vitrivr at the Lifelog Search Challenge 2022

Silvan Heller
University of Basel
Switzerland
silvan.heller@unibas.ch

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

Loris Sauter
University of Basel
Switzerland
loris.sauter@unibas.ch

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

ABSTRACT

In this paper, we present the iteration of the multimedia retrieval system vitrivr participating at LSC 2022. vitrivr is a general-purpose retrieval system which has previously participated at LSC. We describe the system architecture and functionality, and show initial results based on the test and validation topics.

CCS CONCEPTS

• **Information systems** → **Search interfaces**; **Image search**; **Search interfaces**; **Image search**; **Users and interactive retrieval**; *Multimedia databases*; **Information retrieval**; • **Human-centered computing** → **Interactive systems and tools**.

KEYWORDS

Content-based Retrieval, Multimedia Retrieval, Lifelogging, Lifelog Search Challenge

ACM Reference Format:
Silvan Heller, Luca Rossetto, Loris Sauter, and Heiko Schuldt. 2022. vitrivr at the Lifelog Search Challenge 2022. In *Proceedings of the 5th Annual Lifelog Search Challenge (LSC '22), June 27–30, 2022, Newark, NJ, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3512729.3533003

1 INTRODUCTION

The past decade has seen a massive growth in the adoption of devices which allow us to capture and measure more and more personal data. In addition to smartphones, wearables which capture sensor data such as heart rate, blood oxygen, etc., have grown in popularity. The ability to search in one's own *Lifelog* has also inspired fictional works such as Black Mirror [38]. The topic has also gained increasing attention in research starting with the MyLifeBits project [4], with one of the key challenges being the development of systems which enable Lifelog retrieval in an efficient and user-friendly manner.

One of the ways to evaluate research contributions in the area of retrieval systems is through interactive evaluation campaigns [17] such as the Lifelog Search Challenge (LSC) [5–8] or the Video Browser Showdown (VBS) [1, 12, 18–20, 25, 31], which have also been held entirely remotely in recent years [12, 24] thanks to new and improved software for evaluation campaigns [26]. For the fifth edition of LSC, the dataset has increased in size, and now consists of 18 months of lifelog data [6].

In this paper, we present the vitrivr[1] multimedia retrieval system [28], which has participated in previous iterations of LSC [11, 14, 23]. We introduce vitrivr's system architecture and retrieval model and demonstrate how lifelog queries are solved with vitrivr using example queries. Moreover, we also describe how vitrivr make use of the available metadata and the annotations provided with the new LSC dataset.

The remainder of this paper is structured as follows: Section 2 gives an overview of vitrivr's system architecture and describes its functionality. Section 3 shows how vitrivr can be used for the kinds of queries typically found during LSC, and Section 4 concludes.

2 VITRIVR

vitrivr is a "*a scalable, content-based multimedia information retrieval stack*" [3]. Its general purpose nature allows usage in video retrieval [9, 35], cultural heritage [22, 30], and it is a regular participant at LSC [11, 14, 23] with its core architecture unchanged in the last few iterations. The system has three core components: a database specialized for multimedia retrieval, Cottontail DB [2], a retrieval engine, Cineast [27], and the web-based user interface, vitrivr-ng. This modular structure allows for frontends which have different interaction paradigms [13, 29, 32–34]. In this section, we provide a brief self-contained description of the system and its usage for lifelog retrieval.

2.1 Architecture

Figure 1 shows the architecture of vitrivr, with the three major components being the following:

Database: Cottontail DB is a database management system for multimedia retrieval. We make use of the Boolean retrieval functionality for metadata such as weekday, text retrieval for the provided descriptions and vector-based similarity search for visual similarity search and text embeddings.

Retrieval Engine: Cineast is the feature extraction and query processing engine of the vitrivr stack. It supports feature extraction and retrieval across various modalities, supporting different notions of similarity and has support for complex

[1] https://vitrivr.org

LSC '22, June 27–30, 2022, Newark, NJ, USA

ACM ISBN 978-1-4503-9239-6/22/06.
https://doi.org/10.1145/3512729.3533003

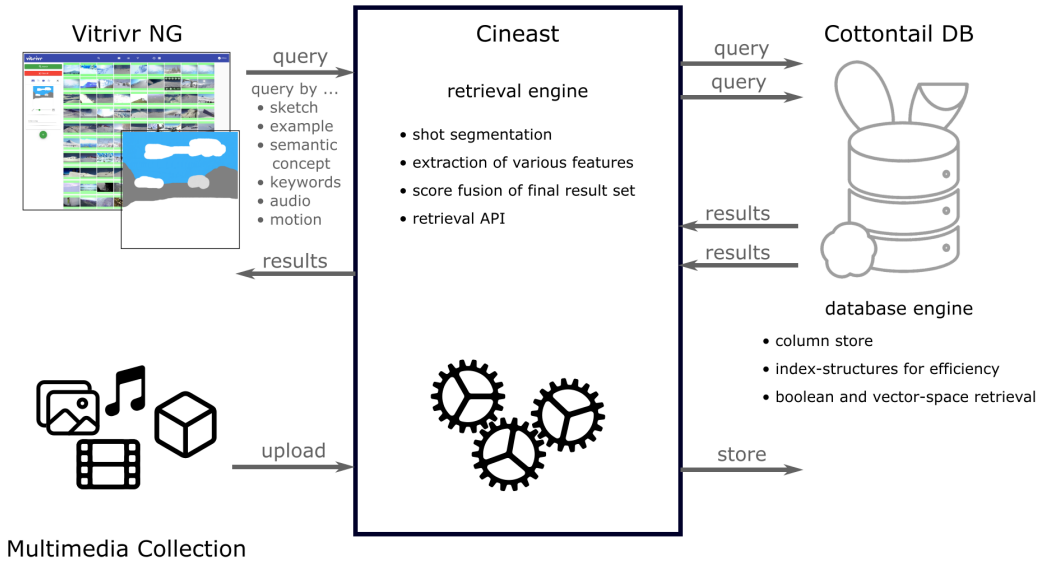

Figure 1: System overview for vitrivr and its three major components: vitrivr-ng, Cineast and Cottontail DB [10]

queries expressing dependencies between modalities [15] and temporal dependencies [9].

Frontend: vitrivr-ng is "responsible for query formulation, result presentation, browsing and filtering. Users can combine various query modalities, for example, first filtering by time and then making a color sketch" [11]. Figures 2 and 3 show the current UI.

2.2 Existing Functionality for Lifelog Retrieval

vitrivr has already participated to the LSC in previous years [11, 14, 23], and in this section we concisely review and summarize the extensions made in previous participations.

Features for Lifelog Retrieval. We use the following features for LSC, which can be combined in various ways [15]:

Text Embedding: We extract textual embeddings from the images [32] using a similar approach to W2VV++ [16], as well as a feature based on OpenAI's CLIP [21], both of which can then be queried using textual input. We use the vector retrieval functionality of Cottontail DB for this purpose.

Text Retrieval: We use the text retrieval functionality of Cottontail DB and Cineast to query the provided textual descriptions and OCR data. We use both the provided OCR data as well as our own HyText [37].

Visual Similarity Search: vitrivr offers several features for content-based image retrieval, such as one based on InceptionResnetV2 [36]. We utilise this functionality for iterative search based on previously retrieved results.

Boolean Retrieval: vitrivr supports classical operators such as range queries and comparisons, which are useful for all provided metadata such as weekdays.

Units of Retrieval. For lifelog retrieval, we return results in three different levels of detail:

Segment: For the smallest unit of retrieval for lifelog, we assign a score to each individual image based on various features mentioned above.

Object: The data model of vitrivr is centered around *media objects* (such as videos, an image, or an audio file). For LSC, we aggregate all images within a day: The score of an object is determined through the fusion of its segments.

Temporal Sequence: We allow users to specify temporal context [9], and return sequences of segments which fit the given query.

Result Filters. We allow users to use Boolean filter predicates (e.g., time of day, weekday) during query formulation time, and additionally provide late filtering for available metadata. Since the format of LSC is such that new information becomes available during the course of a task, this avoids re-formulation and re-execution of a query.

3 EXAMPLE QUERIES AND RESULTS

In this section, we show how example tasks could be solved with vitrivr.

3.1 Textual Known-Item Search

The following topic was part of the LSC 2021 installment and has a set of known true positives within the dataset. The first hint was displayed immediately upon start of the topic, subsequently every 30 seconds, the hint text was updated.

> "*Planning a thesis/dissertation on a whiteboard with my PhD student ...*"

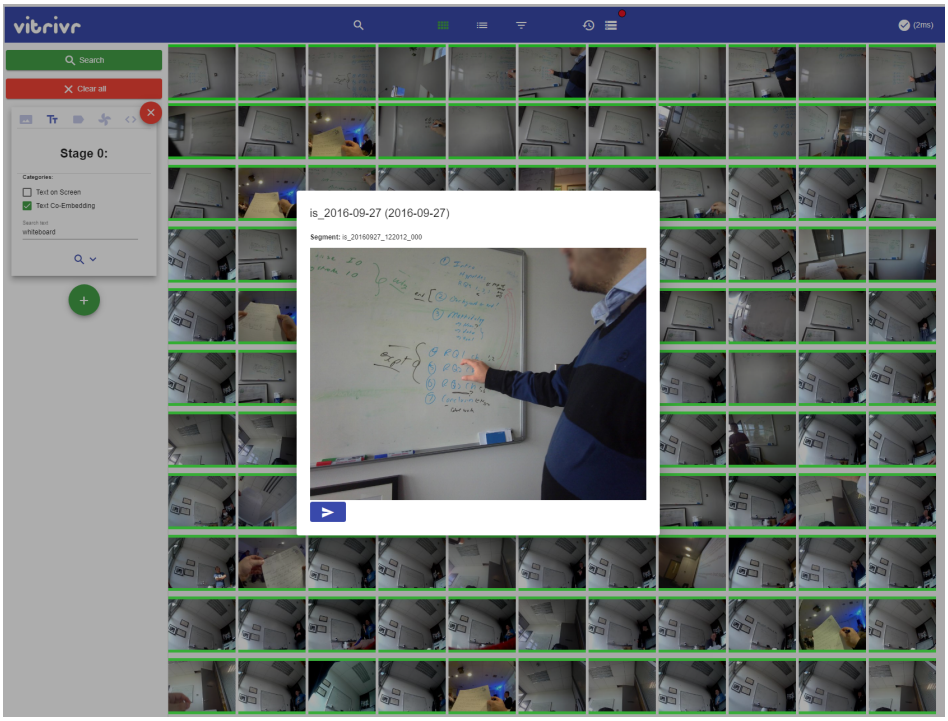

Figure 2: A screenshot of the vitrivr system with the search modality text co-embedding and the query "*whiteboard*". One of the correct solutions for the task is opened for close inspection.

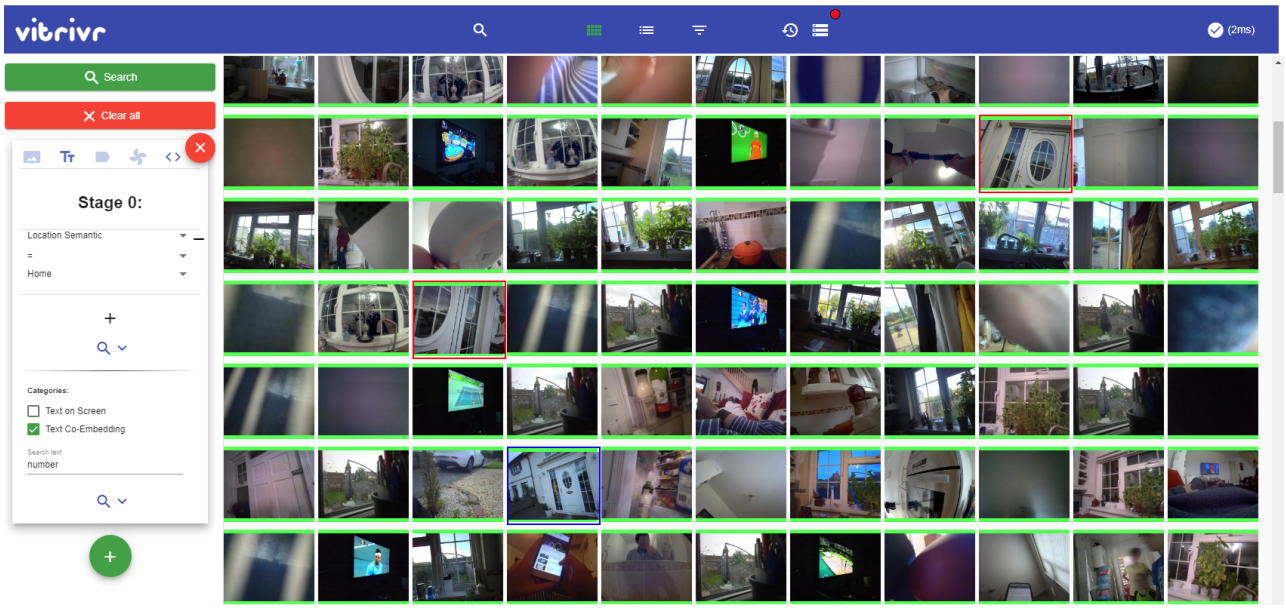

Figure 3: A screenshot of the vitrivr system that shows how to solve the question to find the lifelogger's house number. In particular, there were two results (framed in red) that did not show the entire number. Ultimately, in a result with a slightly lower rank on row six (framed in blue), the number is visible enough for the operator.

"*Planning a thesis/dissertation on a whiteboard with my PhD student, who was wearing a blue and black stripey top ...*"

"*Planning a thesis/dissertation on a whiteboard with my PhD student, who was wearing a blue and black stripey top in my office in 2016 ...*"

"*Planning a thesis/dissertation on a whiteboard with my PhD student, who was wearing a blue and black stripey top in my office in 2016. We were using blue, black and green pens ...*"

"*Planning a thesis/dissertation on a whiteboard with my PhD student, who was wearing a blue and black stripey top in my office in 2016. We were using blue, black and green pens. After this I went back to work at my computer ...*"

"*Planning a thesis/dissertation on a whiteboard with my PhD student, who was wearing a blue and black stripey top in my office in 2016. We were using blue, black and green pens. After this I went back to work at my computer. It was on the 27th September.*"

With the first hint, the keyword "*whiteboard*" stands out to the operator, so the operator chooses to use the modality text co-embedding. Until now, a few seconds have passed and they do not yet know further hints, so they explore the result set. After thirty seconds, the confirmation that the correct results are retrieved comes in the form of the second hint: "*... blue and black stripey top ...*". After close inspection, the operator submits the selected image and immediately receives the feedback that this submission is correct. The last hint would have also confirmed that the date matches the requested one.

3.2 Descriptive Queries

In 2022 two new formats for topics will be introduced, descriptive queries and question and answer topics. We expect queries similar to the following:

"*Find items when I drink coffee in my office in the afternoon*"

To solve such topics, we leverage the multi-modality of vitrivr. A combination of Boolean search (hour between 12 and 18, semantic location of "*work*") and text co-embedding (keyword: "*drinking coffee*") can be used to find such results. In the result exploration phase, the user browses the result set to submit results which match the topic. In doing so, results where the lifelogger is drinking other beverages than coffee are not submitted.

3.3 Q&A Topics

In the new format, there might also be analytical questions which can be answered either through text or judged by humans rather than with a pre-defined known set of true positives from within the dataset.

Based on the description and the LSC21 dataset, we expect topics such as those below:

"*What is my house number?*"

The formulation of this first question enables us to know that the lifelogger's house has a door with a number, presumably the outside door. The data set is annotated with so-called semantic locations, one of them with the label "*home*". These semantic locations are available through Boolean search. Additionally, we are looking for a door with the house number and thus use the keyword "*door*" for a text co-embedding query. As there are too many images of interior doors, we quickly replace the the keyword with "*number*" to find the answer to the question with a little browsing and inspection work, as shown in the screenshot in Figure 3. We do not provide the answer here for privacy reasons.

"*What is my car's registration number?*"

The second question, the one for the car registration, can also be solved with vitrivr by two subsequent, distinct queries. The first one is used, again with the semantic location of "*home*", to gain knowledge about the lifelogger's car (we learn it is a white Volvo). Subsequently, the text co-embedding search for "*number plate*" yields results with the car more in focus. To then finally find the number, one of the results with the car in focus, a More-Like-This query is issued, and the question can be answered with the fifth result (again, for privacy reasons we do not provide the answer here).

4 CONCLUSION

2022 marks the fourth participation of the vitrivr system in the LSC. While no significant changes have been made to vitrivr, such interactive evaluation campaigns offer an exciting opportunity to benchmark retrieval systems in real-world scenarios, and the analysis of the competition results yields interesting results and recommendations for future work. Additionally, vitrivr can serve as a baseline for its virtual reality counterpart, vitrivr-VR [34]. Having achieved the highest score in 2019 and competitive results in 2020 and 2021, we hope to gain further insights into the performance of our general purpose retrieval model and functionality through the participation at LSC.

REFERENCES

[1] Claudiu Cobârzan, Klaus Schoeffmann, Werner Bailer, Wolfgang Hürst, Adam Blažek, Jakub Lokoč, Stefanos Vrochidis, Kai Uwe Barthel, and Luca Rossetto. 2017. Interactive Video Search Tools: A Detailed Analysis of the Video Browser Showdown 2015. *Multimedia Tools and Applications* 76, 4 (2017), 5539–5571. https://doi.org/10.1007/s11042-016-3661-2

[2] 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 (MM)*. Association for Computing Machinery, New York, NY, USA, 4465–4468. https://doi.org/10.1145/3394171.3414538

[3] Ralph Gasser, Luca Rossetto, and Heiko Schuldt. 2019. Multimodal Multimedia Retrieval with Vitrivr. In *International Conference on Multimedia Retrieval (ICMR)*. Association for Computing Machinery, New York, NY, USA, 391–394. https://doi.org/10.1145/3323873.3326921

[4] Jim Gemmell, Gordon Bell, Roger Lueder, Steven Drucker, and Curtis Wong. 2002. MyLifeBits: Fulfilling the Memex Vision. In *International Conference on Multimedia (MM)*. Association for Computing Machinery, New York, NY, USA, 235–238. https://doi.org/10.1145/641007.641053

[5] Cathal Gurrin, Björn Þór Jónsson, Klaus Schöffmann, Duc-Tien Dang-Nguyen, Jakub Lokoč, Minh-Triet Tran, Wolfgang Hürst, Luca Rossetto, and Graham Healy.

2021. Introduction to the Fourth Annual Lifelog Search Challenge, LSC'21. In *International Conference on Multimedia Retrieval (ICMR)*. Association for Computing Machinery, New York, NY, USA, 690–691. https://doi.org/10.1145/3460426.3470945

[6] Cathal Gurrin, Björn Þór Jónsson, Klaus Schöffmann, Duc-Tien Dang-Nguyen, Jakub Lokoč, Minh-Triet Tran, Wolfgang Hürst, Luca Rossetto, and Graham Healy. 2022. Introduction to the Fifth Annual Lifelog Search Challenge, LSC'22. In *International Conference on Multimedia Retrieval (ICMR)*. Association for Computing Machinery, New York, NY, USA.

[7] Cathal Gurrin, Tu-Khiem Le, Van-Tu Ninh, Duc-Tien Dang-Nguyen, Björn Þór Jónsson, Jakub Lokoč, Wolfgang Hürst, Minh-Triet Tran, and Klaus Schöffmann. 2020. Introduction to the Third Annual Lifelog Search Challenge (LSC'20). In *International Conference on Multimedia Retrieval (ICMR)*. Association for Computing Machinery, New York, NY, USA, 584–585. https://doi.org/10.1145/3372278.3388043

[8] Cathal Gurrin, Klaus Schoeffmann, Hideo Joho, Andreas Leibetseder, Liting Zhou, Aaron Duane, Duc-Tien Dang-Nguyen, Michael Riegler, Luca Piras, Minh-Triet Tran, Jakub Lokoč, and Wolfgang Hürst. 2019. [Invited Papers] Comparing Approaches to Interactive Lifelog Search at the Lifelog Search Challenge (LSC2018). *ITE Transactions on Media Technology and Applications* 7, 2 (2019), 46–59. https://doi.org/10.3169/mta.7.46

[9] Silvan Heller, Rahel Arnold, Ralph Gasser, Viktor Gsteiger, Mahnaz Parian-Scherb, Luca Rossetto, Loris Sauter, Florian Spiess, and Heiko Schuldt. 2022. Multi-Modal Interactive Video Retrieval with Temporal Queries. In *MultiMedia Modeling*. Springer International Publishing, Cham, 493–498. https://doi.org/10.1007/978-3-030-98355-0_44

[10] Silvan Heller, Ralph Gasser, Cristina Illi, Maurizio Pasquinelli, Loris Sauter, Florian Spiess, and Heiko Schuldt. 2021. Towards Explainable Interactive Multi-modal Video Retrieval with Vitrivr. In *MultiMedia Modeling*. Springer International Publishing, Cham, 435–440. https://doi.org/10.1007/978-3-030-67835-7_41

[11] 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 *Workshop on Lifelog Search Challenge*. Association for Computing Machinery, New York, NY, USA, 35–39. https://doi.org/10.1145/3463948.3469062

[12] Silvan Heller, Viktor Gsteiger, Werner Bailer, Cathal Gurrin, Björn Þór Jónsson, Jakub Lokoč, Andreas Leibetseder, František Mejzlík, Ladislav Peška, Luca Rossetto, Konstantin Schall, Klaus Schoeffmann, Heiko Schuldt, Florian Spiess, Ly-Duyen Tran, Lucia Vadicamo, Patrik Veselý, Stefanos Vrochidis, and Jiaxin Wu. 2022. Interactive Video Retrieval Evaluation at a Distance: Comparing Sixteen Interactive Video Search Systems in a Remote Setting at the 10th Video Browser Showdown. *International Journal of Multimedia Information Retrieval* 11, 1 (2022), 1–18. https://doi.org/10.1007/s13735-021-00225-2

[13] Silvan Heller, Mahnaz Parian, Maurizio Pasquinelli, and Heiko Schuldt. 2020. Vitrivr-Explore: Guided Multimedia Collection Exploration for Ad-hoc Video Search. In *Similarity Search and Applications (SISAP)*, Vol. 12440. Springer International Publishing, Cham, 379–386. https://doi.org/10.1007/978-3-030-60936-8_30

[14] Silvan Heller, Mahnaz Amiri Parian, Ralph Gasser, Loris Sauter, and Heiko Schuldt. 2020. Interactive Lifelog Retrieval with Vitrivr. In *Third Annual Workshop on Lifelog Search Challenge*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3379172.3391715

[15] Silvan Heller, Loris Sauter, Heiko Schuldt, and Luca Rossetto. 2020. Multi-Stage Queries and Temporal Scoring in Vitrivr. In *IEEE International Conference on Multimedia Expo Workshops (ICMEW)*. IEEE, New Jersey, 1–5. https://doi.org/10.1109/ICMEW46912.2020.9105954

[16] Xirong Li, Chaoxi Xu, Gang Yang, Zhineng Chen, and Jianfeng Dong. 2019. W2VV++: Fully Deep Learning for Ad-hoc Video Search. In *International Conference on Multimedia (MM)*. Association for Computing Machinery, New York, NY, USA, 1786–1794. https://doi.org/10.1145/3343031.3350906

[17] Jakub Lokoč, Werner Bailer, Kai Uwe Barthel, Cathal Gurrin, Silvan Heller, Björn þór Jónsson, Ladislav Peška, Luca Rossetto, Klaus Schoeffmann, Lucia Vadicamo, Stefanos Vrochidis, and Jiaxin Wu. 2022. A Task Category Space for User-Centric Comparative Multimedia Search Evaluations. In *MultiMedia Modeling*. Springer International Publishing, Cham, 193–204. https://doi.org/10.1007/978-3-030-98358-1_16

[18] Jakub Lokoč, Werner Bailer, Klaus Schoeffmann, Bernd Muenzer, and George Awad. 2018. On Influential Trends in Interactive Video Retrieval: Video Browser Showdown 2015–2017. *IEEE Transactions on Multimedia* 20, 12 (2018), 3361–3376. https://doi.org/10.1109/TMM.2018.2830110

[19] Jakub Lokoč, Gregor Kovalčík, Bernd Münzer, Klaus Schöffmann, Werner Bailer, Ralph Gasser, Stefanos Vrochidis, Phuong Anh Nguyen, Sitapa Rujikietgumjorn, and Kai Uwe Barthel. 2019. Interactive Search or Sequential Browsing? A Detailed Analysis of the Video Browser Showdown 2018. *ACM Transactions on Multimedia Computing, Communications, and Applications* 15, 1 (2019), 29:1–29:18. https://doi.org/10.1145/3295663

[20] Jakub Lokoč, Patrik Veselý, František Mejzlík, Gregor Kovalčík, Tomáš Souček, Luca Rossetto, Klaus Schoeffmann, Werner Bailer, Cathal Gurrin, Loris Sauter, Jaeyub Song, Stefanos Vrochidis, Jiaxin Wu, and Björn þóR Jónsson. 2021. Is the Reign of Interactive Search Eternal? Findings from the Video Browser Showdown 2020. *ACM Transactions on Multimedia Computing, Communications, and Applications* 17, 3 (2021), 91:1–91:26. https://doi.org/10.1145/3445031

[21] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. *arXiv:2103.00020 [cs]* (2021).

[22] Laura Rettig, Shaban Shabani, Loris Sauter, Philippe Cudré-Mauroux, Maria Sokhn, and Heiko Schuldt. 2021. City-Stories: Combining Entity Linking, Multimedia Retrieval, and Crowdsourcing to Make Historical Data Accessible. In *Web Engineering*. Springer International Publishing, Cham, 521–524. https://doi.org/10.1007/978-3-030-74296-6_43

[23] Luca Rossetto, Ralph Gasser, Silvan Heller, Mahnaz Amiri Parian, and Heiko Schuldt. 2019. Retrieval of Structured and Unstructured Data with Vitrivr. In *Workshop on Lifelog Search Challenge*. Association for Computing Machinery, New York, NY, USA, 27–31. https://doi.org/10.1145/3326460.3329160

[24] Luca Rossetto, Ralph Gasser, Silvan Heller, Mahnaz Parian-Scherb, Loris Sauter, Florian Spiess, Heiko Schuldt, Ladislav Peška, Tomáš Souček, Miroslav Kratochvíl, František Mejzlík, Patrik Veselý, and Jakub Lokoč. 2021. On the User-Centric Comparative Remote Evaluation of Interactive Video Search Systems. *IEEE MultiMedia* 28, 4 (2021), 18–28. https://doi.org/10.1109/MMUL.2021.3066779

[25] Luca Rossetto, Ralph Gasser, Jakub Lokoč, Werner Bailer, Klaus Schoeffmann, Bernd Muenzer, Tomáš Souček, Phuong Anh Nguyen, Paolo Bolettieri, Andreas Leibetseder, and Stefanos Vrochidis. 2021. Interactive Video Retrieval in the Age of Deep Learning – Detailed Evaluation of VBS 2019. *IEEE Transactions on Multimedia* 23 (2021), 243–256. https://doi.org/10.1109/TMM.2020.2980944

[26] Luca Rossetto, Ralph Gasser, Loris Sauter, Abraham Bernstein, and Heiko Schuldt. 2021. A System for Interactive Multimedia Retrieval Evaluations. In *MultiMedia Modeling*. Springer International Publishing, Cham, 385–390. https://doi.org/10.1007/978-3-030-67835-7_33

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

[28] Luca Rossetto, Ivan Giangreco, Claudiu Tanase, and Heiko Schuldt. 2016. Vitrivr: A Flexible Retrieval Stack Supporting Multiple Query Modes for Searching in Multimedia Collections. In *International Conference on Multimedia (MM)*. Association for Computing Machinery, New York, NY, USA, 1183–1186. https://doi.org/10.1145/2964284.2973797

[29] Luca Rossetto, Ivan Giangreco, Claudiu Tănase, Heiko Schuldt, Stéphane Dupont, Omar Seddati, Metin Sezgin, and Yusuf Sahillioğlu. 2016. iAutoMotion – an Autonomous Content-Based Video Retrieval Engine. In *MultiMedia Modeling*. Springer International Publishing, Cham, 383–387. https://doi.org/10.1007/978-3-319-27674-8_37

[30] Loris Sauter, Luca Rossetto, and Heiko Schuldt. 2018. Exploring Cultural Heritage in Augmented Reality with GoFind!. In *IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. IEEE, New Jersey, 187–188. https://doi.org/10.1109/AIVR.2018.00041

[31] Klaus Schoeffmann. 2019. Video Browser Showdown 2012-2019: A Review. In *International Conference on Content-Based Multimedia Indexing (CBMI)*. IEEE, New Jersey, 1–4. https://doi.org/10.1109/CBMI.2019.8877397

[32] Florian Spiess, Ralph Gasser, Silvan Heller, Mahnaz Parian-Scherb, Luca Rossetto, Loris Sauter, and Heiko Schuldt. 2022. Multi-Modal Video Retrieval in Virtual Reality with Vitrivr-VR. In *MultiMedia Modeling*. Springer International Publishing, Cham, 499–504. https://doi.org/10.1007/978-3-030-98355-0_45

[33] 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 International Publishing, Cham, 441–447. https://doi.org/10.1007/978-3-030-67835-7_42

[34] 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*. Association for Computing Machinery, New York, NY, USA, 17–22. https://doi.org/10.1145/3463948.3469061

[35] Newton Spolaôr, Huei Diana Lee, Weber Shoity Resende Takaki, Leandro Augusto Ensina, Antonio Rafael Sabino Parmezan, Jefferson Tales Oliva, Claudio Saddy Rodrigues Coy, and Feng Chung Wu. 2021. A Video Indexing and Retrieval Computational Prototype Based on Transcribed Speech. *Multimedia Tools and Applications* 80, 25 (2021), 33971–34017. https://doi.org/10.1007/s11042-021-11401-1

[36] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alex Alemi. 2016. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *arXiv:1602.07261 [cs]* (2016).

[37] Alexander Theus, Luca Rossetto, and Abraham Bernstein. 2022. HyText – A Scene-Text Extraction Method for Video Retrieval. In *MultiMedia Modeling*. Springer International Publishing, Cham, 182–193. https://doi.org/10.1007/978-3-030-98355-0_16

[38] Brian Welsh. 2011. Black Mirror: The Entire History of You. Season 1 Episode 3.