, participants preferred simplicity to familiarity, and also performed better in retrieval tasks when using a less complex interface. Usage of a virtual keyboard for conceptual filtering is well liked by participants, however, as expected, interaction with the keyboard can be slow and cumbersome at times.

This research succeeded in providing a framework for lifelog access in VR, both for retrieval and exploitative purposes. While the proof of concept is solid, most limitations lie in the concrete implementation of the *Map* and other interfaces.

IX. FUTURE WORK

While the *TagList* was relatively well received by participants, usage of a virtual keyboard carries fundamental issues that could be improved upon using different interface components. Future work should therefore look at alternative techniques to a virtual keyboard for conceptual filtering, such as a click-based hierarchical interface.

While the *Map* interface was perceived rather negatively, users made various recommendations on how to improve it and make it a valuable addition to the whole system. Future work should thus look at reworking it and adding the requested functionality. The controls and geospatial filtering technique were predominantly met with criticism. Optimizing the control scheme and offering users more accurate tools to select or blacklist locations could improve both user experience and query efficiency.

While not new to lifelogging, the application of an event segmentation algorithm could help flesh out the *ImageView* by providing more context to selected images. It will likely increase user experience as a more chronological display could be applied, and query efficiency could be improved by the ability to query by context (e.g. I was eating my lunch after (...)). Future work should look at the possible applications of event segmentation, and the *ImageView* in general.

REFERENCES

[1] C. Gurrin, A. F. Smeaton, and A. R. Doherty, "Lifelogging: Personal big data," *Foundations and trends in information retrieval*, vol. 8, no. 1, pp. 1–125, 2014.

[2] M. Dodge and R. Kitchin, "'outlines of a world coming into existence': pervasive computing and the ethics of forgetting," *Environment and planning B: planning and design*, vol. 34, no. 3, pp. 431–445, 2007.

[3] C. Gurrin, T.-K. Le, V.-T. Ninh, D.-T. Dang-Nguyen, B. T. Jónsson, J. Lokoč, W. Hurst, M.-T. Tran, and K. Schoeffmann, "An Introduction to the Third Annual Lifelog Search Challenge, LSC'20," in *ICMR '20, The 2020 International Conference on Multimedia Retrieval*, (Dublin, Ireland), ACM, 2020.

[4] A. Duane, “Visual access to lifelog data in a virtual environment,” *Dublin City University*, 2019.

[5] K. Ouwehand, “Geospatial access to lifelogging images in vr,” *Utrecht University*, 2019.

[6] J. van Abeelen, “Visualising lifelogging data in temporal virtual reality environments,” *Utrecht University*, 2019.

[7] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, *et al.*, “A density-based algorithm for discovering clusters in large spatial databases with noise.,” in *kdd*, vol. 96, pp. 226–231, 1996.

[8] J. MacQueen *et al.*, “Some methods for classification and analysis of multivariate observations,” in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, pp. 281–297, Oakland, CA, USA, 1967.

[9] J. Brooke *et al.*, “Sus-a quick and dirty usability scale,” *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.

Appendix A
Literature Study

Abstract*—*This literature study accompanies the "Lifelog Retrieval in VR" research paper, serving as a background analysis and elaboration on the decisions made in the paper. We give a brief explanation of the concept of lifelogging at the start, followed by a review of the state of the art. From that point the study will focus on specific topics that are addressed in the main research.

A. Introduction

A *lifelog*, in its broadest sense, can be considered as a personal record of one's daily life. Those who do the process of recording their life are considered *lifeloggers*. Attempting a more specific definition proves troublesome, as many researchers disagree on the details regarding a lifelog. Gurrin, Smeaton and Doherty [10] define lifelogging as "*the process of passively gathering, processing, and reflecting on life experience data collected by a variety of sensors, and is carried out by an individual, the lifelogger*". Several key aspects of lifelogging are addressed here, such as the utilization of various sensors to obtain the lifelog data, and it being processed to allow the lifelogger to reflect on their previous life experiences. This concept can be traced back to 1945, when Bush envisioned the Memex [11], which was "a portmanteau of memory and index". This device was a type of desk with pulleys and levers that allowed fast retrieval of personal information in the form of archived documents.

It was not until the '90s when Steve Mann, the "father of wearable computing" started the creation of many small, wearable sensors that could be used for the purpose of lifelogging. Devices such as the EyeTap [12] allowed users to record their life as if "the eye were the camera and display". Mann has created these devices for many purposes, such as lifeglogging[2], sousveillance [13][3] and Mediated Reality[4].

The first major research project in the field of lifelogging was the MyLifeBits project, which was created in 2002 [14] and expanded upon in 2006 [15]. This project is further explained in Section A-B. Since then, research in lifelogging has accelerated is multiple directions, the scope of which is well represented in the definition of lifelogging by Dodge and Kitchin [16]: "*(lifelogging is) a form of pervasive computing, consisting of a unified digital record of the totality of an individual's experiences, captured multi-modally through digital sensors and stored permanently as a personal multimedia archive*". This definition can be split up in three different segments, namely (1) it containing the totality of an individual's experiences, (2) it being captured multi-modally through various different sensors and (3) it being permanently stored in a multimedia archive. These concepts are further explained in the following three sections.

This literature study serves to complement the VR (Virtual Reality) lifelog retrieval application introduces in the attached paper. Section A-E will therefore contain a background study on existing state-of-the-art lifelog retrieval applications, in which parallels are drawn to our own implementation. The remainder of the literature study will be used to provide a background study on the main components of our system, and motivate the decisions made.

B. The totality of a user's experience

"The totality of a user's experience" is a key aspect in many lifelogging applications. Many have made it a purpose in itself to obtain as much lifelogging data as possible, while others advocate for only gathering and processing specific data required for certain applications. This has caused a divide between two types of lifelogging [17]: *Total Capture* focuses on capturing as much data as possible. Resulting data sets often include a wide range of different data ranging from biometric data to large streams of images.

Situation-specific capture has a more narrow scope where it focuses on rich data in specific domains. One example of the latter is the *Quantified Self Movement*, that seeks to capture specific data sets most often aimed to assist users with disease prevention and health [**?**], [18]. This concept is also utilized in the fields of gamification [19] and reflective learning [20]. Nowadays, mobile apps like Google Fit®[5] and Apple Health®[6] allow the user to track their health from their smartphones, and Fitbit®[7] provides smartwatches and trackers that focus on exercise related data.

One of the front-runners of the *Total Capture* movement is the MyLifeBits project [14] [15]. The initial goal of the project was to create a lifelong archive in the form of a SQL database on Gordon Bell, one its of the co-creators. Initially, this included audio, email, documents and hyperlinks among others, but it was later expanded to include more such as images, videos, phone calls and even mouse clicks. This data could then be accessed using a Project Interface [15] in which users could observe and query the data. Over time, the project evolved in an attempt to capture everything that could be captured, and since then many more have contributed to this vision. More and more data types were added, and the project required an increasing amount of alterations to support these new advances. Eventually, the Project Interface could no longer support all the adjustments from the large amount of different authors, and became unusable without the utilization of third-party modifications and plugins that often did not go well together. The project became swallowed in its quest to capture and store everything, and its purpose became more and more questioned over time. Sellen and Whittaker have criticized this moment, and the idea of *Total Capture* in general, blaming it as "nothing more than an excuse to show off technological advancements in their fields [17]." This was complemented by their observation that lifelogging applications lacked in widespread use and knowledge, as they are often only utilized by those who are directly invested in the movement. In order to provide perspective, they propose the "five Rs", which are five core aspects of memory that

[2] **life**long cyborg**glogging**, an older term for lifelogging

[3] Inverse surveillance, in which those with authority are observed by individuals.

[4] A variation of Augmented Reality in which images from the real world can be filtered out.

[5] https://www.google.com/intl/en$_us/fit/$

[6] https://www.apple.com/ios/health/

[7] https://www.fitbit.com

could benefit from lifelogging:

Recollection: The first benefit from lifelogging is that it could help users "re-live" certain life experiences, which is often referred to as *episodic memory* [21] [22]. Examples of practical purposes are recollecting faces and people, or remembering the details of a meeting.

Reminiscence: Reminiscence is a specialized form of recollection as it is a form of "re-living" past experiences, this time for sentimental and emotional reasons rather than practicality. There is more emphasis on sharing with others, as examples include watching a home movie or flipping through a photo book.

Retrieval: Retrieval is one of the main purposes of most lifelogging applications and is the process of retrieving specific digital information of a lifelog. Retrieval is closely related to the previous two "Rs", as the retrieval of a document helps recollect its contents, and the retrieval of a specific image might support reminiscence. Lifelog applications seeking to optimize retrieval should look for ways to efficiently find specific data through large heterogeneous data sets through various querying and visualization techniques, some of which are explained in Section A-H. Retrieval in itself is the most commonly investigated aspect in the field of lifelogging, and also the main focus of this research.

Reflection: Looking back on past experiences through lifelog data can assist the user in creating a more abstract view of past events. Reflection is the process of looking at the past with a new perspective, and utilizing this information for the benefit of health, self-identity and learning. This is in contrast to recollection, which is more oriented on memory in itself.

Remembering Intentions: Rather than relating to retrospective memories, these aspects focuses on prospective memories. Through lifelogging data, it is possible to remember activities that have to be done in the future, such as taking medicine or showing up at an appointment.

C. Multimodal Data Capturing

Capturing multimodal data using a wide array of sensors is one of the key aspects of lifelogging. Life itself cannot be captured using photographs, audio and video exclusively, as life itself is a multimodal experience. In practice, this multimodality executed on different levels, as using a wide array of devices can result in practical issues. In the past, capturing large amounts of different data required large and bulky setups to be carried around by the lifelogger (Figure A.1). Over time, these devices became smaller and smaller, and new devices were created to better suit the needs of lifelogging.
One of the first devices for the purpose of lifelogging was the Microsoft SenseCam®[8], which was a portable camera capable of automatically capturing large numbers of photographs per day. It was complemented by a PC-based application called the SenseCam Photo Viewer to manage and view the captured images. The original purpose of this device was to be a retrospective memory aid that helps the wearer recollect experiences that have subsequently been forgotten [23]. At the time of its creation, the concept of lifelogging was still in its early days, and thus the device and software were relatively simple. A more recent device is the Google Glass®[9], which was first released in 2014, discontinued in 2015, with newer versions being released in 2017 and 2019 for industry only. Research [10] anticipated that the Google Glass would revolutionize lifelogging by appealing to a broad audience, despite lifelogging not being one of the main focus points of the Glass. When looking at the current state of affairs, that does not seem to be the case. Nowadays, most of the lifelog capture functionality has been built into the smartphone. With the ability to record audio, video, GPS positions and much more, this device replaces a large range of individual sensors, making lifelogging a lot more accessible to a public audience. Numerous apps are available that utilize these functions for the purpose of lifelogging, most often in the context of Quantified Self.

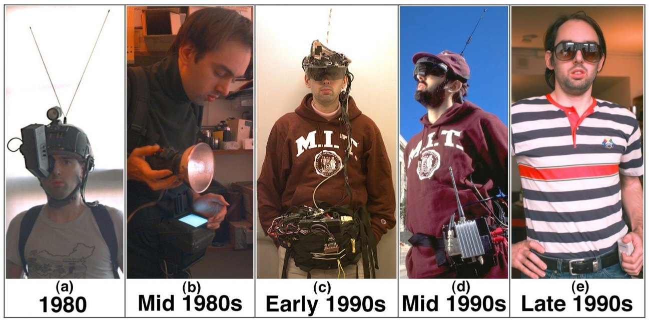


Fig. A.1. The evolution of data capturing in the early days of lifelogging [12].

D. Multimedia Archives

In recent times, lifelogging research has moved away from the technical aspect of gathering and storing data, towards creating interfaces that are capable of visualizing and retrieving it efficiently. One obstacle that quickly comes to light is the sheer amount of data storage required for a lifelogging archive. Gurrin, Smeaton and Doherty [10] have created an illustration of the data sizes of several types of lifelog data, which can be seen in Table A.1.

TABLE A.1
AN ILLUSTRATION OF THE DATA QUANTITIES AND DATA SIZES FOR A SELECTION OF LIFELOG DATA OVER A DAY, YEAR AND A LIFETIME (TYPICAL 85 YEAR JAPANESE LIFESPAN) [10]

[8] https://www.microsoft.com/en-us/research/project/sensecam
[9] https://www.google.com/glass/start/

Content Type	Volume/day	Volume/year	Volume/lifetime
HD Video	5,840 hours	32.8TB	2.65PB
Autographer Camera	1.1 million images	479.6GB	40.8TB
Audio (mono - 22KHz)	5,840 hours audio	227.8GB	19.4TB
Microsoft SenseCam	1.65 million images	30.2GB	2.6TB
Accelerometer (1 Hz)	21 million readings	0.05GB	4.25GB
Locations (0.2 Hz)	3.9 million GPS points	0.01GB	1TB
Bluetooth Interactions	± 150,000 encounters	2GB+	150GB
Reading Log	User dependent	1GB+	80GB

Although intimidating at first, recent developments and future prospects in the field of storage technologies show that the size of lifelogging data does not form a major obstacle. The main challenges remaining in this field are processing of such a large and diverse data set, obtaining relevant information from it and visualizing this in a way that allows for easy interfacing by the lifelogger. This has led to the development of multiple lifelog retrieval applications in recent times, some of which will be discussed in the following section.

E. Lifelogging Applications

Over the years, multiple lifelog retrieval applications have been created for the purpose of querying large datasets with optimal speed and efficiency. These systems face off in the annual Lifelog Search Challenge (LSC) [24], in which each system is tasked with the retrieval of an image based on a textual description. In this contest, state-of-the-art lifelog retrieval systems compete against each other to evaluate their systems. Most of the retrieval systems discussed below have performed well in previous LSC challenges:

lifeXplore: The lifeXplore system created by Leibetseder et al. [25] is based on the diveXplore system for video retrieval [26]. Because of this, the lifeXplore system borrowed many of it features from video retrieval. It preprocesses its image data by grouping sequential frames together, creating one video per day. These videos are then partitioned in a sequence of scenes. Querying techniques include similarity search, drag drop filters and geolocation filtering.

VIRET: The VIRET tool by Kovalčík et al. [27] works similarly to the lifeXplore system, as both found their origins in the domain of video retrieval. Both systems also treat their image data as frames of a video. It has obtained a third place at the LSC'18 by using exclusively visual data, after which is has been improved and tailored to more lifelog specific requirements.

vitrivr: The vitrivr stack is a multimedia retrieval stack that once again has been originally designed for video retrieval [28], but has since expanded to multiple other domains [29], including lifelogging [30]. The vitrivr system is split in three mayor components:

1) *Vitrivr NG*, the user interface used for query formulation and result presentation.
2) *Cineast*, the retrieval engine that handles query processing and feature extraction.
3) *Cottontail DB*, the database that stores all information.

VRLE: The Virtual Reality Lifelog Explorer by Duane et al. [31] [32] [33] is the first system examining the feasibility and potential of lifelog retrieval in VR. The user interfaces with the application through two controllers and a Head-Mounted Display (HMD). The system has proven itself at LSC'18[10], where it was the top performer. Since then, the system has been improved and new features have been added. The field of lifelog retrieval in VR remains mostly unexplored, and is the main focal point of the current research.

Ouwehand & Van Abeelen: The thesis project by Ouwehand [34], which has been expanded upon by Van Abeelen [35], also presents a system that brings lifelogging to a VR environment. Their focus however, was more on user experience and casual browsing as opposed to efficiency like the aforementioned systems. Neither system has therefore participated in a Lifelog Search Challenge. Ouwehand projected a floor map on which pins were placed, each of them representing one or more photos (Figure A.2). Van Abeelen expanded on this by adding a spacial dimension that displays the time of each photo in the same view, as opposed to geospatial data exclusively.

Observations: Except for the VR retrieval applications, most lifelog retrieval systems have the common theme that they are adaptations of existing systems, primarily in the domain of video retrieval. These applications have the advantage that existing and proven concepts can easily be transferred to the domain of lifelogging. At the same time, these systems are bound to an existing framework, making it more difficult to implement more drastic improvements and innovations. Designing a lifelog retrieval application from scratch will open up more windows for the exploration of new features. Therefore, in this research the choice has been made to not use any existing lifelog or multimedia retrieval applications.

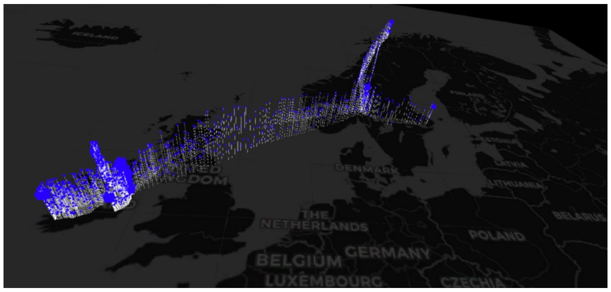


Fig. A.2. Screenshot of the lifelogging application made by Ouwehand [34]. Each pin represents a photo, which on larger zoom levels are clustered based on location.

F. Data Set

In order to evaluate and test our implementation, the LSC'20 Data Set [24] is used. This multimodal data set consists of six months of life experience from a single lifelogger. The main focal point of this set is the image data, as roughly one photo is taken for every minute using a wearable camera. Each

[10] http://lsc.dcu.ie/2018/

photo is timestamped and provided with GPS coordinates. Visual concepts and attributes have been extracted and are attached to each photo. Additional information in the form of biometrics is also provided, such as heart rate and calories burnt. This data will not be used in this research. This data set is chosen as it is the state-of-the-art at the moment of writing. It is also used in the upcoming LSC'21 session, in which the best lifelog systems will compete against each other based on their querying efficiency. Participation in this event could provide valuable insights about the application, however, due to the limitations of the system and schedule of the challenge, participation is not an option.

G. Event Segmentation

As mentioned in Section A-D, the size and diversity of lifelog data can prove difficult when attempting to access it. A continuous stream of raw lifelog data in itself is not suitable for information retrieval due to various reasons. The existence of a *semantic gap*[11] [36] combined with the fact that certain types of sensory data is not searchable using standard types of information retrieval (e.g. speed, heart rate) makes it near impossible to formulate suitable queries for lifelog data retrieval. One way to overcome these barricades is by enhancing lifelogging data though *event segmentation* and *annotation*. Event segmentation can be defined as "*the process by which people parse a continuous stream of activity into meaningful events*" [37], and is a widely explored field extending far beyond the context of lifelogging. Research has shown that individuals that are better able to mentally segment an activity into events, are better able to remember it later [38]. Replicating the psychological event segmentation of the lifelogger in the archive therefore allows for more effective and efficient querying.

Event Segmentation in Lifelogging: Event Segmentation has been a recurring theme in multiple lifelogging applications. Doherty et al. [39] have segmented lifelogs using visual data, temperature and light sensors, and an accelerometer. Gupta and Gurrin implemented a system that uses visual data exclusively [40], utilizing the Caffe framework [41]. In this research, we considered the Contextual Event Segmentation (CES) by del Molino et al [42]. Their implementation utilizes a Visual Context Predictor (VCP), consisting of a type of Recurring Neural Network that is trained to predict the visual feature of either the previous or next image frame. A boundary detector then uses the VCP to detect boundaries of events by finding local maxima in visual context changes. Therefore, smaller changes in visual context such as looking in a different direction can still be considered belonging to the same event. At the end, noise frames are removed as short interruptions in the middle of events should be ignored. An example of CES in action can be seen in Figure A.3.
Due to time restrictions, event segmentation was ultimately not implemented in the application.

H. Image Classification

The key component of any lifelog retrieval system is the querying architecture lying behind it. One of the goals of our implementation includes the ability for the user to formulate queries as precise, accurate and complete as possible. To facilitate this, we classify our lifelog data in three different dimensions: Geospatial, Temporal and Conceptual, each of which can be filtered on separately.

Geospatial Classification: Geospatial Classification allows the user to select or deselect images based on their GPS location. Querying by location allows the user to find images using the recollection of its (approximate) location. In this research, images are represented in clusters on a 2D map placed perpendicular to the user. Similar approaches have been done by Ouwehand [34] and Van Abeelen [35], though they have opted for a floor-based map instead.

Temporal Classification: Each image in the LSC'20 [24] data set is labeled with a date and time. Roughly one image is taken per minute of every hour the lifelogger was awake, over an extended period of time. A timeline was considered to allow the user to filter by time. Outside the domain of lifelogging, Karlsson et al. [43] introduced the concept of a multiscale timeline for mobile photo albums. Photos are clustered and placed on a single timeline, which the user can scale by pinching their fingers. The user can select clusters by tapping their thumbnail which either decreases the scale of the timeline, or displays a cluster if the number of photos are below a given threshold (Figure A.4). We decided to use a button-based filtering interface instead.

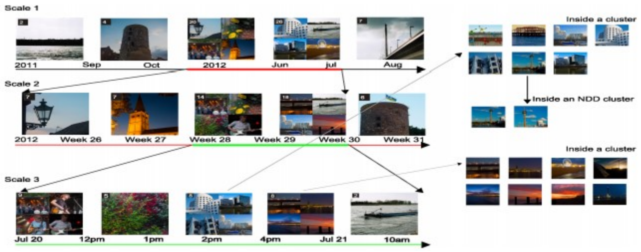
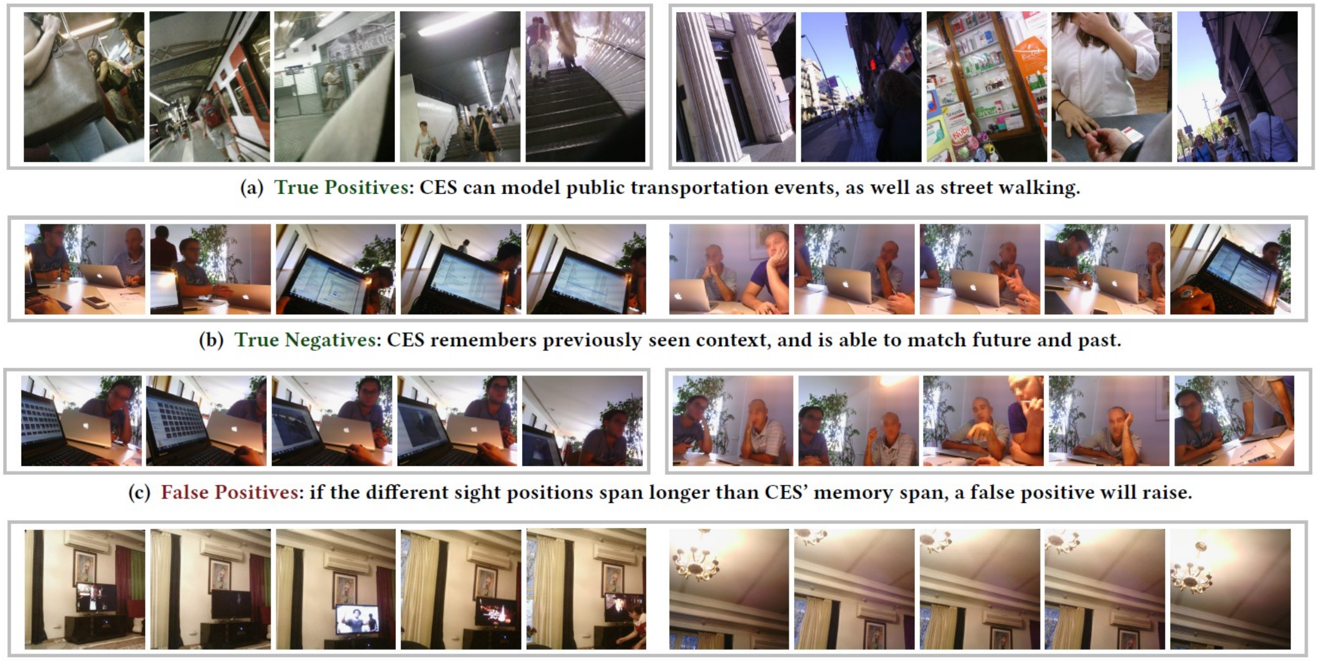


Fig. A.4. The multiscale timeline by Karlsson et al. [43]

Conceptual Classification: The ability to filter by concept requires an intermediate step, that bridges the *semantic gap* that exists between a raw image and the interpretation of the user. Images can be remembered by the presence of certain objects or people, or more abstract attributes such as the weather or the color of a building. Visual concepts and attributes are to be extracted from images and represented as tags that the user can understand. The automated annotation of lifelog data (primarily images) is an important step as manual annotation is not feasible due to the amount of data. Computer Vision techniques are utilized to provide quick and efficient image annotations, right after the image was taken [44]. These annotations can then be used for filtering by concept, obscuring objects to promote privacy [45], localization [46] and more. The LSC'20 [24] data set is already annotated with concepts, categories and attributes, which in this research are all combined as "tags".

[11] The semantic gap is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation.



(a) True Positives: CES can model public transportation events, as well as street walking.

(b) True Negatives: CES remembers previously seen context, and is able to match future and past.

(c) False Positives: if the different sight positions span longer than CES' memory span, a false positive will raise.

(d) False Negatives: two events taking place in the same location can sometimes be understood as a single one.

Fig. A.3. Examples of the capabilities of the Contextual Event Segmentation (CES) [42]. Detected events are framed in seperate boxes.

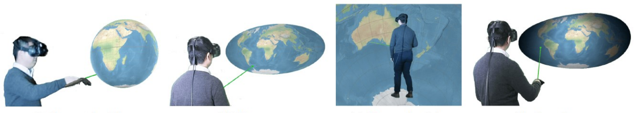
Interaction

Over the years, research has focused on new interaction devices and techniques in relation to VR, with most focus being put on haptic design [47] and body-based interaction [48]. Most of these devices and techniques are not available to consumers, and are often not proven outside the scope of their research. Existing lifelogging applications in VR have therefore opted for commercially available devices like the *HTC Vive®* controllers [32] [34] [35]. When looking at our Query Interface, which in large part consists of a geospatial map, this type of device-based interaction proves itself as a good option. Research has shown that device-based interaction through controllers has an increased performance in terms of time and error rate compared to body-based interaction, when used in the context of Geographical Information Systems (GIS) [49]. In addition, map interfacing in VR through controllers has also proven to hold great benefits over interaction with desktop-based maps [50], showing that a potential loss of performance due to a different type of interaction device is not a large concern.

I. Filtering Interface

Our filtering interface is used to view and apply filters to the lifelogging data. This interface is split in three smaller interfaces, one for each classification type

Geospatial Filtering: The central component of our application is the 2D map through which images can be filtered geospatially. Ouwehand [34] and Van Abeelen [35] have chosen for a flat, floor-based map (Figure A.2), due to the familiarity most people have with flat maps (e.g. Google Maps[12]). We have chosen to place our map perpendicular to the user instead, as the amount of vertical head movement required to use a floor-based map might strain the neck. Research has shown that both an exocentric globe (Figure A.5.a) and curved map (Figure A.5.c) have significant advantages over a flat map when estimating geographical properties such as size and distance [51]. Virtual globes have been a well-studied subject both in- and outside the domain of VR [52], and curved (panorama) maps have been utilized in several domains unrelated to lifelogging [53] [54], showing promising results. Ultimately, we decided to not further explore this domain, as focus shifted to data clustering and visualization instead. Investigating various map types for the purpose of lifelog retrieval may nonetheless provide an interesting topic for future work.



(a) Exocentric globe (b) Flat map (c) Egocentric globe (d) Curved map

Fig. A.5. Four types of geographical map visualization as proposed by Yang et al. [51]

Geospatial Clustering

A lifelog is characterized by its sheer amount of data, as can be seen in Table A.1. Representing each image on a map will result in large amounts of clutter. Utilizing event segmentation partially remedies this issue, but the remaining amount of events could still render the map unusable. Therefore, some level of clustering is required to group images based on their location, in order to aid in the clarity of the map. Previous applications in VR lifelogging that used a map representation [34] [35] have clustered based exclusively on

[12] https://www.google.com/maps

the geographical location of individual photos. This approach has its disadvantages as high density areas are at risk of becoming cluttered, and two geospatially adjacent images could end up in different clusters based on the rounding of their respective GPS coordinates

In our research, we present and compare two existing techniques for the visualization of clusters, namely Markers and Heatmap visualization, and apply the same clustering algorithm to the two of them

Clustering: For our application, we require an algorithm that is able to cluster images based on their geographical distance and adapt in real time to changes to the filtering parameters and zoom level of the map. Generalized clustering algorithms exist that may be suited to the clustering of lifelogging data specifically [55] [56]. These algorithms can be subdivided into multiple categories such as hierarchical clustering [57], partitional clustering such as *k-means* [58], density based clustering [59] and more. For this research the initial choice was made to use Agglomerative Complete-Link Clustering [60], which is a hierarchical clustering algorithm that works as follows:

1) Place each pattern in its own cluster. Construct a list of interpattern distances for all distinct unordered pairs of patterns, and sort this list in ascending order.
2) Step through the sorted list of distances, forming for each distinct dissimilarity value d_k a graph of the patterns where pairs of patterns closer than d_k are connected by a graph edge. If all the patterns are members of a completely connected graph, stop.
3) The output of the algorithm is a nested hierarchy of graphs which can be cut at a desired dissimilarity level forming a partition (clustering) identified by completely connected components in the corresponding graph.

Each pattern consist of an image, and the distance function is the Euclidean distance between the location centers of two clusters p and q:

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2} \tag{1}$$

The result is a *dendrogram*, which is a tree-like diagram that displays each pattern on top of the tree, with the remaining nodes representing clusters to which the data belongs, with arrows displaying the distance (Figure A.8). This distance is 0 on top of the tree, and increases to the point where every pattern is contained in a single cluster. This property makes a dendrogram suitable for supporting a zooming feature, as no further calculations need to be done, thus restricting calls to the algorithm to querying exclusively. The time complexity of the algorithm equals $O(n^2 \log n)$, with a space complexity of $O(n^2)$, with n equal to the number of segmented events. However, due to the algorithm being entirely sequential, it was ultimately too slow to provide real-time clustering and a different algorithm was chosen.

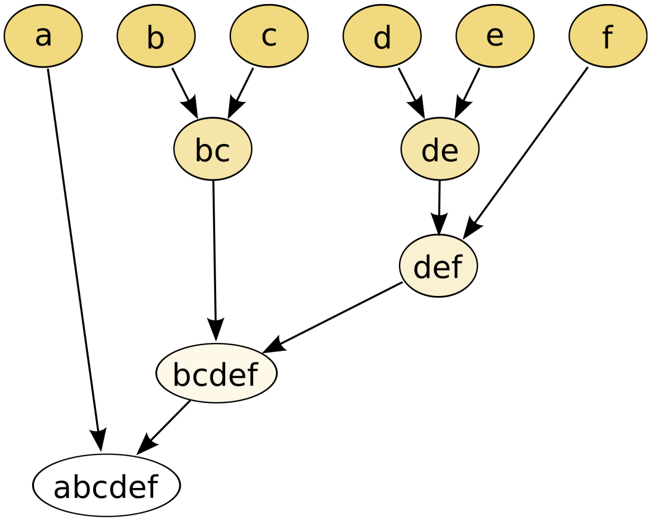


Fig. A.6. An example dendrogram.

The choice fell on the density-based DBSCAN [61] algorithm instead:

```
void CreateBins(){
foreach(Image i in images){
if (bins.Contains(i.coordinates)
    bins[i.coordinates].Add(i)
else { Bin b = new Bin();
b.Add(i)
bins.Add(i.coordinates, b) }
}}

void CreateDistances(){
foreach(Image i in images){
foreach(Image j in images){
if (j == i) continue
distanceMatrix[i,j].Add(
    EuclideanDistance(i,j)
}}}

void CreateNeighbourSets(){
foreach(Bin b1 in bins){
foreach(Bin b2 in bins){
if (b1 == b2) continue
if (distanceMatrix[b1,b2] < epsilon)
    b1.neighbours.add[b2]
}

void CreateDBSCAN(){
foreach(bin b in bins)
if (b.inCluster) continue
Cluster c = new Cluster
c.Add(b)
b.inCluster = true
AddNeighbours(c,b)
}}}

void AddNeighbours(Cluster c, Bin b){
foreach(Bin d in b.neighbours){
d.inCluster = true
```

```
    c.Add(d)
    addNeighbours(c,d)
    }}
```

DBSCAN clusters images based on the distance to their neighbors. Epsilon (ϵ) is a predetermined variable that indicates the minimum distance for two bins to be considered neighbors. In our application, ϵ is based on the current zoom level, high when zoomed out, and low when zoomed in. A second DBSCAN variable, $minPoints$, which indicates the minimum size of clusters to not be considered an outlier, is set to zero and is thus not used. The algorithm is $O(n^3)$ (n equals the number of bins) in the CreateNeighborSets() function if the maximum value in the distance matrix is smaller than ϵ. This and the CreateDistances() function are further optimized by using an Unity ComputeShader. The distanceMatrix is $O(n^2)$ in size. The CreateDBSCAN() and AddNeighbors() functions are the only two that can not be preprocessed, taking $\Theta(n)$ in total.

Marker Visualization: Markers are often used as a representation of points on interest in applications as Google Maps and third-party libraries like *marker clustering* [62], offering a clear overview of the locations of individual clusters. Markers are sensitive to high-density areas, either causing clutter or a loss of information based on the amount of markers.

Heatmap Visualization: Heatmaps visualizes clusters on a square 2D overlay placed over the geospatial map. Heatmaps can be continuous or discrete, with the latter using a 2D grid to indicate density. The former approach shows large similarities to the isopleth map visualization by Toyama et al. [63]. Heatmaps are sensitive to zoom levels as high-density areas could become a big blur when zoomed out. Unlike Markers, the Heatmap is better able to display high density areas, and is better in visualizing them when exact locations are not needed.. A similar clustering approach Toyama, known as Media Dots is not considered due to the large similarities to the discrete version of this approach. Our research will use a continuous heatmap in the form of a shader.

J. Temporal Filtering

Users can filter on date or time using an interface panel on the right of the map. Van Abeelen added a third dimension to his floor-based map to represent the temporal information of images [35], and Duane created a button-based system for date and time selection in his application [32]. While using simple buttons has its merits, there are other methods of filtering by date and time. The Flickr Cities project[13] uses a timeline to query by month, and two circular interfaces to filter by hour and weekday. Our implementation presents two interfaces, one inspired by the simple button system (*Buttons*), and one comprised of a calendar and two circular clocks for date/time selection (*C&C*).

K. Conceptual Filtering

Users can filter by concept or tag by using a panel on the left of the map. Using a virtual keyboard, they can type out (partial) tag names which are auto-completed by our system for easier use. Speicher et al. [64] have tested several forms of textual input in VR, both body-based and device-based. Results show that pointing on a keyboard with a controller outperforms the other methods that they tested, which has been implemented in our *TagList* interface. Boletsis and Kongsvik present a drum-like VR keyboard [65] which can be seen in Figure A.7. In this input method, controllers are used as sticks which through downward movements, "press" the keys of the virtual keyboard, thus providing an alternative method of text input. This method is not implemented in our research.

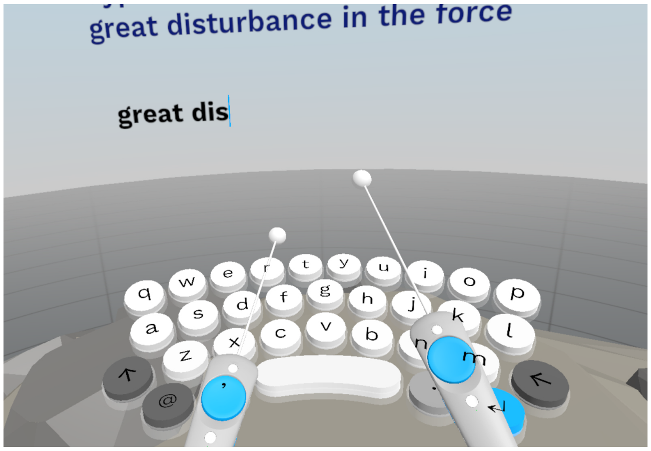


Fig. A.7. The drum-like VR Keyboard by Boletsis and Kongsvik [65]

L. Image Visualization

A separate interface is required to inspect individual images once filtered out. Duane created a so-called *memory wall* in his VR lifelogging application [32], which displays images in a grid perpendicular to the user. However, previous research has shown benefit to the use of peripheral vision in different contexts [51] [53] in VR. Our research will therefore use an interface that is partially curved around the user, allowing a large amount of images to be displayed with equal quality.

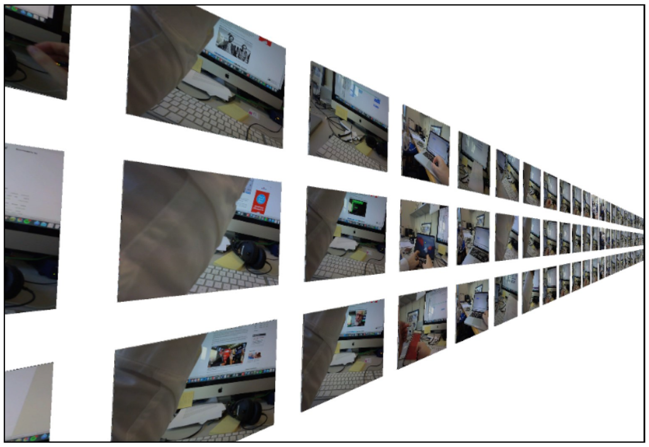


Fig. A.8. The *memory wall* from Duane's lifelogging application, which is displayed as a grid perpendicular to the user [32].

M. Conclusion

In this literature study, we have looked at the topic of lifelogging: its history, benefits and limitations. After that, we

[13] http://www.datainterfaces.org/projects

put our focus on existing lifelog retrieval applications, multiple of which have been discussed, including existing applications in Virtual Reality. The concept of event segmentation is discussed, including its uses for lifelog retrieval. To allow for complete and accurate retrieval queries, our application allows the filtering of images in three dimensions: Geospatial, Temporal and Conceptual. This research focused on the design of filtering interfaces related to the three dimensions. Finally, techniques for visualizing images were briefly discussed, which participants used to inspect individual images.

Appendix B: References

[10] C. Gurrin, A. F. Smeaton, and A. R. Doherty, "Lifelogging: Personal big data," *Foundations and trends in information retrieval*, vol. 8, no. 1, pp. 1–125, 2014.

[11] V. Bush *et al.*, "As we may think," *The atlantic monthly*, vol. 176, no. 1, pp. 101–108, 1945.

[12] S. Mann, "Continuous lifelong capture of personal experience with eyetap," in *Proceedings of the the 1st ACM workshop on Continuous archival and retrieval of personal experiences*, pp. 1–21, 2004.

[13] S. Mann, J. Nolan, and B. Wellman, "Sousveillance: Inventing and using wearable computing devices for data collection in surveillance environments.," *Surveillance & society*, vol. 1, no. 3, pp. 331–355, 2003.

[14] J. Gemmell, G. Bell, R. Lueder, S. Drucker, and C. Wong, "Mylifebits: fulfilling the memex vision," in *Proceedings of the tenth ACM international conference on Multimedia*, pp. 235–238, 2002.

[15] J. Gemmell, G. Bell, and R. Lueder, "Mylifebits: a personal database for everything," *Communications of the ACM*, vol. 49, no. 1, pp. 88–95, 2006.

[16] M. Dodge and R. Kitchin, "'outlines of a world coming into existence': pervasive computing and the ethics of forgetting," *Environment and planning B: planning and design*, vol. 34, no. 3, pp. 431–445, 2007.

[17] A. J. Sellen and S. Whittaker, "Beyond total capture: a constructive critique of lifelogging," *Communications of the ACM*, vol. 53, no. 5, pp. 70–77, 2010.

[18] M. A. Barrett, O. Humblet, R. A. Hiatt, and N. E. Adler, "Big data and disease prevention: from quantified self to quantified communities," *Big data*, vol. 1, no. 3, pp. 168–175, 2013.

[19] J. R. Whitson, "Gaming the quantified self," *Surveillance & Society*, vol. 11, no. 1/2, pp. 163–176, 2013.

[20] V. Rivera-Pelayo, V. Zacharias, L. Müller, and S. Braun, "Applying quantified self approaches to support reflective learning," in *Proceedings of the 2nd international conference on learning analytics and knowledge*, pp. 111–114, 2012.

[21] E. Tulving, "Elements of episodic memory," 1983.

[22] A. J. Sellen, A. Fogg, M. Aitken, S. Hodges, C. Rother, and K. Wood, "Do life-logging technologies support memory for the past? an experimental study using sensecam," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 81–90, 2007.

[23] S. Hodges, L. Williams, E. Berry, S. Izadi, J. Srinivasan, A. Butler, G. Smyth, N. Kapur, and K. Wood, "Sensecam: A retrospective memory aid," in *International conference on ubiquitous computing*, pp. 177–193, Springer, 2006.

[24] C. Gurrin, T.-K. Le, V.-T. Ninh, D.-T. Dang-Nguyen, B. T. Jónsson, J. Lokoč, W. Hurst, M.-T. Tran, and K. Schoeffmann, "An Introduction to the Third Annual Lifelog Search Challenge, LSC'20," in *ICMR '20, The 2020 International Conference on Multimedia Retrieval*, (Dublin, Ireland), ACM, 2020.

[25] A. Leibetseder and K. Schoeffmann, "lifexplore at the lifelog search challenge 2020," in *Proceedings of the Third Annual Workshop on Lifelog Search Challenge*, pp. 37–42, 2020.

[26] K. Schoeffmann, B. Münzer, A. Leibetseder, J. Primus, and S. Kletz, "Autopiloting feature maps: the deep interactive video exploration (diveexplore) system at vbs2019," in *International Conference on Multimedia Modeling*, pp. 585–590, Springer, 2019.

[27] G. Kovalčík, V. Škrhak, T. Souček, and J. Lokoč, "Viret tool with advanced visual browsing and feedback," in *Proceedings of the Third Annual Workshop on Lifelog Search Challenge*, pp. 63–66, 2020.

[28] L. Rossetto, I. Giangreco, C. Tanase, and H. Schuldt, "vitrivr: A flexible retrieval stack supporting multiple query modes for searching in multimedia collections," in *Proceedings of the 24th ACM international conference on Multimedia*, pp. 1183–1186, 2016.

[29] R. Gasser, L. Rossetto, and H. Schuldt, "Multimodal multimedia retrieval with vitrivr," in *Proceedings of the 2019 on International Conference on Multimedia Retrieval*, pp. 391–394, 2019.

[30] S. Heller, M. Amiri Parian, R. Gasser, L. Sauter, and H. Schuldt, "Interactive lifelog retrieval with vitrivr," in *Proceedings of the Third Annual Workshop on Lifelog Search Challenge*, pp. 1–6, 2020.

[31] A. Duane and C. Gurrin, "Lifelog exploration prototype in virtual reality," in *International Conference on Multimedia Modeling*, pp. 377–380, Springer, 2018.

[32] A. Duane, "Visual access to lifelog data in a virtual environment," *Dublin City University*, 2019.

[33] A. Duane, B. Þór Jónsson, and C. Gurrin, "Vrle: Lifelog interaction prototype in virtual reality: Lifelog search challenge at acm icmr 2020," in *Proceedings of the Third Annual Workshop on Lifelog Search Challenge*, pp. 7–12, 2020.

[34] K. Ouwehand, "Geospatial access to lifelogging images in vr," *Utrecht University*, 2019.

[35] J. van Abeelen, "Visualising lifelogging data in temporal virtual reality environments," *Utrecht University*, 2019.

[36] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 12, pp. 1349–1380, 2000.

[37] J. M. Zacks and K. M. Swallow, "Event segmentation," *Current directions in psychological science*, vol. 16, no. 2, pp. 80–84, 2007.

[38] J. M. Zacks, N. K. Speer, J. M. Vettel, and L. L. Jacoby, "Event understanding and memory in healthy aging and dementia of the alzheimer type.," *Psychology and aging*, vol. 21, no. 3, p. 466, 2006.

[39] A. R. Doherty, A. F. Smeaton, K. Lee, and D. P. Ellis, "Multimodal segmentation of lifelog data," 2007.

[40] R. Gupta and C. Gurrin, "Approaches for event segmentation of visual lifelog data," in *International Conference on Multimedia Modeling*, pp. 581–593, Springer, 2018.

[41] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*, pp. 675–678, 2014.

[42] A. Garcia del Molino, J.-H. Lim, and A.-H. Tan, "Predicting visual context for unsupervised event segmentation in continuous photo-streams," in *Proceedings of the 26th ACM international conference on Multimedia*, pp. 10–17, 2018.

[43] K. Karlsson, W. Jiang, and D.-Q. Zhang, "Mobile photo album management with multiscale timeline," in *Proceedings of the 22nd ACM international conference on Multimedia*, pp. 1061–1064, 2014.

[44] P. Wang, L. Sun, A. F. Smeaton, C. Gurrin, and S. Yang, "Computer vision for lifelogging: Characterizing everyday activities based on visual semantics," in *Computer Vision for Assistive Healthcare*, pp. 249–282, Elsevier, 2018.

[45] R. Hasan, P. Shaffer, D. Crandall, E. T. Apu Kapadia, *et al.*, "Cartooning for enhanced privacy in lifelogging and streaming videos," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 29–38, 2017.

[46] V. Bettadapura, I. Essa, and C. Pantofaru, "Egocentric field-of-view localization using first-person point-of-view devices," in *2015 IEEE Winter Conference on Applications of Computer Vision*, pp. 626–633, IEEE, 2015.

[47] N. Magnenat-Thalmann and U. Bonanni, "Haptics in virtual reality and multimedia," *IEEE MultiMedia*, vol. 13, no. 3, pp. 6–11, 2006.

[48] P. Maes, T. Darrell, B. Blumberg, and A. Pentland, "The alive system: Wireless, full-body interaction with autonomous agents," *Multimedia systems*, vol. 5, no. 2, pp. 105–112, 1997.

[49] A. Santos-Torres, T. Zarraonandia, P. Díaz, and I. Aedo, "Exploring interaction mechanisms for map interfaces in virtual reality environments," in *Proceedings of the XIX International Conference on Human Computer Interaction*, pp. 1–7, 2018.

[50] W. Dong, T. Yang, H. Liao, and L. Meng, "How does map use differ in virtual reality and desktop-based environments?," *International Journal of Digital Earth*, pp. 1–20, 2020.

[51] Y. Yang, B. Jenny, T. Dwyer, K. Marriott, H. Chen, and M. Cordeil, "Maps and globes in virtual reality," in *Computer Graphics Forum*, vol. 37, pp. 427–438, Wiley Online Library, 2018.

[52] B. T. Tuttle, S. Anderson, and R. Huff, "Virtual globes: an overview of their history, uses, and future challenges," *Geography Compass*, vol. 2, no. 5, pp. 1478–1505, 2008.

[53] O.-H. Kwon, C. Muelder, K. Lee, and K.-L. Ma, "A study of layout, rendering, and interaction methods for immersive graph visualization,"

IEEE transactions on visualization and computer graphics, vol. 22, no. 7, pp. 1802–1815, 2016.

[54] M. Mengerink, "Geospatial image browsing in virtual reality," *Utrecht University*, 2019.

[55] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM computing surveys (CSUR)*, vol. 31, no. 3, pp. 264–323, 1999.

[56] L. Rokach and O. Maimon, "Clustering methods," in *Data mining and knowledge discovery handbook*, pp. 321–352, Springer, 2005.

[57] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: an overview," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 2, no. 1, pp. 86–97, 2012.

[58] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, pp. 281–297, Oakland, CA, USA, 1967.

[59] H.-P. Kriegel, P. Kröger, J. Sander, and A. Zimek, "Density-based clustering," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 3, pp. 231–240, 2011.

[60] B. King, "Step-wise clustering procedures," *Journal of the American Statistical Association*, vol. 62, no. 317, pp. 86–101, 1967.

[61] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise.," in *kdd*, vol. 96, pp. 226–231, 1996.

[62] R. Netek, J. Brus, and O. Tomecka, "Performance testing on marker clustering and heatmap visualization techniques: A comparative study on javascript mapping libraries," *ISPRS International Journal of Geo-Information*, vol. 8, no. 8, p. 348, 2019.

[63] K. Toyama, R. Logan, and A. Roseway, "Geographic location tags on digital images," in *Proceedings of the eleventh ACM international conference on Multimedia*, pp. 156–166, 2003.

[64] M. Speicher, A. M. Feit, P. Ziegler, and A. Krüger, "Selection-based text entry in virtual reality," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2018.

[65] C. Boletsis and S. Kongsvik, "Text input in virtual reality: A preliminary evaluation of the drum-like vr keyboard," *Technologies*, vol. 7, no. 2, p. 31, 2019.

Appendix C
Implementation Details

Hardware

Target hardware is the HTC Vive®:

TABLE C.1
Technical specifications for the HTC Vive®

Display	OLED
Resolution	$1080\times1200pxpereye$
Refresh rate	90hz
Field of View	110°
Lens Type	Fresnel
IPD	60.8-74.6mm
Tracking Area	$15\times15ft$

Software

The system was made in Unity 2020.1.6f1.

External Libraries

- OpenVR XR Plugin: 1.1.4
- SteamVR 2.7.3
- Unity UI 1.0.0
- TextMexhPro 3.0.1

Appendix D
Tasks

A. *Map Interface*

1) How many images can you see on the map in total? (Answer: 174018)
2) Which country contains the most images? (Answer: Ireland)
3) How many images have been taken in Turkey? (Answer: 178)
4) The lifelogger once took a flight to Russia from Germany. Which countries did he cross? (Answer: Czech Republic, Poland, Belarus)
5) How many images have been taken in China? (Answer: 5331)

B. *Temporal Interface*

All images not taken in China are filtered out.

1) During which months and years was the Lifelogger in China? (Answer: March 2015 May 2018)
2) How many days did he spend in China? (Answer: 5)
3) In March he spent a only a single day in China, what day was that? (Answer: Friday the 20th)
4) What was the last day the Lifelogger spent in China? (Answer: May 25th 2018)
5) How many images were not taken before 6:00? (Answer: 5055)

C. *Tag Interface*

All images not taken in China during May 2018 are filtered out.

1) What tag is the most common for images in China? (Answer: *no horizon*)
2) How many images contain the tag *research*? (Answer: 39)
3) How many images do not contain the tag *airplane_cabin*? (Answer: 1930)
4) How many of those images were taken from inside a car? (Answer: 110)
5) How many images were taken inside a fastfood restaurant? (Answer: 238)

D. *Image List*

All images not taken in China during May 2018 without the *fastfood_restaurant* tag are filtered out.

1) In which fastfood restaurant is the Lifelogger eating? (Answer: McDonalds)
2) Name three items he is having at McDonalds (Answer: Fries, Burger, Drink)
3) Find an image in which the burger is visible. (Answer: ...)
4) In what coffee chain does the lifelogger order his coffee (after the second restaurant) (Answer: Starbucks)
5) What is the name of the street where this is located? (Answer: Hanjie Street)

Appendix E
Queries

A. *Set 1*

1) Q1.1:

- A red car on a driveway. ...
- ... It was beside a white house. ...
- ... on a cloudy day. ...
- ... There was a man standing on the driveway, ...
- ... along with some potted plants. ...
- ... I could also see a tree in the background.

2) Q1.2:

- Checking out of a hotel in Norway...
- ... in the early morning. ...
- ... There was a weather report on a TV on the right. ...
- ... There was a man behind the counter ...
- ... in front of a wooden wall. ...
- ... It was the 5th of September.

3) Q1.3:

- Pulling up grass or weeds ...
- ... in my garden in Ireland. ...
- ... I can see trees on the other side of the road, ...
- ... and several types of plants on my right. ...
- ... It was on a Saturday ...
- ... in September.

4) Q1.4:

- Walking on the tarmac of the airport at Dublin. ...
- ... I can see a person in an orange safety jacket ...
- ... while I was walking towards an airplane. ...
- ... I also saw the back of a bus with the number "105" on its back. ...
- ... It was March 2015 ...
- ... around 15:30 in the afternoon.

5) Q1.5:

- I was looking at an old clock, with flowers visible. ...
- ... It was on a table next to my bed. ...
- ... There was a blue rabbit-like creature inside of the bed. ...
- ... It was on a Monday ...
- ... in September ...
- ... at my home in Dublin.

B. *Set 2*

1) Q2.1:

- Getting some coffee at the airport in Stockholm. ...
- ... I was sitting at a small table in a lounge. ...
- ... There were other people as well, ...
- ... but most chairs were empty. ...
- ... It was August 2016 ...
- ... around 14:10 in the afternoon.

2) Q2.2:

- I was in my office in Ireland taking a skype call on my laptop. ...
- ... It was on a Friday in September, ...
- ... just before 9:00. ...

- ... I can clearly see a large image of a man's face on the screen. ...
- ... My hand may have been covering part of the keyboard. ...
- ... It was the 9th of September.

3) Q2.3:

- Walking in the cabin of an airplane in Shanghai. ...
- ... There was a man with an orange shirt in front of me. ...
- ... He also wore a backpack.
- ... People started taking their seats. ...
- ... It was May 2018 ...
- ... around 15:15. ...

4) Q2.4:

- Waiting at the reception of the Yeats Country Hotel in Ireland. ...
- ... The wall behind the counter was full of clocks. ...
- ... There was a man with a gray sweater in front of me. ...
- ... I remember it being a wooden wall. ...
- ... It was September 2016, ...
- ... around 10:15 in the morning.

5) Q2.5:

- Sitting at a wooden table in an antiques store in the UK. ...
- ... I was having some cake ...
- ... with a cup of tea ...
- ... and a small bottle of milk. ...
- ... It was a Saturday ...
- ... in March 2015.

Appendix F
Evaluation

A. Demographics

The following questionnaire is given to the participants at the start of the experiment. It is used to gather basic demographic information about the participants.

Age:

- Below 18
- 18 - 24
- 25 - 34
- 35 - 44
- 45 - 54
- 55 - 64
- Above 65

Due to the experiment taking place at Utrecht University, it is expected that the majority of will be students, and fall within the 18 - 24 age group. Other participants, such as university staff, are expected to be distributed evenly among the older age groups.

Gender:

- Male
- Female
- Other/do not want to disclose

A 50/50 gender distribution between male and female would be ideal to both identify differences between the two genders, and nullify any gender related effect on the results of the study. The expectation is however that the dominant gender will be male due to the background of most participants, and the difficulty of recruiting a diverse set of participants due to the Covid-19 pandemic.

Primary Language:

- Dutch
- English
- Other

It is expected that most participants will have Dutch as a primary language, even though the implementation and experiment itself are both in English. It is unlikely that a language barrier will affect performance.

Experience with Virtual Reality (VR):

- None
- I have used VR before, but not often
- I use VR regularly (1+ times per month)
- I use VR often (1+ times per week)

It is expected that most participant have some, if limited experience with VR. Those with little to no experience with VR are expected to perform worse on average due to inexperience with the hardware, and require a longer time to become familiar with the system.

Experience with Lifelogging:

- None
- I am aware of the concept of Lifelogging
- I actively work with the processing of Lifelog data
- I actively generate, store and process Lifelog data

It is expected that the majority of participants have never heard of Lifelogging. Those who do have experience are expected to perform better on average, due to their experience with the data used.

Motion Sickness:

- I do not suffer from motion sickness
- I occasionally suffer from (mild) motion sickness
- I often suffer from motion sickness, and the symptoms can be severe.

It is expected that the majority of participants do not suffer from motion sickness, or only experience mild symptoms. Those that have selected the bottom option are prohibited from partaking in the experiment due to the potential risks of cyber sickness, whereas those who have filled in the second option are made extra aware of the concept of cyber sickness, and are stressed upon that they must stop the experiment as soon as they experience any symptoms. This, however, is not expected to occur often due to the participants remaining seated throughout the experiment.

B. *Performance*

The following performance metrics are used:

Time per query & Attempts made: As mentioned, participants have up to 180 seconds of time, or 3 attempts per query. Submitting 3 incorrect images will result in the time being set to 180 seconds to preserve result accuracy, regardless of the actual duration.

Performance differences between two interface variations could be indicative of one interface being more cumbersome or easy to handle. It is expected that performance varies greatly per query and query set, therefore sets are not considered individually.

Time per interface: During each search query, the participant's view is recorded through a screen recording, to later determine the relative time spent in each interface. These times are then normalized by the query time, and compared to other variations of the interface. As individual search queries already lean towards using one type of interface over the over, they are not indicative on their own.

The reason why screen recording is chosen as the method for measurement is that participants tend not to completely move their head over to use other interfaces (*DateTime* and *Tag* most notably), causing a gaze tracker to not be reliable. Eye tracking technology was not available for this study. While a screen recording is not as exact as using a gaze tracker, it is easier to determine where the focus of the user lies and should therefore provide more accurate results.

Selected tags, dates & times: While not directly indicative of the usability of the interface, measuring the selected (or blacklisted) tags, dates and times gives insight on the participant's interpretation of the search query and their path to problem-solving. With hints ranging from specific (... on Monday ...) to nonspecific (... in the Morning ...), it is interesting to see participant's approach on resolving the search queries. Irrelevant tags, misclicks or other errors in the tags, dates or times could be indicative of underlying issues with either the interfaces or the data set. Only selected and blacklisted tags are logged, no keylogger is used. In addition, interaction with the Map interface is not directly logged due to the vast amount of possible interactions with the map.

Inspected & submitted images: The system will keep track of all images that the participant chooses to inspect individually, as well as those that are ultimately submitted. It is expected that the participant will only choose to closely inspect images that they think could be the correct answer. A large amount of inspections in contrast to few submissions could be indicative of the participant's inability to narrow the possibilities, having applied the wrong filters or simply the inability to see the images on the ImageView clear enough. In addition, incorrect submissions are inspected and rated on their likeliness to their correct counterparts, to identify what differences have led to the participant's choice of said submission.

User Evaluation (SUS): After the study, participants are asked to evaluate each of the filtering interfaces (Map, DateTime, Tag) individually using the System Usability Scale (SUS). The scale has been slightly adjusted to support individual interfaces instead of an entire system.

Qualitative Feedback: After the study, participants are asked to give at least one positive, and one negative on each of the filtering interfaces. In addition, special remarks made by the participants during the study are noted down.

APPENDIX G
RESULTS

A. *Demographics*

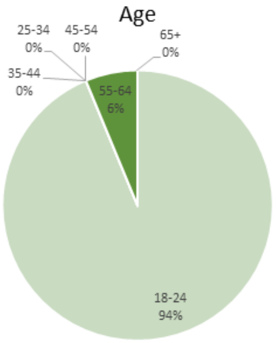


Fig. G.1. Age distribution of participants

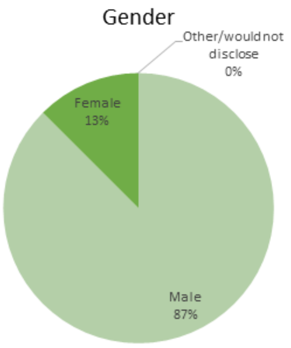


Fig. G.2. Gender distribution of participants

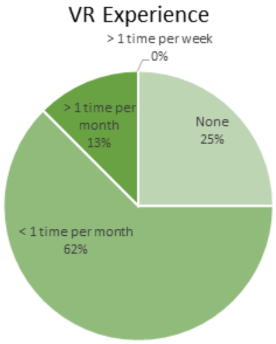


Fig. G.3. Previous VR experience of participants

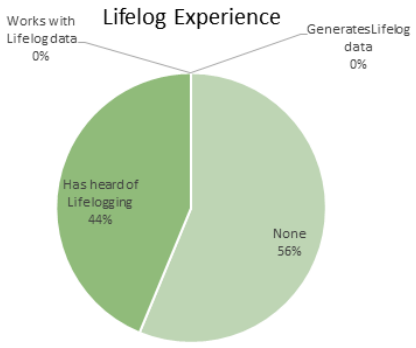


Fig. G.4. Previous Lifelog experience of participants

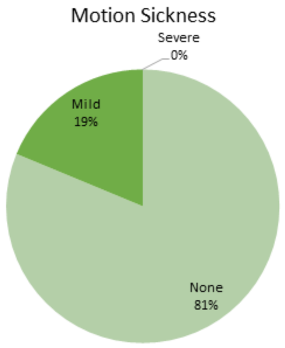


Fig. G.5. Motion sickness distribution of participants

The demographics distribution is not as even as was desired. Only one participant was not of an age between 18-24 (55-64), and only two participants were female (against 14 male). While this is to be expected due to most of the participants being students from Utrecht University (and computer science related studies), ideally a more even gender distribution is required.

Most participants have none to very limited VR experience, with only two participants being somewhat regular users. This has lead to a lot of participants, especially early in the experiment, struggling with the VR controls due to their lack of experience.

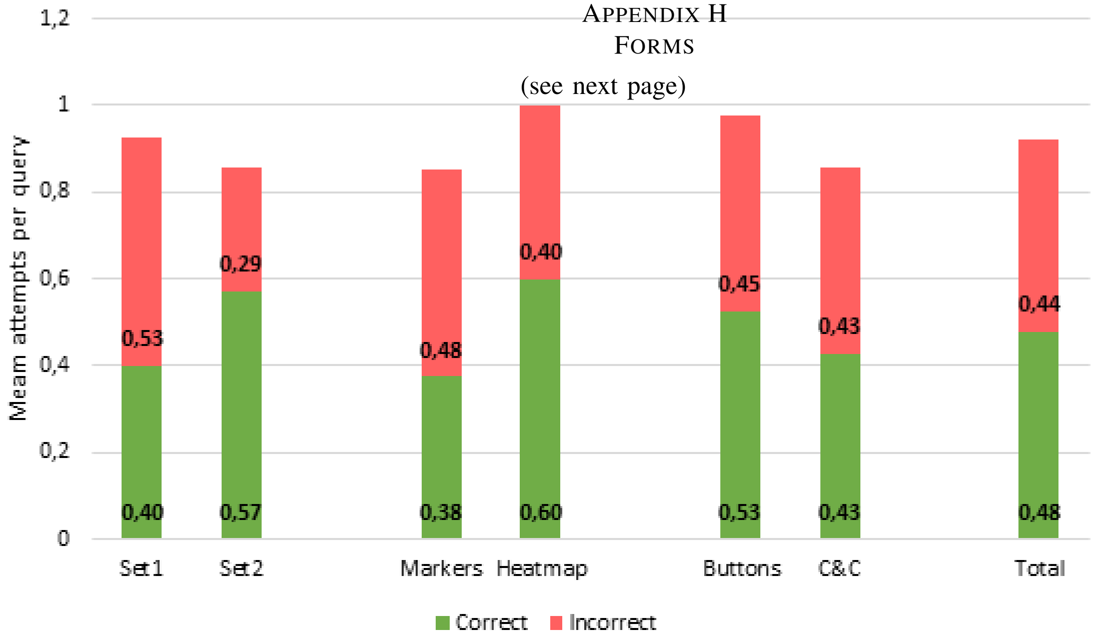
Slightly less than half of the participants has heard of the concept of lifelogging, though none mentioned the same form being used as in this study. One participant mentioned a person taking daily photos of his beard growing as lifelogging. None of the participants is actively involved in lifelogging, which is to be expected given the group of participants.

Of all participants, three mentioned suffering from mild motion sickness from time to time. None of the participants experienced any symptoms during the experiment.

B. *Time per query & Attempts made*

TABLE G.1
MEAN VALUES AND STDS FOR QUERY TIME AND (IN)CORRECT ATTEMPTS PER VARIATION

Variation	$\bar{x} \setminus \Sigma$	Query Time	Correct attempts	Incorrect attempts
Set1	$\bar{x}$	162.41	0.40	0.53
	σ	13.97	0.15	0.44
Set2	$\bar{x}$	124.45	0.57	0.34
	σ	20.87	0.34	0.22
Markers	$\bar{x}$	158.56	0.38	0.48
	σ	16.52	0.27	0.44
Heatmap	$\bar{x}$	143.43	0.60	0.40
	σ	23.29	0.20	0.26
Buttons	$\bar{x}$	148.40	0.53	0.45
	σ	25.21	0.34	0.46
C&C	$\bar{x}$	155.04	0.43	0.43
	σ	15.32	0.14	0.21
Total	$\bar{x}$	151.50	0.48	0.48
	σ	21.26	0.28	0.36



keepaspectratio!keepaspectratio!

Fig. G.6. Average (in)correct query attempts per variation

The mean values and STDs for time per query and attempts made can be seen in Table G.1. A graph visualizing query attempts can be seen in Figure G.6. In an ideal world, the average amount of (correct) answers is one, which will happen if every query is correctly answered in the first attempt. A higher average of answers indicates that participants often make (incorrect) guesses, a lower average means that not every query is answered, correct or incorrect. This graph is ultimately not used in the paper as the standard deviations for attempts made are too high to produce statistically significant results.

C. Time per interface

TABLE G.2
Mean values and STD's for time per interface, per variation

Variation	$\bar{x} \setminus \Sigma$	Map	DateTime	Tag	ImageView	None
Set1	$\bar{x}$	12.62	5.65	48.96	23.66	9.11
	σ	4.33	3.30	3.7	4.02	4.01
Set2	$\bar{x}$	21.18	7.33	38.34	22.59	10.57
	σ	5.26	3.60	5.14	7.58	4.02
Markers	$\bar{x}$	18.86	8.31	42.09	21.02	8.99
	σ	6.12	3.64	7.55	5.78	4.00
Heatmap	$\bar{x}$	14.29	4.27	44.54	25.93	10.98
	σ	6.34	1.82	8.66	5.08	3.87
Buttons	$\bar{x}$	17.66	5.85	43.51	24.40	8.58
	σ	5.36	3.82	7.99	4.05	7.40
C&C	$\bar{x}$	15.88	7.34	43.83	21.43	13.08
	σ	8.04	2.95	6.18	7.75	5.05
Total	$\bar{x}$	16.90	6.49	42.93	23.12	9.84
	σ	6.42	3.43	7.01	5.85	3.93

The mean values and STDs for time per interface (%) can be seen in Table G.2.

D. SUS Average

(See Table 1 in the paper for results)

Appendix H
Forms

(see next page)

Consent Form

Utrecht University

Participant ID

This consent form will inform you (the participant) about your rights in the upcoming experiment. I (the researcher) have explained the purpose and structure of this experiment, which is part of the Master Thesis of Hidde Veer, supervised by Wolfgang Hürst.

You are aware that during the experiment, data will be gathered about the interaction between you and the software. This includes basic demographics (gender, age, etc.), interview responses, performance data and screen captures. All data gathered at the experiment may only be for the purpose of this research, including publishment in the form of a master thesis and scientific paper. All data gathered in the experiment will be anonymized and treated confidentially.

You understand that participation in the experiment is voluntary: You may abort the experiment at any moment when you desire, and you do not have to provide a reason to us. you am aware that you will suffer no negative consequences from aborting, and that all data gathered during the experiment will be destroyed immediately, and therefore not be used in the research.

You are aware that if you decide to partake in this experiment, it is your responsibility to stop it immediately and inform us in case you experience any discomfort or unwellness, such as dizziness or motion sickness.

You may send further questions about the research to Hidde Veer (h.s.veer@students.uu.nl) or Wolfgang Hürst (huest@uu.nl).If you suspect that your rights as participant are violated, you may contact the Research Integrity Committee (vertrouwenspersoon-wi@uu.nl).

□ I (the participant) have read the and understood the above test, and I consent that data collected from my participation may be used and published in this research.

□ I confirm that the researcher has given me a copy of this form for my own use and safekeeping.

□ (Optional) I consent that the researcher may obtain pictures or videos of me using the software, to be used in promotional material.

Signature

____

Date *(dd/mm/yy)*

______/______/__________

Demographics

Utrecht University

Participant ID □

Please fill in exactly one box per question

Age

□ 18 - 24
□ 25 - 34
□ 35 - 44
□ 45 - 54
□ 55 - 64
□ Above 65

Gender

□ Male
□ Female
□ Other / Would not disclose

Primary Language

□ Dutch
□ English
□ Other

Experience with Virtual Reality (VR)

□ None
□ I have used VR before, but not often
□ I use VR regularly (1+ times per month)
□ I use VR often (1+ times per week)

Experience with Lifelogging

□ None
□ I am aware of the concept of Lifelogging
□ I actively work with the processing of Lifelog data.
□ I actively generate, store and process Lifelog data.

Motion Sickness

□ I do not suffer from motion sickness
□ I occasionally suffer from (mild) motion sickness
□ I often suffer from motion sickness, and the symptoms can be severe.
If you have selected the third option, you may not partake in this experiment.

Evaluation

Utrecht University

Participant ID □

Select 1 option

Map Interface / DateTime Interface / Tag Interface

Please tick one box per question.
Give each question a score between one and five based on how much you agree with the question, with 1 being "Strongly Agree" and 5 being "Strongly Disagree"

I think that I would like to use this interface frequently.	1	2	3	4	5
I found the interface unnecessarily complex.	1	2	3	4	5
I thought the interface was easy to use.	1	2	3	4	5
I think that I would need the support of a technical person to be able to use this interface.	1	2	3	4	5
I found the various functions in this interface were well integrated.	1	2	3	4	5
I thought there was too much inconsistency in this interface.	1	2	3	4	5
I would imagine that most people would learn to use this interface very quickly.	1	2	3	4	5
I found the interface very cumbersome to use.	1	2	3	4	5
I felt very confident using the interface.	1	2	3	4	5
I needed to learn a lot of things before I could get going with this interface.	1	2	3	4	5

Name one or more things you liked about the interface

____

____

____

____

Name one or more things you would like to see improved

____

____

____

____