# Interactive Video Search: Where is the User in the Age of Deep Learning?

Klaus Schoeffmann, Werner Bailer, Cathal Gurrin, George Awad, Jakub Lokoč
ks@itec.aau.at,werner.bailer@joanneum.at,cathal.gurrin@dcu.ie,george.awad@nist.gov,lokoc@ksi.mff.cuni.cz

## ABSTRACT

In this tutorial we discuss interactive video search tools and methods, review their need in the age of deep learning, and explore video and multimedia search challenges and their role as evaluation benchmarks in the field of multimedia information retrieval. We cover three different campaigns (TRECVID, Video Browser Showdown, and the Lifelog Search Challenge), discuss their goals and rules, and present their achieved findings over the last half-decade. Moreover, we talk about datasets, tasks, evaluation procedures, and examples of interactive video search tools, as well as how they evolved over the years. Participants of this tutorial will be able to gain collective insights from all three challenges and use them for focusing their research efforts on outstanding problems that still remain unsolved in this area.

**ACM Reference Format:**
Klaus Schoeffmann, Werner Bailer, Cathal Gurrin, George Awad, Jakub Lokoč. 2018. Interactive Video Search:, Where is the User in the Age of Deep Learning?. In *Proceedings of 2018 ACM Multimedia Conference (MM '18).* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3240508.3241473

## 1 MOTIVATION FOR THE TOPIC

With an increasing amount of video content in our daily life, the need for information retrieval in videos increases as well. Although a lot of research has been done on automatic video retrieval tools (following a *query-and-browse results* approach, where the search engine acts more like a black-box for the user), we need more interactive video search tools that provide a *strong integration of the user into the search process* [7, 9, 15]. For example, sometimes users are not able to formulate their search needs through a text-query; also, often they simply want to browse content without any concrete query in mind (and switch to querying later). Such search scenarios are enabled by interactive video search tools, which provide a rich set of flexible content retrieval features, combined with effective video interaction means. They give full control of the search process to the users, who know best what feature (or a combination thereof) to use when and how, in order to solve a specific search problem. Even if there is no "best feature" for a search task, the user could switch back and forth between different interactions and approach the problem in a way of trial-and-error.

However, in order to know the achievable performance of a video search tool, we also need appropriate evaluation methods. It is only possible to assess the performance of specific tools or methods when comparing them to other tools/methods. While there exist several choices for such evaluations – e.g., user studies, user simulations, and evaluation campaigns/search challenges – we argue that the latter are particularly well suited for evaluation of the performance of multimedia retrieval tools. The reason for that is obvious: such competitions make sure that all teams/tools use exactly the same dataset, the same tasks, the same environment (room setting etc.), and the same evaluation metrics, so that the results become truly comparable. Over the last years, several evaluation campaigns and benchmarking initiatives in the field of multimedia have been carried out (e.g., TRECVID[1], VBS[4], LSC - the Lifelog Search Challenge[5], MediaEval[8], etc.), which clearly demonstrate the importance of the topic for the community.

In this tutorial we discuss the need for interactive video search tools and present (and demonstrate) state-of-the-art implementations. We cover details of the above-mentioned search challenges and report on the learned lessons, insights, and conclusions, as well as open challenges that remain to be solved. The evaluation campaigns for interactive video search tools fill a gap between the evaluation of automatic retrieval systems and user studies. With emerging trends such as multimedia analytics [3], the need for interactive search and exploration of multimedia content arises in a range of disciplines, which could benefit from the lessons learned in these evaluation campaigns. In addition, some of the methodology may also be applicable to other problems in multimedia, that share some aspects with interactive search.

## 2 COURSE DESCRIPTION

*Introduction.* In the introduction we review the need for interactive video search, using examples from different application areas. These use cases have in common that the information need is fuzzy or cannot be easily formulated in terms of the metadata or features available, and thus benefit from the user in the loop. We raise the question about the role of the human user in the age of deep learning [15]. The quality of automatic annotations is rapidly increasing, among others thanks to the use of deep neural networks. While such annotations are highly important to improve content search (both automatic and interactive), they cannot always replace the user, who may not even precisely know what to search for in the first place. We discuss requirements for interactive video search tools and discuss the possibilities and limitations of the state-of-the-art tools [9, 14]. The question how to compare the performance of such tools is discussed, and how systems that involve a user in the loop can be assessed in a repeatable manner, while keeping the efforts manageable. This brings in the aspect of evaluation campaigns, and why they should address interactive tools.

*Interactive video search tools.* In this part we discuss the structure and properties of modern interactive video search tools, and show demonstrations of such tools. The demos include the VIRET tool


*MM '18, October 22–26, 2018, Seoul, Republic of Korea*

ACM ISBN 978-1-4503-5665-7/18/10.
https://doi.org/10.1145/3240508.3241473

[10, 11] from Charles University in Prague (winner at VBS 2018), the ITEC tool [12, 13] from Klagenfurt University (2nd place at VBS 2018 and LSC 2018), and the most-recent lifelogging search tool (2018) from Dublin City University [6]. Moreover, in this part we also discuss existing open source software and libraries that can be helpful for researchers starting to work in this area. rep

*Evaluation approaches.* Different approaches of evaluating interactive search and browsing systems such as user studies/surveys, analysis of logs, question answering and indirect/task-based evaluation (e.g. following the Cranfield paradigm as in VBS, LSC, and TRECVID) are discussed, using data comparing the approaches (e.g. [2]). The choice of evaluation approaches is not only influenced by aspects such as repeatability and the reuse of assessments, but also impacted by the setting of the evaluation campaign, i.e., whether the competition is live in front of the audience (as e.g. in VBS) or an offline process (as e.g. in TRECVID). The history of selected evaluation campaigns is briefly described, and examples of tasks from TRECVID, VBS and LSC are reviewed, in order to illustrate specific evaluation goals and task settings. The review of these examples leads to a more structured analysis of task design and data preparation in the next part.

*Task design.* Finally, task design also addresses the question *who* performs the task. While it is usually the developers of the teams who participate in evaluation campaigns, "novice" sessions, in which members of the audience use the tools, provide valuable insights into the complexity and usability of the tools. In many application areas tools are likely to be used by domain experts rather than retrieval experts, thus this condition models real situations. However, it also raises questions of comparability that need to be addressed in the evaluation.

*Datasets.* We then cover the selection and preparation of large-scale datasets, including cleaning the data (e.g., handling broken data and duplicates), consideration of legal issues, and ways to generate ground truth. The aspects of dataset generation are put in relation to task types, e.g., concerning the effort for creating ground truth, covering the complete dataset and the reusability of annotations in other settings. Examples from TRECVID, LSC, VBS, MediaEval and MPEG CDVA are used to compare properties of tasks and the practicalities of defining the task setting and illustrate the challenges of dataset generation and possible ways to handle them. It includes a discussion of crowdsourcing data and the related issues, as well as on-demand (live) annotation of submissions as used in VBS. There are different ways to set up the evaluation procedure, in terms of secrecy of test data and sharing work between participants and organizers, in terms of using pooled results or preparing all ground truth, etc. This part provides an in-depth discussion of the choice and design of evaluation metrics. Examples from TRECVID, VBS, and LSC are discussed, including the comparison of different metrics on the same data from actual submissions. The topic of comparability of metrics from tasks with slightly different settings are also covered.

*Lessons learned from evaluation campaigns.* This part focuses on achievements and observations made from running evaluation campaigns over the years. For example, the VBS – which has been organized for seven times in an annual manner since 2012 – experienced many changes and trends in how users search and how they adapt the systems to the type of tasks. We discuss findings of what worked and did not work in terms of task design and metrics, provide examples of loopholes found by the participants, and how they have shaped the rules of the competition. Moreover, we compare results from automatic retrieval tasks (e.g., TRECVID) to the ones from interactive search tasks (e.g., VBS) for the same dataset. We analyze the influence of (i) the user interface and (ii) concept detection, which has gained increasing performance over the years. We demonstrate how different teams influence each other and mutually push each other to the limits. Moreover, we discuss the issue of managing large datasets, which have been growing with the capabilities of tools over the years, as well as other existing problems and open challenges.

*Conclusions and outlook.* We revisit the question of the role of the user in the age of deep learning. Having shown may settings that require a human in the loop, we discuss where the progress in automatic content analysis can benefit interactive search, and complement the capabilities of the user. Based on that we conclude our tutorial and present an outlook of future challenges for interactive multimedia search and the evaluation of tools.

## ACKNOWLEDGMENTS

This work has been partially supported by Czech Science Foundation (GAČR) project no. 17-22224S, and has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no 761802, MARCONI.

## REFERENCES


[1] George Awad, Asad Butt, Jonathan Fiscus, Martial Michel, David Joy, Wessel Kraaij, Alan F. Smeaton, Georges Quénot, Maria Eskevich, Roeland Ordelman, Gareth J. F. Jones, and Benoit Huet. 2017. TRECVID 2017: Evaluating Ad-hoc and Instance Video Search, Events Detection, Video Captioning and Hyperlinking. In *Proceedings of TRECVID 2017*. NIST, USA.

[2] Werner Bailer and Herwig Rehatschek. 2009. Comparing Fact Finding Tasks and User Survey for Evaluating a Video Browsing Tool. In *Proceedings of ACM Multimedia*. Beijing, CN.

[3] Nancy A Chinchor, James J Thomas, Pak Chung Wong, Michael G Christel, and William Ribarsky. 2010. Multimedia analysis+ visual analytics= multimedia analytics. *IEEE computer graphics and applications* 30, 5 (2010), 52–60.

[4] 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.

[5] Duc-Tien Dang-Nguyen, Klaus Schoeffmann, and Wolfgang Hurst. 2018. LSE2018 Panel - Challenges of Lifelog Search and Access. In *Proceedings of the 2018 ACM Workshop on The Lifelog Search Challenge (LSC '18)*. ACM, New York, NY, USA, 1–2. DOI: http://dx.doi.org/10.1145/3210539.3210540

[6] Aaron Duane, Cathal Gurrin, and Wolfgang Huerst. 2018. Virtual Reality Lifelog Explorer: Lifelog Search Challenge at ACM ICMR 2018. In *Proceedings of the 2018 ACM Workshop on The Lifelog Search Challenge (LSC '18)*. ACM, New York, NY, USA, 20–23. DOI: http://dx.doi.org/10.1145/3210539.3210544

[7] Christoph Kofler, Martha Larson, and Alan Hanjalic. 2016. User intent in multimedia search: a survey of the state of the art and future challenges. *ACM Computing Surveys (CSUR)* 49, 2 (2016), 36.

[8] Martha Larson, Mohammad Soleymani, Guillaume Gravier, Bogdan Ionescu, and Gareth JF Jones. 2017. The benchmarking initiative for multimedia evaluation: MediaEval 2016. *IEEE MultiMedia* 24, 1 (2017), 93–96.

[9] J. Lokoč, W. Bailer, K. Schoeffmann, B. Muenzer, and G. Awad. 2018. On influential trends in interactive video retrieval: Video Browser Showdown 2015-2017. *IEEE Transactions on Multimedia* (2018), 1–1. DOI: http://dx.doi.org/10.1109/TMM.2018.2830110

[10] Jakub Lokoč, Gregor Kovalčík, and Tomáš Souček. 2018. Revisiting SIRET Video Retrieval Tool. In *MultiMedia Modeling - 24th International Conference, MMM 2018, Bangkok, Thailand, February 5-7, 2018, Proceedings, Part II*. 419–424.

[11] Jakub Lokoč, Tomáš Souček, and Gregor Kovalčík. 2018. Using an Interactive Video Retrieval Tool for LifeLog Data. In *Proceedings of the 2018 ACM Workshop on The Lifelog Search Challenge, LSC@ICMR 2018, Yokohama, Japan, June 11, 2018*. 15–19.

[12] Bernd Münzer, Andreas Leibetseder, Sabrina Kletz, Manfred Jürgen Primus, and Klaus Schoeffmann. 2018. lifeXplore at the Lifelog Search Challenge 2018. In *Proceedings of the 2018 ACM Workshop on The Lifelog Search Challenge (LSC '18)*. ACM, New York, NY, USA, 3–8. DOI: http://dx.doi.org/10.1145/3210539.3210541

[13] Manfred Jürgen Primus, Bernd Münzer, Andreas Leibetseder, and Klaus Schoeffmann. 2018. The ITEC Collaborative Video Search System at the Video Browser Showdown 2018. In *International Conference on Multimedia Modeling*. Springer, 438–443.

[14] Klaus Schoeffmann, Marco A. Hudelist, and Jochen Huber. 2015. Video Interaction Tools: A Survey of Recent Work. *ACM Comput. Surv.* 48, 1, Article 14 (Sept. 2015), 34 pages.

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