Interactive Video Search

Klaus Schoeffmann
Klagenfurt University
Universitätsstr. 65-67,
A-9020 Klagenfurt, Austria
ks@itec.aau.at

Frank Hopfgartner
University of Glasgow
11 University Gardens, Glasgow G12
8QH, UK
frank.hopfgartner@glasgow.ac.uk

ABSTRACT

With an increasing amount of video data in our daily life, the need for content-based search in videos increases as well. Though a lot of research has been spent on video retrieval tools and methods which allow for automatic search in videos through content-based queries, still the performance of automatic video retrieval is far from optimal. In this tutorial we discussed (i) proposed solutions for improved video content navigation, (ii) typical interaction of content-based querying features, and (iii) advanced video content visualization methods. Moreover, we discussed interactive video search systems and ways to evaluate their performance.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Experimentation

Keywords

Interactive Video Retrieval; Search; Tutorial

1. INTRODUCTION

Over the last two decades, there has been a lot of research on content-based video retrieval to solve the task of finding the proper scene of interest in a large video archive. However, there are still well-known limitations with video retrieval applications. First, there is the *usability gap*. A user is often not able to express his/her needs and thoughts through text, a challenge that is already apparent for an image ("*An image is worth a thousand words.*") but much worse for a video segment consisting of many images. Additionally, the issue of *polysemy* is a challenging problem in text-based search, which can only be partially solved through relevance feedback. Similarly, an automatic retrieval tool cannot easily determine the user-dependent relative importance (i.e., weight) of a query term and hence

often returns too many irrelevant results. Finally, there is another challenging problem that affects the content-based indexing phase, namely the *semantic gap*, which is the discrepancy between the semantics a user can derive from an image and the information a computer can extract from pixel values [15]. The semantic gap seriously limits the achievable performance of visual information retrieval tools.

There is consensus in the multimedia community that content-based retrieval approaches should have a stronger focus on the human behind the retrieval application [4, 16]. Instead of making small improvements in the field of content-based indexing and retrieval, video search tools should aim at more intense integration of the user into the search process, focusing on interactive video search [2, 3, 7, 8] rather than automatic querying. Interactive video search follows this idea of intense user integration with advanced content interaction and provides a powerful alternative to the common video retrieval approach [12]. It is known as the process of interactive video content exploration with the help of content navigation [9], summarization [1], on-demand querying [14], and browsing of querying results or filtered content [5]. In contrast to typical video retrieval, such interactive video search tools give more control to the user and provide flexible search and browsing features instead of focusing on the *query-and-browse-results* approach. Hence, even if the performance of content analysis is not optimal, there is a chance that a smart user could compensate shortcomings through clever use of available features.

Interactive video search tools support *directed* and *undirected* search scenarios. In the first scenario, users have a clear information need and want to find a specific target segment in the video (e.g., the weather forecast in a news show); such a search is also known as *known-item search* or *target search*. In the second scenario, users have no concrete search goal but want to explore the content in order to learn or find something interesting (e.g., a violent scene in surveillance videos); such a search scenario is known as *exploratory search*. Users need exploratory search when they cannot sufficiently describe the desired content, e.g., through a query in a video retrieval tool, or when they just want to browse instead of search (i.e., without a concrete desired scene).

Over the years many tools for interactive video search (often known as video browsing or video exploration tools) have been proposed in the literature (see [12, 13] for a survey), and it has been shown that these tools can effectively help users find desired content in videos (e.g., [1, 14]). Some of these tools combine sophisticated content analysis methods controlled by the user for their personal needs [10]. Some

others provide rather simple content navigation features but give the users more interactivity to allow them to effectively take advantage of their knowledge about the content and the content structure [11]. Interestingly, it has been shown that tools of the latter kind can even outperform tools of the first kind for some search tasks [10].

2. LEARNING OBJECTIVE

This tutorial gave a broad overview of solutions in two areas: (1) video retrieval, and (2) interactive video search. We discussed state-of-the-art approaches, but also the current shortcomings and outlined important research tasks for the future of multimedia search and retrieval. Our discussion also covered tools and methods that originate from the Human-Computer-Interaction (HCI) community, which many people in the multimedia community might not be aware of yet. The tutorial provided the basis to elevate research on video retrieval to the next level.

Target audience were researchers and practitioners in the field of multimedia analysis and information access. The participants learned best practice on how to interact with multimedia content and should be able to apply this knowledge in their own project upon completion of the tutorial.

Following an introduction of the subject matter, we highlighted the need for interactive video search and outlined well-known issues and current research areas. Then, we presented different graphical user interfaces that are designed to address these challenges. We started by presenting commercial and academic video search systems that represent the state-of-the-art in video search. In order to illustrate the limitations of video search, we began with an introduction to video content analysis. In particular, we focused on the segmentation of videos into different units of retrieval (i.e., video shots and semantically coherent scenes), argued for the selection of appropriate keyframes and explained how to index these video materials. Then, we overviewed methods for video content presentation, namely abstraction, summarization, and visualization of video content. After that, we introduced different methods to interact with video content. We showcased the advantages of browsing and exploration of video content, methods to navigate through the content, querying and sketching interfaces. Following this session, the participants learned how to evaluate the performance of video search engines. We introduced the de-facto standard evaluation protocol that is applied for scientific performance evaluation. Furthermore, we introduced popular Academic evaluation campaigns, namely the Known-Item search task promoted by TRECVid and the Video Browser Showdown. As an outlook in upcoming research challenges, we introduced a special case of multimedia search, namely access to visual lifelogs [6]. With the ever increasing amount of wearable devices that are released nowadays, people have the opportunity to capture vast amount of heterogeneous data (e.g., images, sensor readings, videos) that depict aspects of their personal lives. In the last part of the tutorial, we introduced approaches to access and display these "lifelogs".

Acknowledgments

3. REFERENCES

[1] B. Adams, S. Greenhill, and S. Venkatesh. Towards a video browser for the digital native. In *ICMEW'12*, pages 127–132, July 2012.

[2] M. Christel, C. Huang, N. Moraveji, and N. Papernick. Exploiting multiple modalities for interactive video retrieval. In *ICASSP'04*, volume 3, pages iii–1032, May 2004.

[3] M. G. Christel and R. Yan. Merging storyboard strategies and automatic retrieval for improving interactive video search. In *CIVR'07*, pages 486–493, New York, NY, USA, 2007. ACM.

[4] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 40(2):5:1–5:60, May 2008.

[5] A. Girgensohn, F. Shipman, and L. Wilcox. Adaptive clustering and interactive visualizations to support the selection of video clips. In *ICMR '11*, pages 34:1–34:8, New York, NY, USA, 2011. ACM.

[6] C. Gurrin, A. F. Smeaton, and A. R. Doherty. Lifelogging: Personal big data. *Foundations and Trends in Information Retrieval*, 8(1):1–125, 2014.

[7] F. Hopfgartner. *Understanding Video Retrieval*. VDM Verlag, 2007.

[8] F. Hopfgartner, J. Urban, R. Villa, and J. M. Jose. Simulated testing of an adaptive multimedia information retrieval system. In *CBMI'07*, pages 328–335, 2007.

[9] J. Huber, J. Steimle, and M. Mühlhäuser. Toward more efficient user interfaces for mobile video browsing: An in-depth exploration of the design space. In *MM'10*, pages 341–350, New York, NY, USA, 2010.

[10] K. Schoeffmann, D. Ahlström, W. Bailer, C. Cobârzan, F. Hopfgartner, K. McGuinness, C. Gurrin, C. Frisson, D.-D. Le, M. Del Fabro, H. Bai, and W. Weiss. The video browser showdown: a live evaluation of interactive video search tools. *IJMIR*, 3(2):113–127, 2014.

[11] K. Schoeffmann and L. Boeszoermenyi. Video browsing using interactive navigation summaries. In *CBMI'09*, pages 243–248, Chania, Crete, June 2009.

[12] K. Schoeffmann, F. Hopfgartner, O. Marques, L. Boeszoermenyi, and J. M. Jose. Video browsing interfaces and applications: a review. *SPIE Reviews*, 1(1):1–35(018004), 2010.

[13] K. Schoeffmann, J. Huber, and M. A. Hudelist. Video interaction tools: A survey of recent work. *ACM Computing Surveys*, pages 1–36, 2015. Accepted for publication.

[14] K. Schoeffmann, M. Taschwer, and L. Boeszoermenyi. The video explorer: A tool for navigation and searching within a single video based on fast content analysis. In *MMSys'10*, pages 247–258. ACM, 2010.

[15] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *TPAMI*, 22(12):1349–1380, Dec 2000.

[16] M. Worring, P. Sajda, S. Santini, D. A. Shamma, A. F. Smeaton, and Q. Yang. Where is the user in multimedia retrieval? *IEEE MultiMedia*, 19(4):6–10, 2012.