A New Tool for Collaborative Video Search via Content-based Retrieval and Visual Inspection

Wolfgang Hürst,
Algernon Ip Vai Ching
Utrecht University
Utrecht, The Netherlands
huerst@uu.nl

Marco A. Hudelist,
Manfred J. Primus,
Klaus Schoeffmann
Klagenfurt University
Klagenfurt, Austria
{marco, mprimus,ks}@itec.aau.at

Christian Beecks
RWTH Aachen University
Aachen, Germany
beecks@cs.rwth-aachen.de

ABSTRACT

We present a new approach for collaborative video search and video browsing relying on a combination of traditional, index-based video retrieval complemented with large-scale human-based visual inspection. In particular, a traditional PC interface is used for query-based search using advanced indexing and querying methods (e.g., concept search), whereas a visualization of the database on a tablet is used for pure human-based browsing. Both parts are coupled to compensate for mutual disadvantages; human visual inspection allows for a better, more detailed analysis of the data – also bridging the semantic gap – but comes at the price of an unfiltered database – a disadvantage that is resolved by using the results from the query engine to change visualization order.

Keywords

Video retrieval; feature signatures; collaborative search; human-computer interaction

1. INTRODUCTION

State-of-the-art content-based retrieval tools allow for high quality search in video databases. When entering information about color distribution, concepts, or other features, users are generally presented with a filtered and ranked subset of the data that should contain the information one is looking for. While this works well in many situations, there are still limitations and issues such as the semantic gap [7]. Problems with ambiguous queries or a lack of knowledge about the search goals make it likely that not all of these will ever be resolved completely. Thus, the interface and interaction design plays a crucial role for effective video search. Common standard representations of search results work well if and only if the search engine's ranked results are reliable. Yet, providing more content information to a user can be helpful in situations when this is not the case (e.g., when the target segment is in the same file but at a different position than the one returned by the retrieval tool). We therefore present a combined approach for collaborative video search that takes advantage of both worlds. A high-performing content-based search engine is used for video search by querying. It is coupled with a second tool, operated by a collaborating user. Here, the search results are represented in more detail using a visualization technique optimized for quick, large-scale visual data skimming under consideration of established knowledge about human visual perception.

The basic concept was used in the Video Browser Showdown (VBS) competition [6] in 2016 where it achieved the 2nd place. In this demo, we present its latest incarnation with an extended content-based search, a further optimized visualization tool, and an improved communication between the two collaborating systems. In the following, we will shorty describe the content-based video retrieval part of the system (2.1), illustrate the visualization concept and related interaction of the tablet-based search tool (2.2), and describe how they are integrated in a beneficial way for optimal collaborative video search (2.3).

2. SYSTEM DESCRIPTION
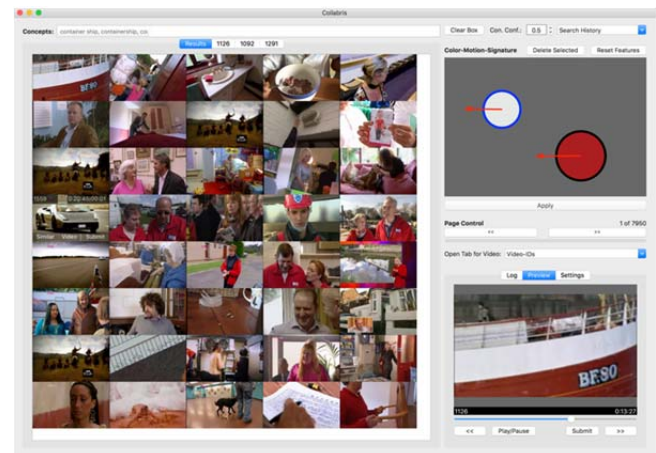
2.1 Content-based Video Retrieval

Figure 1. PC-based CBVR tool.

The content-based video retrieval (CBVB) tool features various query processing functionality, including temporal feature signatures entered via sketching (Fig. 1, top right), concept filtering tools (Fig. 1, top) and chronological segment browsing. Content analysis is done by, for example, using CNNs for concept detection and temporal feature signatures which in turn utilize a variant of the Signature Matching Distance for video shot comparisons [1,5]. The high performance of these search algorithms was demonstrated at the VBS 2016 competition [2] and newer tests done with an enhanced version of the system. Results returned by the search engine are represented in a standard way by ranked segments (Fig. 1, left). Users can further inspect individual results in a small playback area (Fig. 1, bottom right).

Figure 2. Tablet-based visual inspection tool.

2.2 Large-scale Visual Inspection

The traditional way of representing potentially relevant parts of a video as search results is efficient if and only if the target location is part of these visualized segments. Other situations though might require more details. Navigating individual clips using the small playback window to explore a particular video is possible, but often cumbersome and not flexible enough due to the enforced sequential navigation structure. Fig. 2 shows a screenshot of our tablet-based tool for visual inspection that is optimized for quick, efficient visual skimming of large amounts of video at detailed levels. The whole database is visualized as a storyboard representing all videos. The blue bar in Fig. 2 illustrates the end of a single video file. For each video, a key frame is extracted every second and shown as a small thumbnail. Thumbnail sizes are kept extremely small to make maximum usage of the limited screen estate, yet being large enough to ensure a good human classification performance with respect to visual perception [3]. In order to better identify scenes in such a miniaturized visualization, thumbnails are not arranged in the common line-by-line representation, but instead they are shown from left to right in columns of five thumbnails. Fig. 2 illustrates that this results in "clusters" of scenes that are easier and faster to spot. This visualization approach has proven to perform extremely well for known item search tasks (KIS). In the VBS 2015 competition, a similar interface was used with a storyboard containing all 100 hours of the database in a fixed but random order. Although it did not feature any content-based indexing or querying at all but enforced the user to visually skim the entire database, it achieved a remarkable 3rd place in the end [4].

2.3 Combined Collaborative Approach

The results from [4] have proven that an optimized visualization method performs extremely well even at database sizes of up to 100 hours. Yet, it does not scale indefinitely and there are obvious limits to how much data can be visually inspected in a given amount of time (e.g., at VBS 2015, the time limit per task was 5 minutes). Thus, we present an integrated system that combines both approaches in a way that copes with the respective disadvantages and maximizes the benefits of each individual search concept (Fig. 3). In particular, results from the CBVR tool are sent to the tablet and used to update the order of videos in the storyboard. Likewise, files that have been inspected on the tablet and classified as irrelevant are excluded from future retrieval request thus reducing the database and increasing the likelihood of finding relevant results quickly. In our research [8], we investigated different ways to resort the visualizations in the storyboard on the tablet based on the CBVR tool results as well as

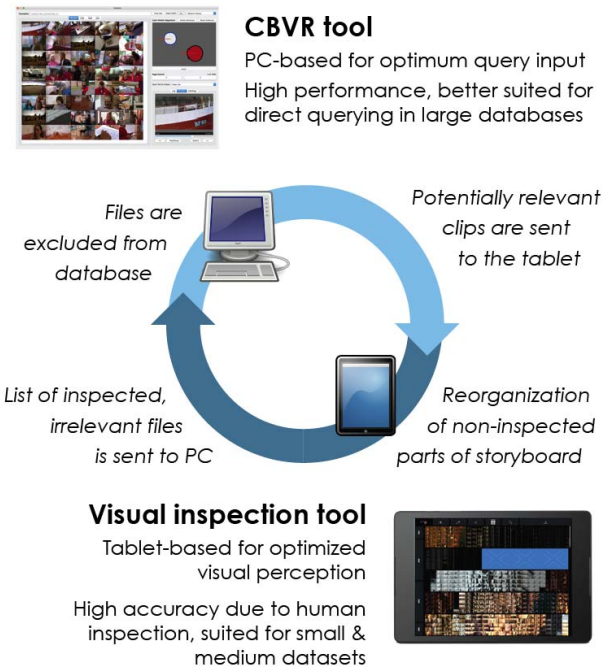how the information sent via the back channel from the tablet to the PC tool can be used in optimal ways.

Figure 3. Collaboration between CBVR & tablet tool.

3. REFERENCES

[1] Christian Beeks, Steffen Kirchhoff, Thomas Seidl. 2013. Signature matching distance for content-based image retrieval. In *Proceedings of the 3rd ACM International conference on multimedia retrieval* (ICMR'13), 41-48.

[2] Marco A. Hudelist, Claudiu Cobârzan, Christian Beecks, Rob van de Werken, Sabrina Kletz, Wolfgang Hürst, Klaus Schoeffmann. 2016. Collaborative Video Search Combining Video Retrieval with Human-based Visual Inspection. *MultiMedia Modeling*, Vol 9517 of Lecture Notes in Computer Science, 400-405.

[3] Wolfgang Hürst, and Rob van de Werken. 2015. Human-based Video Browsing – Investigating interface design for fast video browsing. In *Proceedings of IEEE International Symposium on Multimedia* (ISM 2015), 363-368.

[4] Wolfgang Hürst, Cees G.M. Snoek, Willem-Jan Spoel, and Mate Tomin. 2010. Keep moving!: revisiting thumbnails for mobile video retrieval. In *Proceedings of the 18th ACM international conference on Multimedia* (MM '10). ACM, New York, NY, USA, 963-966.

[5] Jakub Lokoč, Adam Blažek, and Tomáš Skopal. 2014. Signature-Based Video Browser. *MultiMedia Modeling*, Vol 8326 of Lecture Notes in Computer Science, 415-418.

[6] Klaus Schoeffmann. 2014. A User-Centric Media Retrieval Competition: The Video Browser Showdown 2012-2014. *IEEE Multimedia*, Vol. 21, No. 4, Oct.-Dec., 2014, 8-13.

[7] Arnold W. M. Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta, and Ramesh Jain. 2000. Content-Based Image Retrieval at the End of the Early Years. *IEEE Trans. Pattern Anal. Mach. Intell.* 22, 12 (December 2000), 1349-1380.