# Interactive Search in Video & Lifelogging Repositories

Frank Hopfgartner
University of Glasgow
Glasgow, UK
frank.hopfgartner@glasgow.ac.uk

Klaus Schoeffmann
Klagenfurt University
Klagenfurt, Austria
ks@itec.aau.at

## ABSTRACT

Due to increasing possibilities to create digital video, we are facing the emergence of large video archives that are made accessible either online or offline. Though a lot of research has been spent on video retrieval tools and methods, which allow for automatic search in videos, still the performance of automatic video retrieval is far from optimal.

At the same time, the organization of personal data is receiving increasing research attention due to the challenges that are faced in gathering, enriching, searching and visualizing this data. Given the increasing quantities of personal data being gathered by individuals, the concept of a heterogeneous personal digital libraries of rich multimedia and sensory content for every individual is becoming a reality.

Despite the differences between video archives and personal lifelogging libraries, we are facing very similar challenges when accessing these multimedia repositories. For example, users will struggle to find the information they are looking for in either collection if they are not able to formulate their search needs through a query. In this tutorial we discussed (i) proposed solutions for improved video & lifelog content navigation, (ii) typical interaction of content-based querying features, and (iii) advanced content visualization methods. Moreover, we discussed and demonstrate interactive video & lifelog search systems and ways to evaluate their performance.

## 1. MOTIVATION

This tutorial addressed two separate, yet similar challenges, namely interactive search in video and lifelogging repositories. In Section 1.1, we first introduce current challenges in the video retrieval domain. Section 1.2 then introduces open research challenges when dealing with lifelogging material.

### 1.1 Interactive Video Search

Over the last two decades, there has been a lot of research on content-based video retrieval to solve the problem of finding the proper scene of interest in a large video archive (e.g., [1, 4, 5, 6, 11, 16, 17, 20, 22, 25, 15]). Video retrieval tools use content-based indexing methods to perform automatic annotation with content descriptors for color, texture, shape, and semantic concepts. The extracted information is used to provide retrieval functions through different querying modes [13]: *query-by-text*, *query-by-example image* (or example clip), *query-by-sketch*, and by *query-by-filtering* (e.g. through semantic concepts). The user first formulates a query, initiates the retrieval engine that looks for matching video segments, and then browses the results, which are typically lists of keyframes extracted from the result segments.

Retrieval approaches commonly use interactive *relevance feedback* methods to keep the human in the loop after the query and to learn which results are relevant in order to adapt the search in another iteration to the user's needs ([3, 12, 24]). However, the actual search process with video retrieval tools is mostly automatic and a *black box* for the user. Interaction is limited to the *query-formulation* and *results-browsing* phases. Research on video retrieval mainly focuses on improving the performance of the retrieval engine (i.e., the querying itself), with less focus on the user and her interaction with the system [26].

There are well-known issues with video retrieval applications. First, there is the *usability gap*. A user is often not able to express her needs and thoughts through text, a problem 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. In fact, most video retrieval tools operate on images (i.e., keyframes of shots) and provide image query features. 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. When considering the query-by-example approach, it turns out that users rarely have a good example image or example clip at hand, which is similar to the target scene they are looking for. The query-by-sketch approach is also not very convenient for users, because most users typically cannot draw a sufficiently good sketch for a scene they want to find, although they would immediately recognize it when they see it. Query-by-filtering (e.g., for a semantic concept like '*car*', for example) in a large video collection is often also not very helpful for a user, since the returned result list is typically way too long and the confidence of concept detection is still not good enough.



*CHIIR '17 March 07-11, 2017, Oslo, Norway*



ACM ISBN 978-1-4503-4677-1/17/03.

DOI: http://dx.doi.org/10.1145/3020165.3022161

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 [23]. The semantic gap seriously limits the achievable performance of visual information retrieval tools.

## 1.2 Lifelogging

Recent technological advances have introduced new types of sensors (informational sensors, physical sensors) and devices (for example Google Glass or Apple's Watch) which allow the individual to compile vast archives of personal data.

The activity of recording personal bio signals and metrics using software and tools is referred to as self-tracking. Different reasons can be found that motivate people to start recording and analyzing personal data. A large share of self-trackers include people with chronic medical conditions who rely on technology to track their personal well-being. Other self-trackers are motivated to observe and to increase their personal fitness and wellness levels [18]. These individuals can be considered as part of the *Quantified Self* (QS) movement that uses instruments to record numerical data on all aspects of our lives: input (food consumed, surrounding air-quality), states (mood, arousal, blood oxygen levels), and performance (mental, physical). Such self-monitoring and self-sensing, which combines wearable sensors and wearable computing, is also sometimes referred to as *Lifelogging*, although lifelogging describes the process of recording and storing any type of personal data rather than just user-centric sensor data. Captured over a long period of time, such personal data can provide a detailed picture of the activities of an individual and will require search, summarization and knowledge extraction tools to make them valuable. Therefore it comes as no surprise that Lifelong Digital Libraries are receiving increasing attention within the research community [7].

One of the early adopters of lifelogging techniques is Gordon Bell of Microsoft. Within the "MyLifeBits" project [2], they captured all of their personal data in digital form and created software that allowed them to access this data. The goal of this technology was to create a personal archive, or a "portable, infallible, artificial memory" that can be exploited to increase job productivity, serve as a basis for medical treatment, improving performances in school and many other scenarios. Another lifelogging pioneer is Steve Mann, who uses wearable computing to create a record of his life. He performed significant studies on visual memory prosthetics that transformed his eye into a camera and his body into a Web server. He refers to this technique as 'cyborg logging' or 'glogging'. Wearable cameras play an important role in research on lifelogging since analyzing camera data streams can reveal a lot of information. One wearable lifelogging device, the SenseCam [14], automatically captures images every 30 seconds, resulting in thousands of recorded images per day. It is a camera with a fish-eye lens, about the size of a cigarette packet. The camera is usually worn around the neck and can be set to take photographs when triggered by such things as changes in the light, ambient heat or body temperature. The SenseCam has since been improved with new models and competing devices available for purchase. The SenseCam has received a lot of attention from scientific researchers who focused in 100+ papers on various aspects of lifelogging. For a detailed description of state-of-the-art lifelogging techniques, see Gurrin et al. [10].

# 2. OBJECTIVES

As mentioned above, various issues and challenges arise when interacting with multimedia content. In this tutorial, we gave an overview of existing video & lifelog search interfaces that illustrate different methods to approach named challenges and issues. The participants learned best practice on how to interact with multimedia content and how to apply this knowledge in their own project. Following an introduction of the subject matter, we highlighted the need for interactive video search and outlined well-known issues and current research areas. Moreover, we introduce the research area of lifelogging and highlight similarities between both fields.

Focusing on video search, 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 by providing an introduction into 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 and summarization of video content and methods for video content visualization. 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.

Focusing on the lifelogging domain, we first provided an introduction to this upcoming technology. After presenting approaches on how to capture heterogeneous datastreams of lifelogs, we outlined state-of-the-art techniques to analyze this data. Then, we presented different methodologies to visualize and access lifelog data.

Following this session, the participants learned how to evaluate the performance of video & lifelog 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 [21], the Video Browser Showdown [19] which has been organized as part of the Multimedia Modeling Conference, and the Personal Lifelog Access & Retrieval Task NTCIR-Lifelog [9] that is organized as part of the Japanese conference series on the Evaluation of Information Access Systems (NTCIR). The aim of this task is to begin the comparative evaluation of information access and retrieval systems operating over personal lifelog data. As part of this, a dataset consisting of anonymized lifeless gathered by a number of individuals over an extended period is released [8]. In 2017, NTCIR-Lifelog includes four subtasks, two of which focus on unser-centred interaction and retrieval: In the Lifelog Semantic Access Task (1), participants have to retrieve a number of specific moments in a lifelogger's life. The Lifelog Insight Task (2) aims to gain insights into a lifelogger's life as promoted by the Quantified Self community.

## 3. REFERENCES


[1] Y.A. Aslandogan and C.T. Yu. Techniques and systems for image and video retrieval. *Knowledge and Data Engineering, IEEE Transactions on*, 11(1):56–63, Jan 1999.

[2] Gordon Bell and J. Gemmell. *Your Life, Uploaded: The Digital Way to Better Memory, Health, and Productivity*. Plume, 2010.

[3] C. Breiteneder and H. Eidenberger. Content-based image retrieval in digital libraries. In *Digital Libraries: Research and Practice, 2000 Kyoto, International Conference on.*, pages 288–295, 2000.

[4] Shih-Fu Chang, John R. Smith, Mandis Beigi, and Ana Benitez. Visual information retrieval from large distributed online repositories. *Communications of the ACM*, 40(12):63–71, December 1997.

[5] Tat-Seng Chua and Li-Qun Ruan. A video retrieval and sequencing system. *ACM Transactions on Information Systems*, 13(4):373–407, October 1995.

[6] P Geetha and Vasumathi Narayanan. A survey of content-based video retrieval. *Journal of Computer Science*, 4(6):474, 2008.

[7] Cathal Gurrin and Frank Hopfgartner. Lifelong digital libraries. In *Proceedings of the 15th ACM/IEEE-CS Joint Conference on Digital Libraries*, JCDL '15, pages 287–287. ACM, 2015.

[8] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, and Rami Albatal. NTCIR lifelog: The first test collection for lifelog research. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, SIGIR 2016, Pisa, Italy, July 17-21, 2016*, pages 705–708, 2016.

[9] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, and Rami Albatal. Overview of ntcir-12 lifelog task, 2016.

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

[11] Alexander G. Hauptmann and Michael G. Christel. Successful approaches in the TREC video retrieval evaluations. In *Proceedings of ACM Multimedia*, pages 668–675, New York, NY, USA, 2004. ACM.

[12] Alexander G. Hauptmann, Wei-Hao Lin, Rong Yan, Jun Yang, and Ming-Yu Chen. Extreme video retrieval: Joint maximization of human and computer performance. In *Proceedings of ACM Multimedia*, pages 385–394. ACM, 2006.

[13] Daniel Heesch, Peter Howarth, Joao Magalhaes, Alexander May, Marcus Pickering, Alexei Yavlinsky, and Stefan Rüger. Video retrieval using search and browsing. In *TREC Video Retrieval Evaluation Online Proceedings*, 2004.

[14] Steve Hodges, Lyndsay Williams, Emma Berry, Shahram Izadi, James Srinivasan, Alex Butler, Gavin Smyth, Narinder Kapur, and Ken Wood. Sensecam: A retrospective memory aid. In *Proceedings of the 8th International Conference on Ubiquitous Computing*, UbiComp'06, pages 177–193, Berlin, Heidelberg, 2006. Springer-Verlag.

[15] Frank Hopfgartner, Jana Urban, Robert Villa, and Joemon M. Jose. Simulated testing of an adaptive multimedia information retrieval system. In *International Workshop on Content-Based Multimedia Indexing, CBMI '07, Bordeaux, France, June 25-27, 2007*, pages 328–335, 2007.

[16] Yu-Gang Jiang, Chong-Wah Ngo, and Jun Yang. Towards optimal bag-of-features for object categorization and semantic video retrieval. In *Proceedings of the 6th ACM International Conference on Image and Video Retrieval*, CIVR '07, pages 494–501, New York, NY, USA, 2007. ACM.

[17] Michael S. Lew, Nicu Sebe, Chabane Djeraba, and Ramesh Jain. Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2(1):1–19, February 2006.

[18] Na Li and Frank Hopfgartner. To log or not to log? swot analysis of self-tracking. In Stefan Selke, editor, *Lifelogging: Theoretical Approaches and Case Studies About Self-Tracking*. Springer Verlag, Zurish, June 2016.

[19] Klaus Schoeffmann, David Ahlström, Werner Bailer, Claudiu Cobârzan, Frank Hopfgartner, Kevin McGuinness, Cathal Gurrin, Christian Frisson, Duy-Dinh Le, Manfred del Fabro, Hongliang Bai, and Wolfgang Weiss. The video browser showdown: a live evaluation of interactive video search tools. *IJMIR*, 3(2):113–127, 2014.

[20] Nicu Sebe, MichaelS. Lew, Xiang Zhou, ThomasS. Huang, and ErwinM. Bakker. The state of the art in image and video retrieval. In ErwinM. Bakker, MichaelS. Lew, ThomasS. Huang, Nicu Sebe, and XiangSean Zhou, editors, *Image and Video Retrieval*, volume 2728 of *Lecture Notes in Computer Science*, pages 1–8. Springer Berlin Heidelberg, 2003.

[21] Alan F. Smeaton, Paul Over, and Wessel Kraaij. Evaluation campaigns and trecvid. In *MIR '06: Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*, pages 321–330, New York, NY, USA, 2006. ACM Press.

[22] Alan F. Smeaton, Peter Wilkins, Marcel Worring, Ork de Rooij, Tat-Seng Chua, and Huanbo Luan. Content-based video retrieval: Three example systems from trecvid. *International Journal of Imaging Systems and Technology*, 18(2-3):195–201, 2008.

[23] A. W M Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(12):1349–1380, Dec 2000.

[24] Cees Snoek, Kvd Sande, OD Rooij, Bouke Huurnink, J Uijlings, van M Liempt, M Bugalhoy, I Trancosoy, F Yan, M Tahir, et al. The mediamill trecvid 2009 semantic video search engine. In *TRECVID*, 2009.

[25] Cees G. M. Snoek and Marcel Worring. Concept-based video retrieval. *Foundations and Trends in Information Retrieval*, 2(4):215–322, April 2009.

[26] Marcel Worring, Paul Sajda, Simone Santini, David A. Shamma, Alan F. Smeaton, and Qiang Yang. Where is the user in multimedia retrieval? *MultiMedia, IEEE*, 19(4):6–10, Oct 2012.