Competitive Video Retrieval with vitrivr

Luca Rossetto, Ivan Giangreco, Ralph Gasser, and Heiko Schuldt

Databases and Information Systems Research Group
Department of Mathematics and Computer Science
University of Basel, Switzerland
{firstname.lastname}@unibas.ch

Abstract. This paper presents the competitive video retrieval capabilities of vitrivr. The vitrivr stack is the continuation of the IMOTION system which participated to the Video Browser Showdown competitions since 2015. The primary focus of vitrivr and its participation in this competition is to simplify and generalize the system's individual components, making them easier to deploy and use. The entire vitrivr stack is made available as open source software.

1 Introduction

In this paper we present the current iteration of vitrivr [6], an open-source content-based multimedia retrieval stack. The vitrivr stack is the continuation of the IMOTION system [3, 5, 7, 8] which participated in previous iterations of the Video Browser Showdown [1]. Despite offering some new functionality, the primary focus for this years participation lies in the simplification and generalization of the retrieval stack in order to make it easier to adapt, deploy, and use by both experts and laymen. The vitrivr stack is available in its entirety from https://vitrivr.org.

The remainder of this paper is structured as follows: Section 2 provides a brief overview of the overall system architecture and Section 3 summarizes all query types supported by vitrivr. Section 4 provides details on the functionalities introduced in the current version. In Section 5, we briefly outline our reasoning behind the open sourcing of vitrivr and Section 6 concludes.

2 Architectural Overview

The vitrivr stack – like its predecessor IMOTION – consists of three primary system components: the storage layer ADAM_{pro} [2], the retrieval engine Cineast [4], and a browser-based user interface. Additionally, a web server is used to serve static content such as videos and thumbnail images. Additional details on the architecture of the entire stack can be found in [6].

3 Interaction and Query Types

The vitrivr stack offers various ways in which queries can be specified. Basically, they can for the most part be grouped into two categories: *visual* and *textual*. The visual query modes include *Query-by-Sketch* and *Query-by-Example* as well as *Relevance Feedback* which are based on visual input such as user generated sketches of a scene or one or multiple previously retrieved scenes. These queries are performed based on data extracted directly from the video frames. The textual queries are based on information which can be extracted from the video content and represented as text, such as spoken language, text on screen, or the provided textual video meta data. For this we use the ASR data provided with the video data set as well as several object detectors to produce labels for the shots. OCR is applied in order to make text which might appear on screen searchable as well.

4 New Functionality

While the IMOTION system that has participated in previous instances of VBS has always been a specialized piece of software, purpose-built for the competition, the functionality we added to vitrivr in preparation for this iteration of VBS are such that they are also useful for other use cases of vitrivr.

4.1 New User Interface

The most salient difference to the IMOTION System of the previous year is the new user interface. While still browser-based, the latest iteration of the UI is based upon the Angular framework[1]. Its modular structure makes it easy to customize the entire UI or parts thereof to shift its focus from general purpose multimedia retrieval to, in this case, competitive video retrieval.

As in the past, the UI enables result streaming in order to be able to already present partial results to the user while the query is still being processed by the backend. It, however, achieves this no longer via AJAX requests but rather uses a WebSocket connection to Cineast. A REST-API is also available. The stack still includes a web server which provides the static content such as shot thumbnails and the videos themselves, but it is no longer required to act as a proxy between the browser and Cineast. The screenshot in Figure 1 depicts the current version of the UI.

4.2 Approximate Retrieval

The underlying storage engine ADAM_{pro} [2] supports multiple index structures for efficient vector space retrieval. Many of these index structures achieve their high efficiency by approximating results rather than producing the true nearest
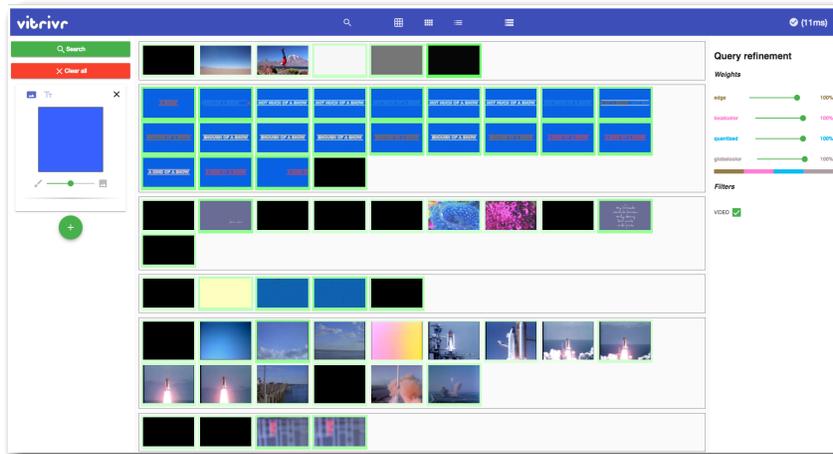
---

[1] https://angular.io/

Fig. 1. Screenshot of the vitrivr UI for large-scale competitive video search

neighbors of a query vector. In previous system iterations, we only made use of exact query results which lead to longer query times. In the current iteration of the system, the choice as to whether exact or approximate queries should be used can be made at query time. Hence, the user can sacrifice some accuracy to gain major speed-ups.

5 Open Source

The entire vitrivr stack [6] is published under the MIT license, the source code of all its components is available from their individual GitHub[2] repositories, additional documentation can be found on https://vitrivr.org. Being a general-purpose multimedia retrieval stack, vitrivr has many applications outside of competitive video retrieval as it also supports other domains such as Images, Audio, and 3D-models. With this flexible open source stack, we hope to offer the community the basis for future research in many areas and domains of multimedia retrieval.

6 Conclusions

With the competitive video retrieval version of vitrivr, we plan to continue the successful participations with we had with the IMOTION system in the past. It is our hope that by publishing the entire retrieval stack as open source software, we lower the entry hurdle for future participants and provide some acceleration for the testing of new ideas in the context of large-scale video retrieval.

---

[2] https://github.com/vitrivr

Acknowledgements

References

1. Claudiu Cobârzan, Klaus Schoeffmann, Werner Bailer, Wolfgang Hürst, Adam Blažek, Jakub Lokoč, Stefanos Vrochidis, Kai Uwe Barthel, and Luca Rossetto. Interactive Video Search Tools: a detailed Analysis of the Video Browser Showdown 2015. *Multimedia Tools and Applications*, 76(4):5539–5571, 2017.
2. Ivan Giangreco and Heiko Schuldt. ADAM$_{pro}$: Database Support for Big Multimedia Retrieval. *Datenbank-Spektrum*, 16(1):17–26, 2016.
3. Luca Rossetto, Ivan Giangreco, Silvan Heller, Claudiu Tănase, Heiko Schuldt, Stéphane Dupont, Omar Seddati, Metin Sezgin, Ozan Can Altıok, and Yusuf Sahillioğlu. IMOTION – Searching for Video Sequences using Multi-shot Sketch Queries. In *Proceedings of the 22nd International Conference on Multimedia Modeling (MMM 2016)*, volume 9517 of *Lecture Notes in Computer Science*, pages 377–382, Miami, FL, USA, January 2016. Springer.
4. Luca Rossetto, Ivan Giangreco, and Heiko Schuldt. Cineast: a Multi-feature Sketch-based Video Retrieval Engine. In *Proceedings of the 2014 IEEE International Symposium on Multimedia (ISM 2014)*, pages 18–23, Taichung, Taiwan, December 2014. IEEE Computer Society.
5. Luca Rossetto, Ivan Giangreco, Heiko Schuldt, Stéphane Dupont, Omar Seddati, Metin Sezgin, and Yusuf Sahillioğlu. IMOTION - A Content-based Video Retrieval Engine. In *Proceedings of the 21st International Conference on Multimedia Modeling (MMM 2015)*, volume 8936 of *Lecture Notes in Computer Science*, pages 255–260, Sydney, NSW, Australia, January 2015. Springer.
6. Luca Rossetto, Ivan Giangreco, Claudiu Tănase, and Heiko Schuldt. vitrivr: A Flexible Retrieval Stack Supporting Multiple Query Modes for Searching in Multimedia Collections. In *Proceedings of the 2016 ACM Conference on Multimedia Conference (ACM MM 2016)*, pages 1183–1186, Amsterdam, The Netherlands, October 2016. ACM.
7. Luca Rossetto, Ivan Giangreco, Claudiu Tănase, and Heiko Schuldt. Multimodal Video Retrieval with the 2017 IMOTION System. In *Proceedings of the 2017 ACM International Conference on Multimedia Retrieval (ICMR 2017)*, pages 457–460, Bucharest, Romania, June 2017. ACM.
8. Luca Rossetto, Ivan Giangreco, Claudiu Tănase, Heiko Schuldt, Stéphane Dupont, and Omar Seddati. Enhanced Retrieval and Browsing in the IMOTION System. In *Proceedings of the 23rd International Conference on Multimedia Modeling (MMM 2017)*, volume 10133 of *Lecture Notes in Computer Science*, pages 469–474, Reykjavik, Iceland, January 2017. Springer.