

User Interaction for Visual Lifelog Retrieval in a Virtual Environment

Aaron Duane and Cathal Gurrin

`aaron.duane@insight-centre.org, cathal.gurrin@dcu.ie`

Abstract. Efficient retrieval of lifelog information is an ongoing area of research due to the multifaceted nature, and ever increasing size of lifelog datasets. Previous studies have examined lifelog exploration on conventional hardware platforms, but in this paper we describe a novel approach to lifelog retrieval using virtual reality. The focus of this research is to identify what aspects of lifelog retrieval can be effectively translated from a conventional to a virtual environment and if it provides any benefit to the user. The most widely available lifelog datasets for research are primarily image-based and focus on continuous capture from a first-person perspective. These large image corpora are often enhanced by image processing techniques and various other metadata. Despite the rapidly maturing nature of virtual reality as a platform, there has been very little investigation into user interaction within the context of lifelogging. The experiment outlined in this work seeks to evaluate four different virtual reality user interaction approaches to lifelog retrieval. The prototype system used in this experiment also competed at the Lifelog Search Challenge at ACM ICMR 2018 where it ranked first place.

Keywords: virtual reality, lifelog, retrieval, user interaction

1 Introduction

The most prevalent form of lifelog data at present is visual data captured from wearable cameras. This data is typically captured continuously from the first-person perspective and a single lifelogger can produce thousands of images per day. Distilling this huge and ever-increasing dataset of images into actionable insights about the individual’s life is a core aspect of lifelogging research. Most often this is assisted by automated image processing techniques such as concept detection and event segmentation. The enhanced metadata generated by these techniques then needs to be exposed intuitively to users alongside the visual imagery in order to support effective retrieval of lifelog information.

Previous research [1] in this area has focused on various hardware platforms such as desktops, tablets, and smart-phones; where each device was investigated for its impact on lifelog exploration use cases. In this paper we expand on that research by evaluating the potential of virtual reality (using the HTC Vive) as a platform for visual lifelog retrieval. Though virtual reality has yet to become as ubiquitous as phones or tablet computers, the hardware is continually becoming

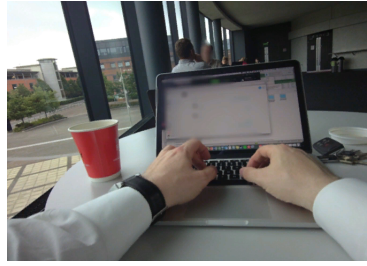
more sophisticated and affordable. It is our hypothesis that visualising complex multi-faceted data in three dimensions alongside a broader field of view could be a more intuitive and efficient method of visual lifelog exploration.

While there are numerous potential benefits to interacting with a lifelog system in a virtual environment, the focus of this research is specifically on lifelog retrieval, defined as the ability of a lifelog system to "retrieve specific digital information" [2]. Unlike some other lifelogging use cases, such as reminiscence or reflection [2], retrieval is the most suitable due to the ease of evaluation and potential for use as a daily life assistance or memory support tool [3]. Evaluating these lifelog retrieval systems is most often accomplished by means of a known-item search task where a set of topics are defined based on events that appear in an individual's lifelog (e.g. waiting for a bus, drinking a coffee, etc). Participants then attempt to search for these topics using their respective lifelog retrieval systems.

The goal of the experiment outlined in this paper was to determine a quantitative measure of the effectiveness of four different approaches to user interaction within a virtual environment and infer which one would be most suited, if any, to visual lifelog retrieval. In addition, we also wanted to determine a qualitative reflection of the system and interaction methodologies as a whole to help improve the virtual interface and user experience. To the best of our knowledge, this is the first lifelog interaction mechanism that has been developed for an environment in virtual reality [4]. This research is conducted as part of a larger study which also explores different data visualisation techniques for lifelog data in virtual reality.

2 Dataset

One of the largest obstacles in lifelogging research is the availability of test collections of sufficient quality and size. This is because there are significant technical challenges to overcome from the gathering of the data, its semantic enrichment and also ensuring the privacy of the individuals captured in the personal archive. The dataset utilised in this paper was sourced from the NTCIR conference [3] where it was originally released as part of the conference's lifelog tasks, which included a known-item search task referred to as the Lifelog Semantic Access Task (LSAT) [5]. The collection contains 90 days of data from 2 lifeloggers who together captured about 114,000 images. These images were then semantically enriched with automatic image processing techniques. The most valuable enrichment was the concept detection which resulted in each image in the dataset being tagged with an average of 5-10 concepts to describe its content (see Figure 1). A total of 48 known-item topics were released alongside the test collection, 24 for each of the two lifeloggers present in the dataset. These topics encompassed a broad range of life experience from the mundane (e.g. eating pasta for lunch) to the unusual (e.g. being interviewed for television). For the scope of our experiment, we determined using one lifelogger and their corresponding 24 known-item topics would be sufficient.



```

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

```

Fig. 1: Example Image from Test Collection with Concepts

3 Virtual Reality

The rationale for investigating the impact of virtual reality on visual lifelog exploration is based on its highly immersive quality. There are numerous benefits to operating in highly immersive environments; the most obvious being the ability for individuals to garner first-hand experience in an activity without actually engaging in said activity. For example in healthcare, a surgeon could practice an operation without risk of patient injury. However, there are other benefits to immersion that are less obvious. For example, actively using more of the human sensory capability and motor skills has been known to increase understanding and learning [6] and new research has suggested that immersion greatly improves user recall [7]. Also our ability to engage with digital elements directly in an open three dimensional environment more closely simulates our natural environment more so than a two dimensional analogue. This could suggest user interactions in virtual reality have the potential to be more intuitive, especially for novice users. It is difficult to speculate on every potential impact virtual reality might have on lifelog exploration, especially at this early stage, but we feel there is sufficient potential to warrant an exploratory examination of its applicability.

Virtual reality platforms, when compared to more conventional platforms such as laptops and phones, are in their relative infancy and this is compounded by the cost of the hardware to date. It still requires significant computing resources to generate high resolution virtual environments. However, the cost of the first generation of head-mounted displays has already notably reduced and more virtual reality platforms enter the consumer market each year, generating increased competition and more affordable pricing. It is reasonable to predict that as the hardware becomes more accessible to consumers, its application and use cases will become more sophisticated and nuanced. Using virtual reality to explore lifelog data may seem like a niche area today, but if we envision a future where virtual reality is as simple as equipping a mobile device and lightweight headset, it has potential to be preferable and more intuitive than previous conventional platforms.

Though there has been almost no research targeting the exploration of lifelogs in virtual reality, there has been some applications developed for the platform that could facilitate aspects of exploring and retrieving life experiences. One ob-

vious example is the playback of 360-degree video which is considerably more immersive when viewed in virtual reality and is especially so when the footage is recorded from a more familiar first-person perspective. This evolution of immersion within virtual reality can extend to many interaction methodologies that could better facilitate lifelog exploration. This is not to suggest that explicit examples of lifelog interaction in virtual reality do not exist at all. For example, an art installation by Alan Kwan titled 'Bad Trip'¹ was developed in 2012 which enables users to explore a manifestation of the creator's mind and life experience within a virtual environment. There are also non-lifelog related image retrieval systems developed for virtual reality which do things like map the virtual environment's three axes to facets of image content [8].

For the scope of research carried out in this paper, it was decided to use the HTC Vive², developed by HTC and Valve, as the virtual reality (VR) platform. At the time of writing, it is one of the most technically sophisticated virtual reality platforms available to consumers. However, it is important to acknowledge that the work undertaken in this research area is intended to be applicable to virtual reality as a whole and not strictly limited to the scope of what is possible with the HTC Vive. Therefore where possible, the evaluation criteria implemented in this work has been adapted to account for any virtual reality platform, with the caveat that it should also be equipped with two wireless controllers that are tracked in real-time alongside the head-mounted display.

4 System Overview

As previously stated, the focus of this experiment was to determine a quantitative measure of the effectiveness of four different approaches (see Figure 2) to user interaction performing lifelog retrieval tasks in a virtual environment. We also wanted to determine a qualitative reflection of the system and interaction methodologies to help address any notable flaws in the virtual user interface that may have been overlooked. The prototype system has two primary components, each of which needed to be optimised for virtual reality. The querying component was a virtual interface designed to provide a quick and efficient means for a user to generate a faceted query within the prototype system. While there are many approaches that one could take to input queries, a decision was made to focus on gesture-based interaction, as opposed to other forms of interaction.

The gesture-based querying interface consists of two sub-menus, one for selecting lifelog concepts of interest and the second for selecting the temporal aspect of the query (e.g. hours of the day or days of the week). A typical query to the system, such as "using the computer on a Saturday afternoon" would require the user to use the concept sub-menu to select the appropriate visual descriptors (e.g. computer or laptop) and the temporal sub-menu to select the time range (afternoon) and the day of the week (Saturday). The user then hits the submit button and the query is executed and the result is displayed for the

¹ Alan Kwan's 'Bad Trip' - <https://www.kwanalan.com/blank>

² HTC Vive - <https://www.vive.com/eu/product/>

maximum of 10 concepts per query. In our experimentation, no user has ever selected ten concepts, so this is a reasonable upper-bound for the current work. The temporal sub-menu presents the user with the 7 days of the week and the 24 hours of the day. These days and hours can be selected in any combination to generate a temporal facet for the query.

An important aspect of developing a prototype for visual lifelog exploration in a virtual environment is to identify the most efficient and preferred methods of interacting with that environment’s user interface. At present, there is not a clear answer for how to best interact with a user interface in this context; there are no well defined and understood interaction best practices to implement (e.g. point-and-click in the desktop environment, or swipe-a-finger in a touchscreen environment). Without such normative guidance, we developed two high-level interaction methodologies to interfacing with our prototype which we refer to as ‘distance-based’ and ‘contact-based’ user interaction. These two methodologies were further divided into two low-level variations for a total of four interaction modes in total.

4.1 Distance-Based Interaction

The distance-based approach utilises interactive beams which originate at the the top of the user’s wireless controllers. These beams are projected when the controllers are pointed at any relevant interface in the virtual environment and directly interact with that interface’s elements (see Figure 5). This method of interaction is comparable to a lean-back style of lifelog browsing as introduced in [9] and is functionally similar to using a television remote or other such device. Pressing a button on the controllers selects the concept or time-range that is being pointed at. Naturally, it is possible to use both hands to select concepts in parallel, should a sufficiently dexterous user be generating queries.

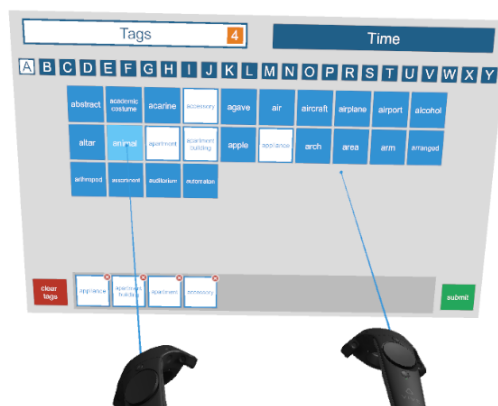


Fig. 5: Distance-Based User Interaction

The low-level variations within the distance-based approach to interaction differ in the positioning of the user interface within the virtual environment. One variation orients the menu vertically and across from the user, which we refer to as the *billboard* style of interaction. The second variation places the menu horizontally and beneath the user, which we refer to as the *floorboard* style of interaction.

4.2 Contact-Based Interaction

The contact-based approach utilises a much more direct form of interaction where the user must physically touch the interface elements with their controllers. To facilitate this process, the controllers are outfitted with a drumstick-like device protruding from the head of each controller (see Figure 6). This object was added to enhance precision and fidelity when contacting interface elements. This method of interaction is reminiscent of a more conventional style of lifelog browsing where the controller drumsticks mimic how our fingers interact with a keyboard or touchscreen. Tactile feedback is provided via the controllers to reflect hitting the keys.

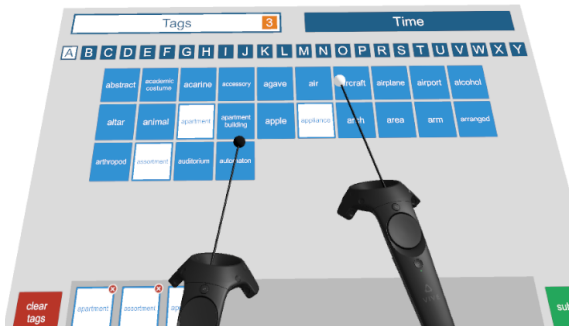


Fig. 6: Contact-Based User Interaction

Similar to before, the low-level variations within the contact-based approach to interaction differ in the positioning of the user interface. One variation orients the menu at a slight angle in front of the user and they have the option of interacting with it using both controllers. We refer to this as the *dashboard* style of interaction. The second variation attaches the menu directly to one of the user's controllers (their choice) and the user interacts with the menu using the opposing controller. We refer to this as the *clipboard* style of interaction.

The two high-level interaction methodologies, distance and contact, are based on real-world analogues (television, keyboard, touchscreen, etc.) and can be observed in various forms in industry-standard virtual reality applications such as

the HTC Vive’s main menu ³ or Google’s popular Tilt Brush interface ⁴. The low-level variations within these two methodologies were developed to further expand on how different interaction types impacted user experience.

4.3 Lifelog Data Ranking and Visualisation

As previously stated, after a faceted query is submitted to the system, the querying interface disappears and the user is presented with the highly-ranked filtered images (see Figure 7) in decreasing rank order. These images are ranked using a combination of concept relevance and the time of capture (maintaining the temporal organisation of the data), where concept relevance takes precedence over the temporal arrangement. For example, if the user creates a query containing 3 different concepts, then images containing all 3 concepts will be ranked first in the list, followed by images containing 2, and then 1. When multiple images contain the same amount of relevant concepts, those images are ranked temporally according to the image capture time.

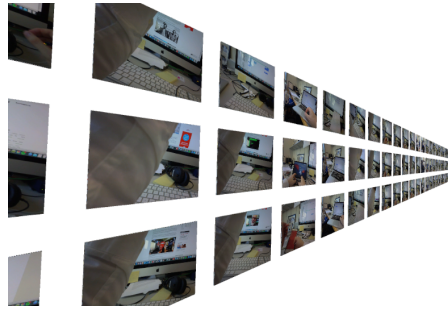


Fig. 7: Ranked List of Images

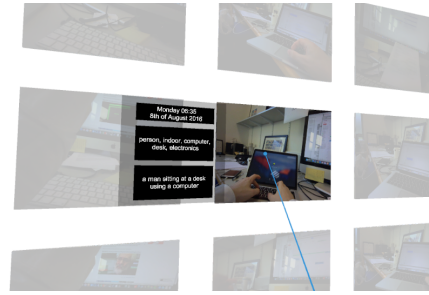


Fig. 8: Image Metadata

Any image displayed in the ranked list can be selected for further exploration by pointing the user’s controller at it and pressing a button. This displays additional metadata about the image such as the specific capture date and time and what concepts have been detected (see Figure 8). Additional filtering options are also made available along with this metadata. For example, the user can choose to see other images contained in the manually annotated event this image was labelled under or they can simply view all the images captured before and after the target image within a specific timespan.

5 Experiment Configuration

This experiment utilises known-item search tasks as the evaluation methodology to quantitatively compare different approaches to lifelog retrieval in virtual

³ SteamVR - <http://store.steampowered.com/steamvr>

⁴ Google Tiltbrush - <https://www.tiltbrush.com>

reality. Each participant also answers a post-experiment user feedback questionnaire (containing an open input field) to qualitatively evaluate how the system performed.

A total of 16 participants volunteered to take part in the experiment. The minimum criteria to participate was a strong understanding of the English language and rudimentary computer skills. It was not a requirement for participants to have any knowledge or experience with virtual reality prior to testing. To reduce any potential cognitive bias, each user was given a thorough walkthrough of the system prior to testing and needed to successfully complete a trial topic on each interaction type before proceeding with the experiment.

Each person attempted to identify a subset of 16 topics from the NTCIR test collection [3]. Since they were wearing a VR headset, the topics were described by an assistant. The description of each topic was taken directly from the test collection so every user received an identical definition of the topic prior to testing. The users were timed and given a maximum of 180 seconds to identify a relevant image from the dataset, reflecting the currently described topic, before moving onto the next topic.

To evenly assess each of the four interaction types (billboard, floorboard, dashboard and clipboard), the 16 topics were divided into four groups of four. Each user attempted to identify four topics on each interaction type until all four interaction types and all 16 topics were used. The experiment was purposefully configured so that each topic would be explored on each of the four interaction types a total of four times and that the ordering for each user would account for any learning bias.

6 Results

6.1 User Performance

The 180 second time limit per topic was imposed to prevent a topic taking an excessive amount of time and was also the same number of seconds allocated in the Lifelog Search Challenge [10] at ACM ICMR 2018 which used a subset of the NTCIR test collection employed in this work. If the user exceeded this limit, they would immediately stop and proceed to the next topic.

In figure 9 we can see a visualisation displaying the time taken to identify a topic on each of the four interaction types. Each of the 16 topics are labelled on the horizontal axis (with the topic ids taken from the test collection) and the vertical axis represents the average time in seconds each topic took to complete. The four interaction types are represented by four coloured bars for each topic and there is an indication beneath each topic of how many times a user failed to find relevant content (by exceeding the 180 second limit).

For the majority of topics, the interaction approaches performed similarly, suggesting that there is no clearly superior interaction approach. However there was some inconsistency in topics T2, T17 and T22. The fact that these are also the only topics which were failed by a number of users suggests this inconsistency

is unrelated to the interaction type and is more likely the result of how some users interpreted the topic description. For example, T17 was a topic describing the lifelogger being recorded for a television show, and many participants correctly used 'camera' as a concept to filter with, as logically the lifelogger's personal camera would capture the television camera recording them. However, many participants failed to make this connection and instead used the 'television' concept which resulted in a significant number of false positives being returned.

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

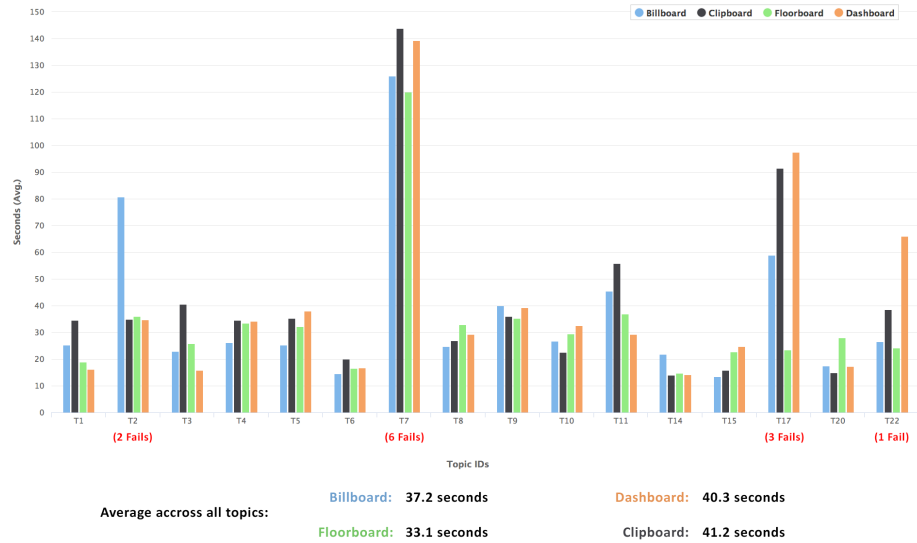


Fig. 9: Average Seconds Taken Per Topic on each Interaction Type

6.2 User Feedback

The experiment participants were asked to fill out a user experience questionnaire after each group of four topics, corresponding to the interaction type they had just used (billboard, dashboard, etc.). Each questionnaire contained usability statements which the users needed to state their level of agreement with on a five-point Likert [11] scale. Most importantly, the users were asked an open

question about the usability of each interaction type and if they felt it could be improved. Finally, at the end of the experiment, the participants were asked to rank the four interaction types in order of their preference.

The most popular distance-based approach was the 'billboard' style interaction and the most popular contact-based approach was the 'dashboard' style interaction. There was a slight overall preference towards the distance-based approaches, which we suspect is due the familiar nature of the point-and-click interaction. Pointing and clicking came a lot more naturally to participants due to their experience with televisions and remote controls, whereas physically contacting digital interface elements required more practice to become accustomed with. Despite the general preference for the distance-based interaction, many users expressed positive sentiment for the contact-based approach for specific use cases, like selecting many interface elements in a short amount of time. However, there was notable discomfort using the 'clipboard' style interaction as it relied on controlling two separate interactive elements at once (the menu and the drumstick) which some users found challenging to coordinate.

Based on user feedback, we suspect a hybrid system utilising the elements of both the 'billboard' and 'dashboard' modes of user interaction would be the most effective interaction methodology. For example, a distance-based approach is most suited to more casual user interactions, such as browsing, whereas more complex user interactions, such as typing, would be most suited to a contact-based approach. Furthermore, ensuring that the user interface's position is static in the virtual environment, but adjustable by the user at any time, was a recurring sentiment.

7 Conclusion

In this paper we outlined our work developing a quantitative and qualitative evaluation methodology to develop a state of the art user interface for a lifelog retrieval system in virtual reality. This work did not extend to evaluating the data visualisation aspect of a virtual reality lifelog retrieval system as this will be addressed in a future work. Some of the key insights we determined during this study were to direct the attention of the users to newly exposed interface elements within the virtual environment to prevent any user getting lost in the virtual space. Also ensure all interface elements are resizeable and repositionable by the user to maximise content legibility and reduce eye strain. Where relevant, it is suggested to utilise a point and click interaction system for low precision tasks and a contact-based interaction system for high precision tasks. Clearly label and highlight the VR controller buttons when they are contextually relevant to the user interaction. These insights, and the remainder of the work outlined in this paper, contributed to the refinements of our virtual reality lifelog retrieval platform that enabled it to perform effectively at the Lifelog Search Challenge (LSC) [10] at ACM ICMR 2018 where it ranked first place among the other challenge participants [12]. It was the only virtual reality based system present at the conference; all other participants utilised conventional laptops or computers.

References

- [1] Yang Yang, Hyowon Lee, and Cathal Gurrin. “Visualizing lifelog data for different interaction platforms”. In: *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13* (2013), p. 1785. DOI: 10.1145/2468356.2468676.
- [2] Abigail J. Sellen and Steve Whittaker. “Beyond total capture”. In: *Communications of the ACM* 53.5 (May 2010), p. 70. ISSN: 00010782. DOI: 10.1145/1735223.1735243.
- [3] Cathal Gurrin et al. “NTCIR Lifelog: The First Test Collection for Lifelog Research”. In: *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. . ACM. ISBN 978-1-4503-4069-4 (2016), pp. 705–708. DOI: 10.1145/2911451.2914680.
- [4] Aaron Duane and Cathal Gurrin. “Lifelog Exploration Prototype in Virtual Reality”. In: 24th International Conference on Multimedia Modeling (MMM 2018), Proceedings. Lecture Notes in Computer Science 10704. Springer., 2018.
- [5] Cathal Gurrin et al. “Overview of NTCIR-13 Lifelog-2 Task”. In: *Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies*. . NTCIR. ISBN 978-4-86049-075-1 (2017), pp. 6–11.
- [6] Dale. E. *Audio-Visual Methods in Teaching (3rd Ed.)* Dryden Press, 1969, pp. 12, 13.
- [7] Eric Krokos, Catherine Plaisant, and Amitabh Varshney. “Virtual memory palaces: immersion aids recall”. In: *Virtual Reality* 0123456789 (2018), pp. 1–15. DOI: 10.1007/s10055-018-0346-3.
- [8] Munehiro Nakazato and Thomas S Huang. “3D MARS: Immersive Virtual Reality for Content-Based Image Retrieval”. In: *Proceedings of 2001 IEEE International Conference on Multimedia and Expo (ICME2001)*. 2001.
- [9] Cathal Gurrin et al. “Browsing Large Personal Multimedia Archives in a Lean- back Environment”. In: *International Conference on Multimedia Modeling* (2010).
- [10] “LSC '18: Proceedings of the 2018 ACM Workshop on The Lifelog Search Challenge”. In: Yokohama, Japan: ACM, 2018. ISBN: 978-1-4503-5796-8.
- [11] Rensis Likert. *A Technique for the Measurement of Attitudes*. 1932, p. 55.
- [12] Aaron Duane, Wolfgang Huerst, and Cathal Gurrin. “Virtual Reality Lifelog Explorer”. In: Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval. ACM., 2018.