The Best of Both Worlds: Lifelog Retrieval with a Desktop-Virtual Reality Hybrid System

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

Ralph Gasser
University of Basel
Switzerland
ralph.gasser@unibas.ch

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

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

ABSTRACT

Personal lifelog data collections are becoming more common as a memory aid, as well as for analytical tasks, such as health and fitness analysis. Due to the multimodal and personal nature of lifelog data, interactive multimedia retrieval approaches are required to facilitate flexible and iterative query formulation and result exploration for retrieval and analysis. In recent years, novel user interface modalities have emerged, that allow new ways for users to interact with a retrieval system. Virtual reality, one such new modality, provides advantages as well as challenges for interactive multimedia retrieval in comparison to conventional desktop-based interfaces.

This paper describes a novel desktop-virtual reality hybrid system participating in the Lifelog Search Challenge 2023. The system, which is based on the components of the vitrivr stack, is described with a focus on query formulation in the web-based desktop user interface vitrivr-ng, and result exploration in the virtual reality-based vitrivr-VR.

CCS CONCEPTS

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

KEYWORDS

Lifelog Search Challenge, Virtual Reality, Interactive Lifelog Retrieval

ACM Reference Format:
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 *6th Annual ACM Lifelog Search Challenge (LSC '23), June 12, 2023, Thessaloniki, Greece.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3592573.3593107

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
LSC '23 , June 12, 2023, Thessaloniki, Greece
© 2023 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0188-7/23/06.
https://doi.org/10.1145/3592573.3593107

1 INTRODUCTION

Increasing numbers of people are recording their lives through images, videos and health data using increasingly affordable devices for the recording of this data, such as smartphones, action cameras, smart watches and fitness trackers. As a result, the size of personal collections of these kinds of data are ever increasing and are reaching sizes and velocities of growth beyond anything manageable by manual annotation.

While some of this data is primarily collected to be aggregated and observe trends, such as fitness and health related data, much of it is also used as a memory aid or to save cherished memories. To allow appropriate use of this so-called lifelogging data, it must be accessible not only by exact file name or specific date and time, but also based on other attributes such as the content and the temporal series of events recorded by it.

Many approaches already exist to analyze and retrieve lifelogging data. Some of these approaches are fully automated, requiring only an initial query to return results, but many more are interactive, allowing a user to iteratively refine their query and explore the data. Fully automated approaches may seem like the more convenient solution, however, in most cases interactive approaches are more desirable for lifelog analytics tasks, as results of a query may lead to insights that can be used to refine the search.

Virtual reality (VR) is a relatively new user interface modality, but has already shown great promise for multimedia retrieval interfaces [2, 3]. Previous multimedia retrieval evaluations have shown, that VR provides both advantages as well as challenges in comparison to desktop user interfaces. In particular, while the available space in VR is a great advantage for results exploration, text-based query formulation is much slower when using methods available in VR as compared to using a physical keyboard.

Both vitrivr [7, 8] and vitrivr-VR [13, 15] have participated in previous instances of the Lifelog Search Challenge (LSC) [6] and demonstrated competitive performance. As indicated by analyses of previous multimedia retrieval evaluations [14], while the more traditional browser-based user interface of vitrivr enables the user to express complex queries quickly with a multitude of modalities, the immersive result interaction modes of vitrivr-VR allow for efficient and intuitive browsing of results. For the 2023 edition of the LSC [6], we present a hybrid approach based on these findings that combines the two interfaces into a single system leveraging the advantages of both, to overcome the challenges of either. The hybrid system,

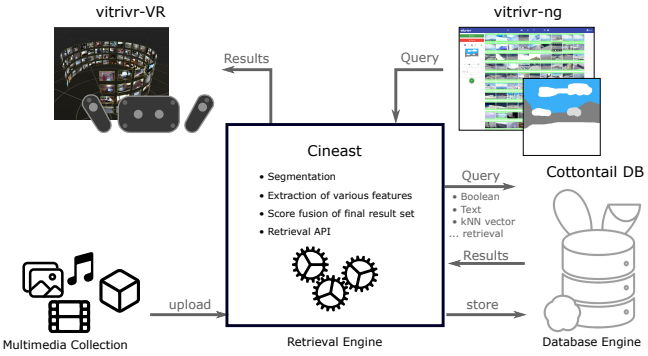

Figure 1: Architecture of the vitrivr desktop-VR hybrid system. Diagram adapted from [10].

which is operated by two users, seeks to combine the strengths of the two types of interfaces. Queries are expressed through the desktop interface while results are explored and ultimately submitted to the evaluation server in virtual reality.

The remainder of this paper is structured as follows: Section 2 details the architecture and setup of the hybrid system while Section 3 outlines the cooperative interaction mode. Finally, Section 4 concludes.

2 SYSTEM OVERVIEW

The vitrivr stack generally consists of three components: the data persistence layer Cottontail DB [4], the feature extraction and query processing engine Cineast [9], and a user interface. In the past, the user interface used would be *either* the browser-based vitrivr-ng [5] *or* the vitrual reality interface vitrivr-VR [13]. In this version of the system however, we use both of these user interfaces simultaneously for different parts of the multimedia retrieval process.

In order for the usually independently operated user interfaces to be able to benefit from each other's strengths, only minor extensions to the backend were necessary. A newly introduced query and result caching mechanism stores relevant query information as well as all results that were transmitted to a frontend during query execution. These cached results are made available through additional API endpoints. This allows any client to request previously retrieved result sets, even if they were generated by a query sent by a different client.

In our setting, vitrivr-VR uses these new API endpoints to poll for new results and make them available to the operator for exploration. The only additionally required changes were relevant configuration options to disable the query formulation functionality on vitrivr-VR and the communication with the evaluation server on vitrivr-ng, both of which could already be achieved using appropriate configuration and required no further change to the systems. The architecture diagram in Figure 1 provides an overview of the hybrid system.

Coordination in this prototype hybrid system works in a very analog manner: the desktop operator formulates and executes a query, that is then displayed in VR for exploration. While the VR operator is exploring the result set, the desktop operator is able to specify alternative, more refined queries. To aid in this refinement process, the VR operator can provide verbal feedback of the results to the desktop operator. To give the VR operator greater control over which result set they are exploring, new query results from the desktop operator are queued up rather than pushed directly into virtual space, allowing the VR operator to switch back and forth between result sets at will.

By allowing the desktop operator to focus on query formulation and refinement, and the VR operator to focus on exploration and submission we expect to overcome the challenges of one interface modality with the strengths of the other. Through this simultaneous query refinement and result exploration, we expect to increase efficiency of the interactive retrieval process.

3 INTERACTION MODES

The following outlines the interaction modes available to the two operators through each of the two interfaces and how they can be used jointly in order to retrieve desired results.

3.1 Query Formulation

As discussed in the context of prior participation to this evaluation campaign [7, 8], vitrivr offers a broad range of query modalities, all of which are accessible through vitrivr-ng. The following provides and overview of the modalities used in this instance.

Query-by-Sketch (QbS) uses rough visual approximations for querying. Sketches can contain color or edge information that is compared to the dataset images without taking any higher order semantic information into account. Due to the types of tasks used in this challenge, where no visual information is shown, sketch-based queries are rarely useful by themselves. They can however be of use in combination with other query modalities to further narrow a search.

Tag-based querying uses a fixed set of pre-defined tags from which a user can select. Since the number of tags known to the retrieval system can be large, the UI offers a text-completion mechanism for easy selection. The tags used in this instance come from the tag and location columns of the provided metadata.

Map-based querying uses the GPS coordinates provided with the dataset and compares them with a user-defined point specified on a map.

Full-text search is used to search for relevant overlaps between a user-provided query string and the text provided in the OCR and caption-columns of the dataset.

Visual-text co-embeddings also use textual input, but rather than comparing strings directly, they use several multi-modal embeddings to project the input's semantics into a vector space which is populated by information extracted from the images. In this instance, we use both a custom visual-text co-embedding [12] as well as an OpenCLIP [1] model trained on LAION-5B [11].

Temporal querying enables the combination of several of the above query modalities and to specify a temporal dependence between multiple instances of them.

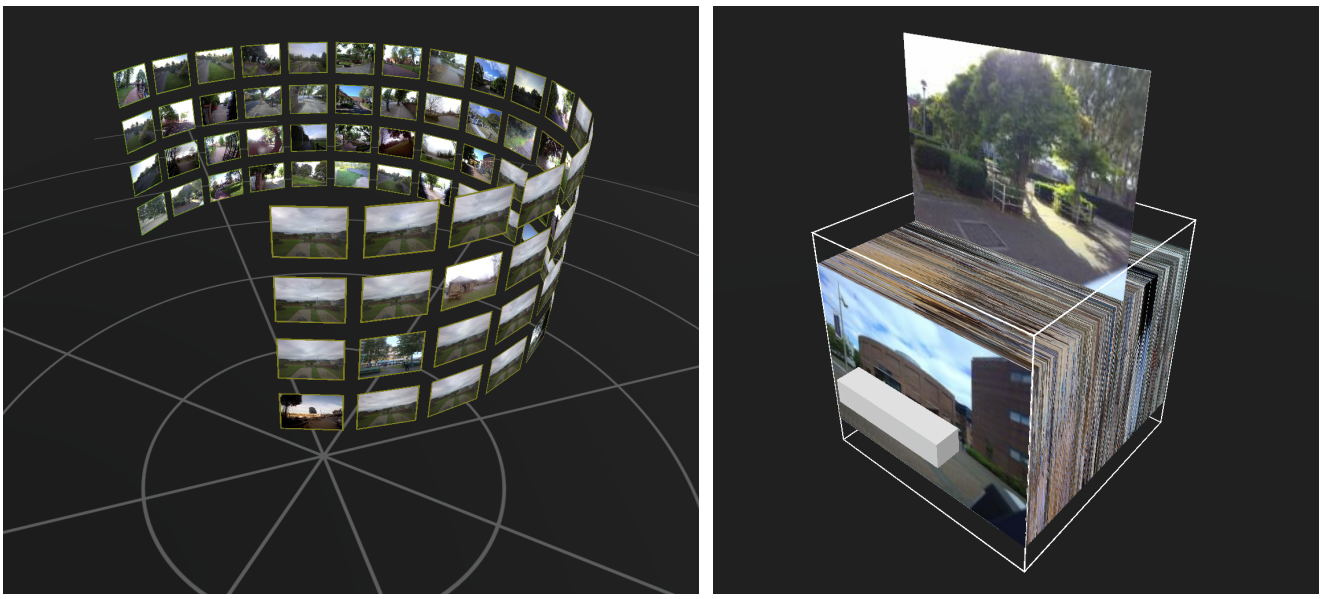

Figure 2: Result exploration methods in VR. Left: cylindrical results display showing result segments cylindrically around the user. Right: multimedia drawer showing lifelog images temporally ordered within the wireframe drawer.

3.2 Result exploration

Results exploration for the hybrid system is provided by vitrivr-VR. Many of the available exploration modes have been described in the context of previous participation in the LSC [13, 15]. In the following, we give an overview of the result exploration methods available through vitrivr-VR.

The cylindrical result view provides a first level of results exploration. Results are arranged in a grid-like fashion cylindrically around the user, ordered by their query score. An example of this is shown on the left side of Figure 2. Different versions of the result view allow grouping by days or for temporal queries by temporal sequence. The view grouped by day displays the highest scoring results for each day as stacks outwards of the cylinder ordered by the maximum score of all items of that day. The temporal sequence grouped view displays the items belonging to a retrieved temporal sequence in a similar stacked way to the day grouped results, ordered by the score of the temporal sequence.

The detail view allows the detailed examination of an item from the results after an item of interest has been discovered in the result view. Besides inspection of result images in higher resolution, this view gives access to additional functions. It allows the display of associated metadata, which can be used to gain insight into the viewed item, and to open the multimedia drawer view, which shows the selected image in the temporal context of the images recorded before and after.

The multimedia drawer enables the exploration of the images of a single day in temporal order. Resembling a wireframe drawer in appearance, the multimedia drawer contains the temporally ordered images of a single day as a horizontal stack, reminiscent of files in a real drawer, as shown on the right side of Figure 2. Hovering over an image in 3D space raises it above the drawer, allowing a sequence of images to be explored quickly by moving a hand through the drawer, similar to a flip-book. By extending the drawer via its handle it can be expanded, making it easier to view individual images. Any image of interest can be selected in the multimedia drawer to create a detail view for further inspection.

Both detail views and multimedia drawers are grab- and placeable 3D objects in the virtual space, allowing them to be positioned anywhere in virtual space. Unless closed manually, these persist between queries and can be used for reference or simply saved for later submission.

4 CONCLUSION

In this paper, we described a novel desktop-virtual reality hybrid multimedia retrieval system built on the vitrivr stack. Past research has shown the advantages and the challenges of desktop and VR interfaces when used independently. By leveraging the fast query formulation of the desktop-based vitrivr-ng and the immersive browsing capabilities afforded by vitrivr-VR in virtual reality, we aim to combine the advantages of both modalities. We hope to gain insight into if and how the challenges of individual user interface modalities can be overcome, by utilizing the strengths of others.

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] Mehdi Cherti, Romain Beaumont, Ross Wightman, Mitchell Wortsman, Gabriel Ilharco, Cade Gordon, Christoph Schuhmann, Ludwig Schmidt, and Jenia Jitsev. 2022. Reproducible scaling laws for contrastive language-image learning. *CoRR* abs/2212.07143 (2022), 39 pages. https://doi.org/10.48550/arXiv.2212.07143 arXiv:2212.07143

[2] Aaron Duane and Björn Pór Jónsson. 2022. ViRMA: Virtual Reality Multimedia Analytics. In *Proceedings of the 2022 International Conference on Multimedia Retrieval*. ACM, Newark NJ USA, 211–214. https://doi.org/10.1145/3512527.3531352

[3] 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

[4] Ralph Gasser, Luca Rossetto, Silvan Heller, and Heiko Schuldt. 2020. Cottontail DB: An Open Source Database System for Multimedia Retrieval and Analysis. In *MM '20: The 28th ACM International Conference on Multimedia, Virtual Event / Seattle, WA, USA, October 12-16, 2020*. ACM, Seattle WA USA, 4465–4468. https://doi.org/10.1145/3394171.3414538

[5] Ralph Gasser, Luca Rossetto, and Heiko Schuldt. 2019. Multimodal Multimedia Retrieval with vitrivr. In *Proceedings of the 2019 on International Conference on Multimedia Retrieval, ICMR 2019, Ottawa, ON, Canada, June 10-13, 2019*, Abdulmotaleb El-Saddik, Alberto Del Bimbo, Zhongfei Zhang, Alexander G. Hauptmann, K. Selçuk Candan, Marco Bertini, Lexing Xie, and Xiao-Yong Wei (Eds.). ACM, Ottawa, ON, Canada, 391–394. https://doi.org/10.1145/3323873.3326921

[6] Cathal Gurrin, Björn Þór Jónsson, Duc Tien Dang Nguyen, Graham Healy, Jakub Lokoc, Liting Zhou, Luca Rossetto, Minh-Triet Tran, Wolfgang Hürst, Werner Bailer, and Klaus Schoeffmann. 2023. Introduction to the Sixth Annual Lifelog Search Challenge, LSC'23. In *Proceedings of the 2023 International Conference on Multimedia Retrieval* (Thessaloniki, Greece) *(ICMR '23)*. Association for Computing Machinery, New York, NY, USA.

[7] Silvan Heller, Ralph Gasser, Mahnaz Parian-Scherb, Sanja Popovic, Luca Rossetto, Loris Sauter, Florian Spiess, and Heiko Schuldt. 2021. Interactive Multimodal Lifelog Retrieval with vitrivr at LSC 2021. In *Proceedings of the 4th Annual on Lifelog Search Challenge, LSC@ICMR 2021, Taipei, Taiwan, 21 August 2021*. ACM, Taipei, Taiwan, 35–39. https://doi.org/10.1145/3463948.3469062

[8] Silvan Heller, Luca Rossetto, Loris Sauter, and Heiko Schuldt. 2022. vitrivr at the Lifelog Search Challenge 2022. In *LSC@ICMR 2022: Proceedings of the 5th Annual on Lifelog Search Challenge, Newark, NJ, USA, June 27 - 30, 2022*. ACM, Newark, NJ, USA, 27–31. https://doi.org/10.1145/3512729.3533003

[9] Luca Rossetto, Ivan Giangreco, and Heiko Schuldt. 2014. Cineast: A Multi-feature Sketch-Based Video Retrieval Engine. In *2014 IEEE International Symposium on Multimedia, ISM 2014, Taichung, Taiwan, December 10-12, 2014*. IEEE Computer Society, Taichung, Taiwan, 18–23. https://doi.org/10.1109/ISM.2014.38

[10] Loris Sauter, Ralph Gasser, Silvan Heller, Luca Rossetto, Colin Saladin, Florian Spiess, and Heiko Schuldt. 2021. Exploring Effective Interactive Text-based Video Search in vitrivr. In *MultiMedia Modeling - 29th International Conference, MMM 2023, Bergen, Norway, January 9-12, 2023, Proceedings, Part II (Lecture Notes in Computer Science)*. Springer, Bergen, Norway, 6 pages.

[11] 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. *CoRR* abs/2210.08402 (2022), 50 pages. https://doi.org/10.48550/arXiv.2210.08402 arXiv:2210.08402

[12] Florian Spiess, Ralph Gasser, Silvan Heller, Luca Rossetto, Loris Sauter, and Heiko Schuldt. 2021. Competitive Interactive Video Retrieval in Virtual Reality with Vitrivr-VR. In *MultiMedia Modeling*. Springer, Prague, Czech Republic, 441–447. https://doi.org/10.1007/978-3-030-67835-7_42

[13] 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 *Proceedings of the 4th Annual on Lifelog Search Challenge, LSC@ICMR 2021, Taipei, Taiwan, 21 August 2021*. ACM, Taipei, Taiwan, 17–22. https://doi.org/10.1145/3463948.3469061

[14] Florian Spiess, Ralph Gasser, Silvan Heller, Heiko Schuldt, and Luca Rossetto. 2023. A Comparison of Video Browsing Performance between Desktop and Virtual Reality Interfaces. In *Proceedings of the 2023 International Conference on Multimedia Retrieval* (Thessaloniki, Greece) *(ICMR '23)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3591106.3592292

[15] Florian Spiess and Heiko Schuldt. 2022. Multimodal Interactive Lifelog Retrieval with vitrivr-VR. In *LSC@ICMR 2022: Proceedings of the 5th Annual on Lifelog Search Challenge, Newark, NJ, USA, June 27 - 30, 2022*. ACM, Newark, NJ, USA, 38–42. https://doi.org/10.1145/3512729.3533008