Spatiotemporal Lifelog Analytics in Virtual Reality with vitrivr-VR

Florian Spiess
University of Basel
Basel, Switzerland
florian.spiess@unibas.ch

Heiko Schuldt
University of Basel
Basel, Switzerland
heiko.schuldt@unibas.ch

Luca Rossetto
University of Zurich
Zurich, Switzerland
rossetto@ifi.uzh.ch

ABSTRACT

Modern wearables and smart devices make it easier than ever to collect a detailed, digital record of biometric as well as visual and aural information. Reasons to collect such a lifelog range from health applications to vacation documentation. With the large quantities of data that can be collected in very short periods of time, it remains a challenging problem to find specific events and answer questions based on such lifelogs. In order to support lifelog multimedia analytics within collections of large sizes, interactive analytics methods must be developed that take advantage of the diverse and multimodal data available.

Through rapid technological improvement, immersive interfaces, such as virtual reality (VR) devices, are quickly becoming more affordable and accessible to the general public. Due to their immersive capabilities, these interfaces are uniquely suited to visualizing and allowing interaction with diverse multimodal data. Even so, research on interactive lifelog analytics has been heavily focused on conventional desktop interfaces with pointer controls, and little research has been conducted on interfaces for virtual reality. While many interfaces developed for conventional interfaces can also be used within VR, the advantages of immersive interfaces can only be utilized by methods tailored to this new interface modality.

In this paper, we describe the vitrivr-VR virtual reality multimedia analytics system in the form in which it will participate in the Lifelog Search Challenge (LSC) 2024. In order to take greater advantage of the variety of available lifelog data and the affordances of immersive interfaces, we implement new search interfaces that allow easier and more flexible temporal and spatial query formulation as well as spatiotemporally contextualized results visualization in VR.

CCS CONCEPTS

• **Information systems → Image search**; **Search interfaces**; **Query representation**; • **Human-centered computing → Virtual reality**.

KEYWORDS

Lifelog Search Challenge, Virtual Reality, Interactive Lifelog Retrieval

LSC '24, June 10, 2024, Phuket, Thailand

ACM ISBN 979-8-4007-0550-2/24/06
https://doi.org/10.1145/3643489.3661113

ACM Reference Format:
Florian Spiess, Heiko Schuldt, and Luca Rossetto. 2024. Spatiotemporal Lifelog Analytics in Virtual Reality with vitrivr-VR. In *The 7th Annual ACM Lifelog Search Challenge (LSC '24), June 10, 2024, Phuket, Thailand.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3643489.3661113

1 INTRODUCTION

Lifelogging, the act of recording one's own life through a variety of sensors and devices, is becoming more and more popular. With the increasing convenience of recording a variety of data ranging from health and fitness data to photos and videos, many people are passively collecting a lifelog without actively thinking about it. While this data is collected in ever greater quantities, appropriate methods are required to adequately analyze and utilize these highly heterogeneous collections. Many approaches have already been developed to analyze lifelog data, many of which extract information from the individual modalities separately, but in order to answer realistic analytics questions for lifelog collections, interactive analytics methods are needed.

To develop methods to best facilitate interactive lifelog analytics for large lifelog collections, a research community has formed around multimedia and lifelog analytics evaluation campaigns such as the Lifelog Search Challenge (LSC) [6]. These evaluation campaigns serve as a much-needed proving ground to test improvements to existing methods and experiment with novel approaches. As such, the LSC has seen a number of participants attempting lifelog analytics with novel interfaces, including virtual reality (VR) interfaces [1–4, 7]. Due to their rapid technological improvement and quickly increasing affordability, VR interfaces are becoming available to a large group of users and are being used in a wide range of applications. Through their immersive visualization and interaction capabilities, VR interfaces are also ideally suited to enable interaction with the multimodal and heterogeneous data contained in large lifelog collections. Despite the necessity of exploring methods for an emerging user interface modality and the potential benefits it can offer, only little research has been conducted on the best way to enable multimedia and lifelog analytics on such platforms.

In this paper, we describe the latest version of vitrivr-VR, a virtual reality multimedia analytics system, with a focus on changes and approaches developed for testing at the Lifelog Search Challenge 2024. vitrivr-VR has participated in different forms in several previous LSCs [11, 12, 14]. Having been designed as a general-purpose multimedia analytics system, previous versions of vitrivr-VR have had only limited support for much of the information provided in lifelog data. In this newest version, we focus on making better use of the spatiotemporal information contained within and integral to lifelog data. In the remainder of this paper, we provide an overview

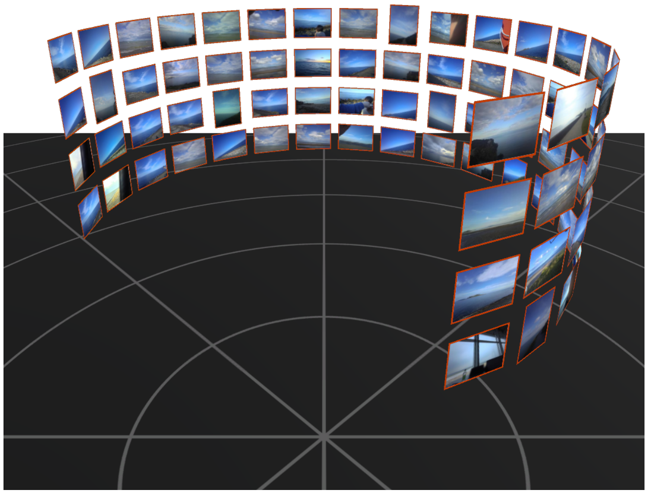

Figure 1: Cylindrical results display. Shows the results of a similarity query ordered by similarity in a grid around the user. Depicted are the results of a query for "a view of the sea".

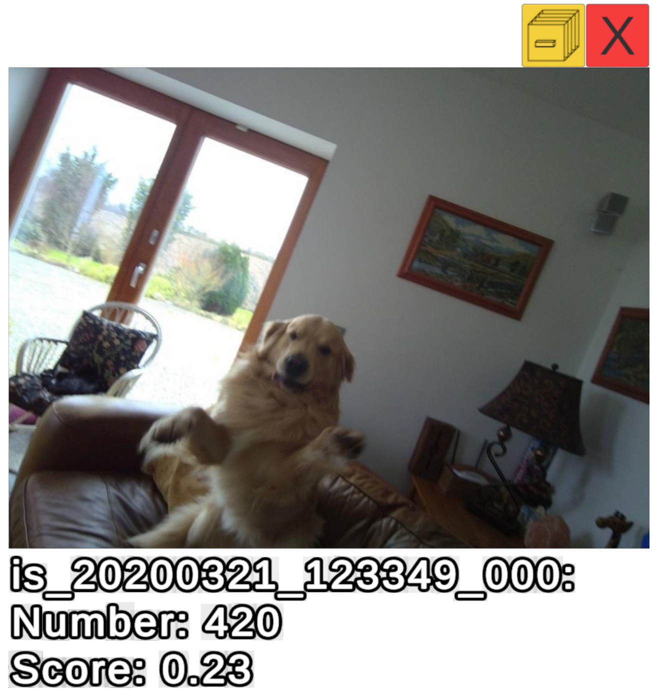

Figure 2: Detailed view of an image result. Shows the image in higher resolution and with additional information and the ability to open the multimedia drawer view.

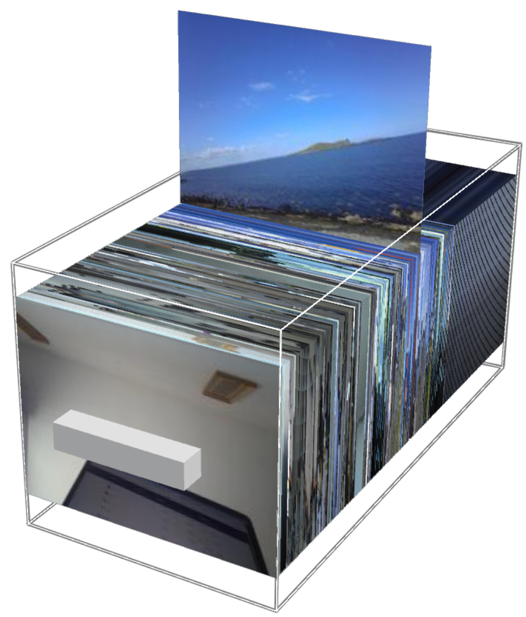

Figure 3: Multimedia drawer view of a part of the lifelog image series. One of the thumbnails is hovered by the user's hand and, therefore, raised above the rest of the drawer. The drawer handle has been slightly pulled to extend the drawer beyond its usual cube shape.

of the vitrivr-VR system architecture in section 2, present our new spatiotemporal interfaces for querying and browsing in section 3, and finally conclude in section 4.

2 SYSTEM OVERVIEW

The vitrivr-VR system consists of three main components: *Cottontail DB* [5], the vector store database, *Cineast* [9], the feature extraction and retrieval engine, and *vitrivr-VR*, the virtual reality analytics interface. The Cottontail database allows vector space retrieval through nearest neighbor queries, as well as Boolean retrieval. Cineast provides feature extraction capabilities for a wide range of features, as well as retrieval support. As the user interface, vitrivr-VR provides the query formulation and results browsing capabilities. All three components are part of the vitrivr stack and are available as open-source projects on GitHub[1].

The main feature used for retrieval in vitrivr-VR is the OpenCLIP [8] model `CLIP ViT-B/32 xlm roberta base`[2] trained on the LAION-5B dataset [10]. In order to support textual queries, vitrivr-VR supports speech-to-text and a word-gesture keyboard [15]. Results are primarily presented on a cylindrical results display that shows previews of retrieved results curved around the user. An example of this is shown in Figure 1. More detailed results displays can be created from this results display. The more detailed displays can then be moved, rotated, and placed anywhere in the virtual space for later viewing. Figure 2 shows an example of the more detailed results view.

From a detailed results display, a multimedia drawer can be created showing nearby items in the lifelog image time series. This drawer view allows the user to move their hand through the images in the time series and have the hovered image displayed above the drawer. By moving their hand through the drawer in this way,

[1]https://github.com/vitrivr
[2]https://huggingface.co/laion/CLIP-ViT-B-32-xlm-roberta-base-laion5B-s13B-b90k

Figure 4: The vitrivr-VR map interface. It consists of a globe minimap (top left), a pan- and zoomable detail map (center), and a pin box.

a user can riffle through the time series in a fashion reminiscent of a flip book, providing an easy and intuitive way of navigating the image series. An example of the multimedia drawer with a hovered frame is shown in Figure 3 and a more detailed description is provided in [13].

In addition to the interfaces discussed in this section, vitrivr-VR also supports spatiotemporal interfaces for both formulating queries as well as visualizing results. These interfaces are described in the following section.

3 SPATIOTEMPORAL INTERFACES

Spatiotemporal information is central to lifelog data. As such, it is imperative for lifelog analytics systems to support dedicated interfaces to express and visualize these aspects of lifelog data. Previous versions of vitrivr-VR have introduced query interfaces for both spatial query specification as well as temporal filtering during search. While these interfaces allow intuitive specification of some aspects of spatiotemporal context, they are quite limited in their expressiveness and particularly in how results are displayed.

3.1 Spatial Interfaces

For spatial queries, vitrivr-VR supports a map-based interface. The map interface consists of three main parts: a drag- and zoomable *detail map*, a *globe minimap*, and a *pin box*. All parts of the interface can be grabbed, moved, and rotated independently. A screenshot of the map interface is shown in Figure 4. The data for the map is loaded from Mapbox[3] through the Mapbox UPM package[4]. The detailed map is the main part of the interface and shows a flat, zoomable view of the map. By dragging the map with one hand, the part of the map that is shown can be panned to reveal the surrounding area. Using two hands to mimic a pinching gesture allows the part of the map that is shown to be zoomed in and out in order to show a larger area or show a smaller area in greater detail. Locations on the map can be marked by grabbing a pin from the pin box and placing it on the chosen location on the map.

The globe minimap is a supporting interface for the detailed map. Zooming in and out of the detailed map to pan between different parts of the world map is time-consuming and tedious. To make this process much faster and more comfortable, the globe minimap can be used to quickly jump to any location on the globe. By placing a pin on the globe minimap, the detailed map is immediately panned to the selected location, retaining the same zoom level as before.

In previous versions of vitrivr-VR, this interface could only be used to specify points on the map to use for distance-based similarity queries. In order to improve the functionality of the spatiotemporal interfaces in vitrivr-VR, the spatial interface now also has two additional functionalities: specification of a specific region and the ability to display results directly on the map interface. By circling a region on the map, it can be selected as a Boolean-like query requirement, limiting the search to results from this area. This makes it much easier to restrict the search to a clearly defined area rather than mixing the spatial similarity of a distance to a point with the similarity score from other aspects of the query. By allowing the results from a query to be visualized directly on the map, it is possible to identify regions of interest, allowing nearby results to be explored or the region to be used in follow-up queries.

3.2 Temporal Interfaces

Previous versions of vitrivr-VR supported only very limited temporal functionality. Through a Boolean search window, years, months, days, days of the week, and hours could be individually selected, as shown in Figure 5. While this selection allows conditions on each one of these categories to be quickly and easily specified individually for each category, several conditions requiring multiple categories, e.g., "the 5$^{\text{th}}$ of March or the 23$^{\text{rd}}$ of April", cannot be expressed. Furthermore, as a result of this limitation, only the selection of years represents the true availability of data in the system. All other categories offer all theoretical values for selection, even if no data exists for them.

Our new temporal interface makes two main improvements over previous versions: a more flexible, calendar-like query selection that reflects the true availability of data in the system and a results display mode that shows results in stacks for each day in a calendar-like arrangement rather than sorted by similarity score, to make their temporal order more apparent. By presenting temporal query selection options in the form of a calendar-like interface, it becomes possible to query for different dates from different months and years. By only showing months for which data is available and visually indicating which dates within these months have associated data, it becomes easy to see at a glance which data is available to query. Similarly, by allowing results to be displayed in a calendar-like manner sorted into stacks by date, it becomes much easier to scan

[3]https://www.mapbox.com
[4]https://github.com/Spiess/mapbox-upm-sdk

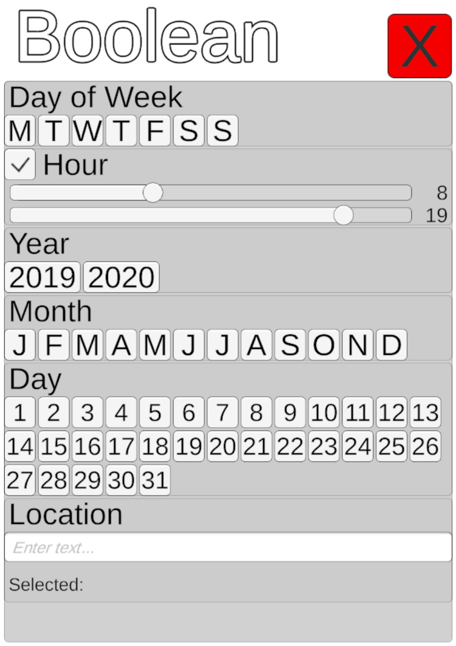

Figure 5: The Boolean search interface of previous versions of vitrivr-VR. Conditions on years, months, days, and hours are independent and cannot be coupled.

Figure 6: The calendar-like results display mode. Results are arranged by day of week and date resembling a calendar.

the results in a structured way with reference to the dates. An example of this results display is shown in Figure 6.

4 CONCLUSION

In this paper, we present an updated version of vitrivr-VR for our participation in the LSC 2024, including much improved spatiotemporal interfaces, with the goal of making it easier to both specify the spatiotemporal context in virtual reality, as well as view the spatiotemporal context of retrieval results. We update our map interface to allow the specification of query bounds and to allow the spatial visualization of results. We update our temporal retrieval interface to allow much greater flexibility and ease of use when specifying temporal ranges and provide visual information on the available data during query specification. In addition, we implement an additional result sorting order that shows results sorted by date. With these improvements, we aim to greatly improve the spatiotemporal lifelog analytics capabilities of vitrivr-VR.

ACKNOWLEDGMENTS

This work was partly supported by the Swiss National Science Foundation through projects "Participatory Knowledge Practices in Analog and Digital Image Archives" (contract no. 193788) and "MediaGraph" (contract no. 202125).

REFERENCES

[1] Aaron Duane and Cathal Gurrin. 2019. User Interaction for Visual Lifelog Retrieval in a Virtual Environment. In *MultiMedia Modeling*. Springer International Publishing, Cham, 239–250. https://doi.org/10.1007/978-3-030-05710-7_20

[2] Aaron Duane, Cathal Gurrin, and Wolfgang Huerst. 2018. Virtual Reality Lifelog Explorer: Lifelog Search Challenge at ACM ICMR 2018. In *Proceedings of the 2018 ACM Workshop on The Lifelog Search Challenge*. Association for Computing Machinery, Yokohama, Japan, 20–23. https://doi.org/10.1145/3210539.3210544

[3] Aaron Duane and Bjorn Þór Jónsson. 2021. ViRMA: Virtual Reality Multimedia Analytics at LSC 2021. In *Proceedings of the 4th Annual on Lifelog Search Challenge*. Association for Computing Machinery, New York, NY, USA, 29–34.

[4] Aaron Duane, Björn Þór Jónsson, and Cathal Gurrin. 2020. 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*. Association for Computing Machinery, Dublin, Ireland, 7–12. https://doi.org/10.1145/3379172.3391716

[5] Ralph Gasser, Luca Rossetto, Silvan Heller, and Heiko Schuldt. 2020. Cottontail DB: An Open Source Database System for Multimedia Retrieval and Analysis. In *International Conference on Multimedia*. Association for Computing Machinery, New York, NY, USA, 4465–4468. https://doi.org/10.1145/3394171.3414538

[6] Cathal Gurrin, Liting Zhou, Graham Healy, Werner Bailer, Duc-Tien Dang-Nguyen, Steve Hodges, Björn Þór, Jakub Lokoc, Luca Rossetto, Minh-Triet Tran, and Klaus Schoeffmann. 2024. Introduction to the Seventh Annual Lifelog Search Challenge, LSC'24. In *Proceedings of the 2024 International Conference on Multimedia Retrieval (ICMR'24)* (Phuket, Thailand) *(ICMR '24)*. Association for Computing Machinery, New York, NY, USA, 2.

[7] Wolfgang Hürst, Kevin Ouwehand, Marijn Mengerink, Aaron Duane, and Cathal Gurrin. 2018. Geospatial Access to Lifelogging Photos in Virtual Reality. In *Proceedings of the 2018 ACM Workshop on The Lifelog Search Challenge*. Association for Computing Machinery, Yokohama, Japan, 33–37. https://doi.org/10.1145/3210539.3210547

[8] Gabriel Ilharco, Mitchell Wortsman, Nicholas Carlini, Rohan Taori, Achal Dave, Vaishaal Shankar, Hongseok Namkoong, John Miller, Hannaneh Hajishirzi, Ali Farhadi, and Ludwig Schmidt. 2021. OpenCLIP. Zenodo. https://doi.org/10.5281/ZENODO.5143773

[9] Luca Rossetto, Ivan Giangreco, and Heiko Schuldt. 2014. Cineast: A Multi-feature Sketch-Based Video Retrieval Engine. In *IEEE International Symposium on Multimedia*. IEEE, New York, NY, USA, 18–23. https://doi.org/10.1109/ISM.2014.38

[10] Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, Patrick Schramowski, Srivatsa Kundurthy, Katherine Crowson, Ludwig Schmidt, Robert Kaczmarczyk, and Jenia Jitsev. 2022. LAION-5B: An Open

Large-Scale Dataset for Training next Generation Image-Text Models. In *Advances in Neural Information Processing Systems*, Vol. 35. Curran Associates, Inc., New York, NY, USA, 25278–25294.

[11] Florian Spiess, Ralph Gasser, Silvan Heller, Luca Rossetto, Loris Sauter, Milan van Zanten, and Heiko Schuldt. 2021. Exploring Intuitive Lifelog Retrieval and Interaction Modes in Virtual Reality with Vitrivr-VR. In *Workshop on Lifelog Search Challenge* (New York, NY, USA, 2021). Association for Computing Machinery, New York, NY, USA, 17–22. https://doi.org/10.1145/3463948.3469061

[12] Florian Spiess, Ralph Gasser, Heiko Schuldt, and Luca Rossetto. 2023. The Best of Both Worlds: Lifelog Retrieval with a Desktop-Virtual Reality Hybrid System. In *Proceedings of the 6th Annual ACM Lifelog Search Challenge* (Thessaloniki Greece, 2023-06-12). ACM, New York, NY, USA, 65–68. https://doi.org/10.1145/3592573.3593107

[13] Florian Spiess, Nicolas Scharowski, Ariane Haller, Zgjim Memeti, Heiko Schuldt, and Florian Brühlmann. 2024. Bringing Video Browsing to Virtual Reality: Empirical Evaluation of a Novel Multimedia Drawer. In *Proceedings of the 2024 International Conference on Multimedia Retrieval* (Phuket, Thailand, 2024-06). ACM, New York, NY, USA, 9. https://doi.org/10.1145/3652583.3658077

[14] Florian Spiess and Heiko Schuldt. 2022. Multimodal Interactive Lifelog Retrieval with Vitrivr-VR. In *Proceedings of the 5th Annual on Lifelog Search Challenge*. Association for Computing Machinery, New York, NY, USA, 38–42. https://doi.org/10.1145/3512729.3533008

[15] Florian Spiess, Philipp Weber, and Heiko Schuldt. 2022. Direct Interaction Word-Gesture Text Input in Virtual Reality. In *2022 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. IEEE, CA, USA, 140–143. https://doi.org/10.1109/AIVR56993.2022.00028