Habilitationsschrift

Interactive Video Search

eingereicht an der

Alpen-Adria-Universität Klagenfurt
Fakultät für Technische Wissenschaften

von

Dipl.-Ing. Dr. Klaus Schöffmann

Klagenfurt, am 20.8.2014

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 video retrieval is far from optimal. More importantly, there are situations where common video retrieval cannot be employed, for example when users are not able to formulate their search needs through a query – or when they simply want to browse the content without any concrete query in mind. Interactive video search tools and methods provide an alternative way of content-based search in videos. Instead of the *query-and-browse-results* approach, they provide many flexible content interaction features and give full control of the search process to the user, who will know best which features to use and how, in order to solve a search problem. This habilitation thesis summarizes my research contributions spent in the four key areas of interactive video search: (i) video presentation, (ii) video interaction, (iii) evaluation of interactive video search tools, and (iv) video processing.

Contents

1 Introduction **3**

1.1 The Need for Content Search in Video 3

1.2 Video Retrieval 4

1.3 Shortcomings of Current Video Retrieval Tools 4

1.4 Interactive Video Search as an Alternative 5

1.5 Key Areas of Interactive Video Search 7

1.6 Organization of this Habilitation Thesis 8

1.6.1 Video Presentation 9

1.6.2 Video Interaction on Mobile Devices 11

1.6.3 Performance Evaluation of Interactive Video Search . 15

1.6.4 Video Processing in Medical Endoscopy 18

1.7 Summary and Outlook 22

1.8 Acknowledgement 23

1.9 Bibliography 24

1 Introduction

This document contains a selection of my publications constituting my habilitation thesis. It provides a summary of my research on interactive video search, which includes the following key areas: (i) video presentation, (ii) video interaction, (iii) evaluation of interactive video search, and (iv) video processing to support video search.

The document is organized as follows. It starts with a motivation why interactive video search is an interesting and challenging topic and has not been addressed sufficiently in previous research. Next, I summarize the four different parts of this habilitation thesis and explain how they are related to interactive video search, before I describe each part in detail, as well as the corresponding contributions.

1.1 The Need for Content Search in Video

When users need to search for content in popular videos in the public domain, the task of finding a specific video or video segment is usually straightforward. They simply have to go to a sharing platform on the web, for example to a video portal like YouTube[1], and perform a textual query. This will typically already bring up results including the desired content, because popular videos in the public domain often contain textual annotations. These meta-data are either produced by the content creators themselves or by users of the sharing platform and are usually sufficient for a text search to produce satisfactory results.

However, the situation changes dramatically when users want to search in private or unpopular video content that contains no or only minor metadata. For example, when parents want to find a specific scene of their child, recorded several years ago and stored in a large private video collection without any annotations. Here the search task becomes highly challenging and time-consuming and often results in a trial-and-error approach, where video files are opened one after another and are tediously checked, typically with a video player by using playback and seeker-bar navigation. The situation becomes even worse if the videos are very long and are added to the collection on a daily basis, like in the medical domain where surgeons nowadays save recordings of their endoscopic interventions in a long-term storage archive [34, 35, 37] (see also Section 1.6.4).

Tools for content-based search in video are needed to support such search scenarios. There are basically two extremes for content-based video search. First, automatic search with a "smart" system that indexes the content and asks the user what she wants (through a query). Secondly, manual search with a "dumb" video player that allows the user to inspect the content,

[1] http://www.youtube.com

e.g., by playback. The first way would be an optimal solution but unfortunately has several issues (see Section 1.3). Also the second way is not convenient in practice, since it obviously cannot efficiently cope with large video archives. Interactive search tools combine methods from both sides to provide a powerful and interactive solution with enough flexibility to allow for "smart usage" by a user for the needs of her search task.

1.2 Video Retrieval

Video retrieval tools and methods [6, 52, 58] were originally invented to support content search situations as described above. 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., [3, 7, 10, 17, 20, 29, 31, 48, 50, 54]). Video retrieval tools use content-based indexing methods (i.e., content analysis) 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 [22]: *query-by-text*, *query-by-example image* (or example clip), *query-by-sketch*, and by *query-by-filtering*. 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 ([5, 21, 53]). However, the actual search process with video retrieval tools is mostly automatic. 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 [57].

1.3 Shortcomings of Current Video Retrieval Tools

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.*", cf. Figure 1) 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*[2] 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

[2] Definition of 'polysemy' in the Oxford Dictionary: The coexistence of many possible meanings for a word or phrase.

Figure 1: Example pictures that are hard to describe in detail by text, but easy to recognize and retrieve interactively (once known by the user).

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 instead of query-by-text, 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 (otherwise they would probably not search for it). 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.

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

Though a lot of progress has been made in video retrieval over the years, the performance of video retrieval tools is still far from optimal, as also concluded in [19, 54]. Moreover, most of them are too much focused on the *query-and-browse-results* approach and do not allow the user to circumvent shortcomings of the video retrieval system.

1.4 Interactive Video Search as an Alternative

Due to the problems mentioned above, users are often not able to perform content-based video search with an automatic video retrieval tool. It is not only a problem that there are situations when users simply cannot formulate a good query for a desired scene but would easily find it by exploratory search ("*I will know it when I see it.*"). More importantly, there are situations where users have no concrete query in mind but rather want to browse through videos in order to explore the content; a use scenario that is not handled by video retrieval applications. There is consensus in the multimedia community that content-based retrieval approaches should have a stronger focus on the human behind the retrieval application [11, 57]. 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 retrieval [8, 9, 16, 23] rather than automatic querying.

Interactive video search follows this idea of intense user integration with sophisticated content interaction and provides a powerful alternative to the common video retrieval approach [2, 14, 32, 33, 45, 56]. It is known as the process of interactive video content exploration with the help of content navigation [25], summarization [1], on-demand querying [46], and browsing of querying results or filtered content [18]. 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 video content in order to learn or find something interesting (e.g., a violent scene in recordings of surveillance cameras); such a search scenario is known as *exploratory search*. Users need exploratory search when they cannot sufficiently describe the content they want to find, 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 [45] for a review), and it has been shown that these tools can effectively help users find desired content in videos [1, 46]. Some of these tools combine sophisticated content analysis methods controlled by the user for her personal needs [41]. Some others provide rather simple content navigation features but give the user more interactivity to allow her to effectively take advantage of her knowledge about the content and the content structure [12, 13, 27, 43]. Interestingly, it has been shown that tools of the latter kind can even outperform tools of the first kind for some search tasks [4].

1.5 Key Areas of Interactive Video Search

Interactive video search comprises four key areas that cover several research topics in the multimedia and human-computer-interaction communities.

As shown in Figure 2, the first area is video *presentation*, which investigates different presentation methods for video content. It covers various ways of content visualization, including video surrogates, video summarization and abstraction. It focuses on content presentation in general, i.e., without user interaction. The second key area is about *interaction* with video content and focuses on interaction methods/models for content browsing, navigation, as well as query and manipulation features. It somehow overlaps with the first area but addresses the interactive use of presented visual content (so, for example, interactive video summarization and skimming would also belong to this area). The third key area is *evaluation* of interactive video search tools, where different evaluation methods (e.g., user studies, user simulations, and evaluation campaigns) are employed to assess the performance of search tools. Finally, the last key area of interactive video search is video *processing*. It focuses on processing of video content for various purposes: analysis and indexing of the video content, segmentation into meaningful units, filtering of the content, but also image and video coding. It is an important part of interactive video search, since the other areas, e.g., (1) and (2), often build on the results of content processing.

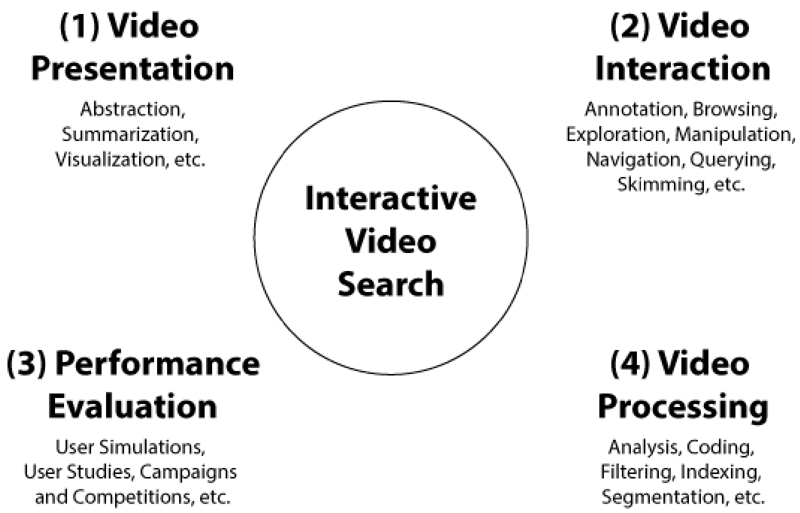

Figure 2: The key areas of Interactive Video Search.

1.6 Organization of this Habilitation Thesis

My habilitation thesis focuses on several problems of the four key areas of interactive video search, as defined in Figure 2.

The first problem, which I have addressed in various publications, is the issue of how to present content of visual data (i.e., videos and images) such that it allows for better performance at visual search. In particular, several novel 3D interfaces that allow for faster visual search in image collections than a common 2D grid interface were designed and evaluated. The proposed interfaces can also be used for video search [42], since videos and video segments are typically represented by images/keyframes.

Another problem addressed in this habilitation thesis is how to take advantage of the available interaction features of mobile multimedia devices (e.g., multi-touch input) for the purpose of video search. Therefore, the second part focuses on video interaction on mobile multimedia devices, such as tablets and smartphones. Video browsers with novel interfaces and special interaction models, tailored to the requirements of such devices, are presented and evaluated.

The third part focuses on the evaluation of interactive video search tools itself. More precisely, it addresses the evaluation of video browsing tools that participated in the Video Browser Showdown[3], which is an annual evaluation campaign for interactive video search tools, co-founded in 2012 by myself (together with Werner Bailer from Joanneum Research). The main goal of the Video Browser Showdown [47] is to push research on interactive video search and to allow for a comparative evaluation of the performance of search tools on a shared dataset.

Finally, the last part of my habilitation thesis addresses video processing in the special field of medical endoscopy. In this challenging domain I focus on segmentation of endoscopic video as well as content filtering and keyframe extraction. The results of these methods can be used as input for interactive video search tools.

[3]http://www.videobrowsershowdown.org/

1.6.1 Video Presentation

Interfaces designed for content-based search in video need to visualize the content to the user. This is true for highly interactive search tools but also for video retrieval tools, which need to present the results of an automatic search query. Most interfaces proposed in the literature, however, use a grid interface, i.e., a matrix-like arrangement of thumbnails of video segments, for that purpose. This interface for browsing images and videos was invented decades ago, when the performance and graphic capabilities of desktop PCs were much lower. Therefore, in a series of experiments I have developed alternative interfaces, using 3D graphics and a color-sorted arrangement of images, which should allow for faster visual search by the user. Together with colleagues I performed several user studies to evaluate the achievable performance of these interfaces. We concentrated on the practical scenario of known-item search [36, 49], which is a target search use case where someone knows of an item and knows that it is contained in the data set, but does not know where it is located.

As the first in the literature, we could show that 3D interfaces allow for significantly faster search in a list of thumbnails, when directly compared to the default 2D grid interface. Additionally, while other works in the literature (e.g., [38, 39, 40]) already indicated that color-based arrangements of images have advantages for image browsing, to best of our knowledge we were the first to provide empirical evidence that it can also significantly increase visual search performance for known-item search tasks.

Moreover, we have also investigated if thumbnails of images (or keyframes of video segments) need to be presented with the correct aspect ratio. This is an important question for the design of search interfaces for visual data, which often have some constraints on the content presentation and arrangement (e.g., vertical and horizontal vanishing lines). The work described above is given in Section **??** and contains a selection of the following papers:

VP-1.1 Klaus Schoeffmann, David Ahlström, and Laszlo Böszörmenyi, **A User Study of Visual Search Performance with Interactive 2D and 3D Storyboards**, *Adaptive Multimedia Retrieval. Large-Scale Multimedia Retrieval and Evaluation*, Marcin Detyniecki, Ana García-Serrano, Andreas Nürnberger, Sebastian Stober (Eds.), LNCS 7836, Springer, Berlin Heidelberg, 2013, pp. 18-32.

This is our first paper to present evaluation results of a 3D interface for browsing image/video thumbnails in direct comparison to a common, scrollable 2D grid interface as well as to a static grid with very small thumbnails. The 3D interface uses an “inside view” to the background of a vertical ring of thumbnails and no color sorting (as later improved versions of the interface do). The results showed no significant improvement on search time over the 2D interfaces but a better subjective ranking of the 3D interface (this paper was published in post-proceedings of selected and revised papers presented at the *9th International Workshop on Adaptive Multimedia Retrieval (AMR 2011)* in Barcelona, Spain).

Own contribution: concept/idea: 80%, implementation: 100%, study design: 50%, execution: 100%, evaluation: 40%, manuscript: 70% (est.)

VP-1.2 Klaus Schoeffmann, David Ahlström, and Laszlo Böszörmenyi, **3D Storyboards for Interactive Visual Search**, in *Proceedings of the 2012 IEEE International Conference on Multimedia and Expo (ICME)*, Melbourne, Australia, 2012, pp. 848-853.

In this paper we have designed and evaluated different versions of an improved interface that has been developed according to the results of our first paper (VP-1.1, see above). In particular, we evaluated a smooth ring of thumbnails in outer (i) vertical view, (ii) horizontal view, and (iii) frontal view. All interfaces use a color sorted arrangement of thumbnails for better visual orientation. Since no appropriate algorithm for general color sorting of image thumbnails could be found at that time in the literature, we developed an own algorithm for that purpose (VP-1.3, see below). The results of the study presented in this paper have shown that all three versions perform similarly in terms of achievable search time but the horizontal arrangement performed best in terms of subjective ratings; therefore, it has been used for all further studies.

Own contribution: concept/idea: 80%, implementation: 100%, study design: 50%, execution: 50%, evaluation: 50%, manuscript: 50% (est.)

VP-1.3 Klaus Schoeffmann and David Ahlström, **An Evaluation of Color Sorting for Image Browsing**, in *International Journal of Multimedia Data Engineering and Management*, Vol.3, No.1, IGI Publishing, Hershey, PA 17033, USA, pp. 49-62, 2012.

This paper describes and evaluates a color sorting algorithm for image thumbnails that was used for all our further studies on the visual search performance of 3D interfaces (VP-1.2, VP-1.5). The algorithm produces an intuitive arrangement of thumbnails based on their dominant color, such that a user should be able to quickly narrow down known-item search tasks by color.

Own contribution: concept/idea: 100%, implementation: 100%, study design: 60%, execution: 80%, evaluation: 50%, manuscript: 70% (est.)

VP-1.4 David Ahlström and Klaus Schoeffmann, **A Visual Search User Study on the Influences of Aspect Ratio Distortion of Preview Thumbnails**, in *Proceedings of the 2012 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, pp. 546-551, 2012.

Many interfaces for image and video search are designed with a predefined size (e.g., width) of thumbnails in order to follow visual design principles to increase usability, such as vertical and horizontal vanishing lines. However, many images and videos use different aspect ratios (e.g., 4:3, 16:9, 16:10, etc.), therefore image and video browsing applications cannot follow the design principle of vanishing lines in a diverse data set. Hence, we performed a user study to find out whether distorted visualization of thumbnails (i.e., visualization of all diverse thumbnails with the same aspect ratio, e.g., 4:3) affects visual search performance. Our results show that such a distorted visualization of image thumbnails, from an original aspect ratio of 16:9 down to 4:3, does not affect visual search in neither terms of search time nor in terms of errors.

Own contribution: concept/idea: 60%, implementation: 80%, study design: 50%, execution: 50%, evaluation: 40%, manuscript: 40% (est.)

VP-1.5 Klaus Schoeffmann and David Ahlström, **Using a 3D Cylindrical Interface for Image Browsing to Improve Visual Search Performance**, in *Proceedings of the 13th IEEE International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, pp. 1-4, 2012.

This is the first paper where we could provide evidence that our proposed 3D ring interface can significantly outperform a common 2D grid interface in terms of search time for known-item search tasks.

Own contribution: concept/idea: 80%, implementation: 100%, study design: 50%, execution: 50%, evaluation: 50%, manuscript: 80% (est.)

1.6.2 Video Interaction on Mobile Devices

The findings from the work on video presentation (Section 1.6.1) were further used for a number of investigations of image and video browsing interfaces on mobile multimedia devices. The main goal in this part was to design and evaluate interaction models and interfaces that take advantage of the possibilities of these new devices, which are equipped with various sensors like multi-touch input, gyroscope, accelerometer, etc. Most of these investigations were performed in the context of the FWF Translational Research Project (TRP 273-N15) on *Next-Generation Video Browsing* (NGVB), also supported by Lakeside Labs Klagenfurt. Several prototypes were implemented for demonstrating different interaction models as well as for the evaluation with various user studies (three demo papers showing a few of these prototypes are included at the end of this section). The described work is given in Section **??** of this thesis, which contains a selection of the following papers:

VI-2.1 Klaus Schoeffmann, David Ahlström, and Marco Andrea Hudelist, **3D Interfaces to Improve the Performance of Visual Known-Item Search**, *IEEE Transactions on Multimedia (TMM)*, 2014, pp. 1-10, *to appear.*

This journal paper summarizes the most important findings on the performance of 3D vs. 2D interfaces for the purpose of visual known-item search on tablets (VI-2.4), smartphones [26], and desktop PCs (VP-1.5). It presents a cross-study evaluation, draws conclusions from all our investigations and outlines avenues for further work.

Own contribution: concept/idea: 80%, implementation: n/a, study design: n/a, execution: n/a, evaluation: 20%, manuscript: 65% (est.)

VI-2.2 Klaus Schoeffmann, **The Stack-of-Rings Interface for Large-Scale Image Browsing on Mobile Touch Devices**, in *Proceedings of the 22nd ACM International Conference on Multimedia (MM 2014)*, Orlando, Florida, 2014, pp. 1-4, *to appear.*

This paper presents results from another study on 3D interfaces for large-scale image browsing on tablet computers. It evaluates an extension of the 3D ring interface (VP-1.5) for browsing an unlimited number of images. The findings of the user study show that this "*Stack-of-Rings*" interface can significantly outperform the common grid interface in both achievable search time and subjective ratings.

Acknowledgement: Many thanks to Johannes Zlattinger, who implemented and evaluated this interface in the course of his master thesis (under my supervision).

Own contribution: concept/idea: 100%, implementation: 0%, study design: 70%, execution: 0%, evaluation: 50%, manuscript: 100% (est.)

VI-2.3 Klaus Schoeffmann, Kevin Chromik, and Laszlo Böszörmenyi, **Video Navigation on Tablets with Multi-Touch Gestures**, in *Proceedings of the IEEE 2014 Conference on Multimedia and Expo Workshops (ICMEW)*, Chengdu, China, 2014, pp. 1-6.

In this paper we propose to use wipe gestures for video navigation on tablet computers, because this would make video interaction consistent to interaction with other media on tablets. The navigation concept presented in the paper combines the previously proposed ideas of

Wolfgang Hürst et al. [28] and Huber et al. [24]. The results from a user study showed that the vast majority of users prefer this kind of navigation interaction in videos over the commonly used seeker-bar navigation.

Acknowledgement: Many thanks to Kevin Chromik, who implemented and evaluated this interface in the course of an internship (under my supervision).

Own contribution: concept/idea: 100%, implementation: 0%, study design: 70%, execution: 0%, evaluation: 100%, manuscript: 80% (est.)

VI-2.4 David Ahlström, Marco Andrea Hudelist, Klaus Schoeffmann, and Gerald Schaefer, **A User Study on Image Browsing on Touchscreens**, in *Proceedings of the 20th ACM International Conference on Multimedia (MM 2012)*, Nara, Japan, 2012, pp. 925-928.

In this paper, which originated from the master thesis of Marco Andrea Hudelist, we evaluate an adapted version of a 3D globe interface for image browsing on tablet computers. The work was inspired by the paper of Gerald Schaefer [40], from Loughborough University, U.K., who originally proposed to use a globe interface with color-based arrangement of thumbnails for large-scale image browsing on the desktop PC. Therefore, we performed our study together with him. The findings of the study show that the adapted globe interface can significantly outperform a common grid interface on a tablet computer in terms of both search time and subjective ratings.

Own contribution: concept/idea: 70%, implementation: 20%, study design: 40%, execution: 0%, evaluation: 10%, manuscript: 20% (est.)

VI-2.5 Marco Andrea Hudelist, Klaus Schoeffmann, and David Ahlström, **Evaluating Alternatives to the 2D Grid Interface for Mobile Image Browsing**, *International Journal of Semantic Computing*, World Scientific, Vol. 8, No. 2, 2014, pp. 25, *to appear*.

In the first part of this journal paper we compare the 3D globe interface (VI-2.4) to an adapted version of the 3D ring interface (VP-1.5) on tablet computers. Results from a user study show that both interfaces perform equally well on tablets. In the second part we present results from user simulations[4] and another user study with smartphones, where we performed a comparative evaluation of the 3D globe,

[4]In this context, user simulations evaluate interfaces by comparing interaction effort for specific given search situations (e.g., the number of required drag/pinch/tap gestures to bring a specific image of the list onto the screen). Although user simulations can never reflect the real user behavior – and also cannot show unexpected behavior like real user studies – they provide theoretical clues about the usefulness of an interface under specific assumptions. In this user simulation we counted the number of required gesture interactions for the considered interfaces in order to bring particular areas of an image list

the 3D ring, the 2D grid, and a zoomable 2D grid interface (called *ImagePane* in the paper). The results of the user simulation show that the ImagePane requires least interaction steps for visual search tasks, followed by the ring, globe, and grid, with most interaction steps. The user study, which was performed with 100, 200, 300, and 400 images, finally showed that the 3D interfaces as well as the ImagePane could outperform the Grid for 300 and in particular for 400 images in terms of search time. However, when evaluating results over all data set sizes, no significant difference in terms of search time could be found. The questionnaires of the user study showed that users perceived the ImagePane significantly worse than the grid in terms of interaction effort and the globe as significantly less fun to use than the grid (there were no other significant differences).

Own contribution: concept/idea: 80%, implementation: 30%, study design: 60%, execution: 0%, evaluation: 20%, manuscript: 20% (est.)

VI-2.6 Klaus Schoeffmann, Marco Andrea Hudelist, Manfred Del Fabro, and Gerald Schaefer, **Demo: Mobile Image Browsing on a 3D Globe**, in *Proceedings of the ACM International Conference on Multimedia Retrieval (ICMR)*, Hong Kong, China, 2012, pp. 61:1-61:2.

In this demo paper we present different versions of a mobile image browser that uses the metaphor of a globe for visualizing color-sorted images. The paper was written together with Gerald Schaefer from Loughborough University, who performed several related studies on this topic.

Own contribution: concept/idea: 80%, implementation: 0%, study design: n/a, execution: n/a, evaluation: n/a, manuscript: 50% (est.)

VI-2.7 Marco Andrea Hudelist, Klaus Schoeffmann, and Laszlo Böszörmenyi, **Demo: Mobile Video Browsing with a 3D Filmstrip**, in *Proceedings of the ACM International Conference on Multimedia Retrieval (ICMR)*, Dallas, Texas, USA, 2013, pp. 68-69.

In this demo paper we present the concept of video browsing through the metaphor of a filmstrip. A video is represented through an interactive 3D filmstrip that can be moved (i.e., scrolled), scaled, and tilted. Also, a user can start playback for several segments in parallel by simple tap gestures. We plan to evaluate this filmstrip through a user study in future work.

Own contribution: concept/idea: 80%, implementation: 0%, study design: n/a, execution: n/a, evaluation: n/a, manuscript: 30% (est.)

to the screen (it is assumed that a user will recognize/find an image once it is visible on the screen, although this might not always be the case in practice). A discussion about user simulations is also given in Section 1.6.3.

VI-2.8 Marco Andrea Hudelist, Klaus Schoeffmann, and Laszlo Böszörmenyi, **Demo: Mobile Video Browsing with the ThumbBrowser**, in *Proceedings of the 21st ACM International Conference on Multimedia*, Barcelona, Spain, 2013, pp. 405-406.

This is another demo paper that demonstrates our ideas for more convenient use of video players on tablet computers. The presented video player assumes that a user holds her tablet computer with two hands and is therefore designed for thumb-based interaction. We plan to evaluate this video player through a user study in future work.

This paper was shortlisted for the best technical demo award at ACM MM 2013.

Own contribution: concept/idea: 50%, implementation: 0%, study design: n/a, execution: n/a, evaluation: n/a, manuscript: 30% (est.)

1.6.3 Performance Evaluation of Interactive Video Search

Video retrieval tools and methods with automatic content search features usually can be evaluated automatically, for a predefined data and task set. Unfortunately, interactive video search tools are not so easy to evaluate since they strongly build on the user and her interaction. Basically, there are three ways to evaluate such interactive video search tools:

- User studies
- User simulations
- Evaluation campaigns

Evaluation through user studies has several advantages. First, it reflects the real user behavior and allows to get specific feedback. Next, it can show unexpected behavior and reveal usability issues of the user interface. Moreover, it allows to measure several aspects of an interface, for example achievable performance and mental load. The downside, however, is the tedious and time-consuming execution of a user study. Also, if there is any problem in the setup (or a software issue in the interface to test), all data collected so far need to be discarded and the user study needs to be repeated.

User simulations on the other hand are much quicker to perform. However, simulating usage of an interactive interface is a non-trivial task and often based on assumptions that might not hold in practice. Moreover, user simulations cannot provide subjective feedback, which would help to improve an interface, nor can they reveal non-obvious issues of an interface or show unexpected behavior. Although user simulations are often a helpful tool to assess specific objective aspects of an application, they are only an approximation of real usage.

There is a third category in the evaluation methods of interactive search tools, namely evaluation campaigns. One example of such an evaluation campaign is TRECVID [49], which is an annual evaluation competition for video retrieval tools that has been running for more than 12 years already. Another annual evaluation campaign for multimedia retrieval is MediaEval, originally known as VideoCLEF [30]. Although both TRECVID and MediaEval contained interactive tasks for some years (e.g., the *Known-Item Search* task at TRECVID [36] from 2010 through 2012, or the still active *Seach & Hyperlinking* task at MediaEval [15]), their main focus is automatic retrieval.

A more interactive evaluation campaign was the *VideOlympics Showcase* [55], which ran from 2007 through 2009 in conjunction with the *ACM International Conference on Image and Video Retrieval (CIVR)*, nowadays known as *International Conference on Multimedia Retrieval (ICMR)*. While VideOlympics had less focus on effectiveness of retrieval results but more focus on the visualization of video content as well as on the demo of the actual retrieval tools, the content-based search was still based on querying (i.e., on the query-and-browse-results approach).

Inspired by these evaluation campaigns, especially by VideOlympics, together with Werner Bailer from Joanneum Research we founded an own campaign in 2012, namely the *Video Browser Showdown* (VBS)[5]. In the first year, also Cees Snoek from the University of Amsterdam in the Netherlands, who is the organizer of the VideOlympics, joined our team. Different from VideOlympics, the VBS focuses entirely on interactive search in video and, for example, does not allow textual queries. The VBS is an annual evaluation competition for video browsing tools, which runs in conjunction with the *International Conference on Multimedia Modeling (MMM)*. It evaluates the performance of interactive video search tools for visual known-item search tasks as well as textual description-based search tasks in single videos and video collections. In addition to the actual event on-site, a post-hoc evaluation of the achieved search performance is conducted and published as a paper in a collaborative effort with the participants of the VBS [4, 41]. The 4th edition of the VBS will be held together with MMM2015 under the new name "*Video Search Showcase*" in Sydney, Australia, in January 2015.

The papers in Section **??** contain publications that originated from the Video Browser Showdown, or were performed as related research.

[5]http://www.videobrowsershowdown.org/

EVA-3.1 Klaus Schoeffmann, **A User-Centric Media Retrieval Competition: The Video Browser Showdown 2012-2014**, *IEEE Multimedia*, Oct.-Dec. 2014, pp. 5, *to appear.*

This article describes the Video Browser Showdown (VBS) event and summarizes the first three years. It explains the rationale behind this annual evaluation competition and gives details on the rules, data sets, tasks, and the teams that participated over the years. Based on the results of the first three years, it also underlines the importance of the user for interactive video search tools and gives some outlook on future events of the VBS.

EVA-3.2 Klaus Schoeffmann, David Ahlström, Werner Bailer, Claudiu Cobarzan, Frank Hopfgartner, Kevin McGuinness, 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**, *International Journal of Multimedia Information Retrieval*, Vol. 3, No. 2, Springer, 2014, pp. 113-127.

This journal paper presents an evaluation of the second Video Browser Showdown competition (VBS 2013). It gives details on the tools of the teams that participated in VBS 2013 and evaluates the achieved performance. Moreover, the performance is also compared to the one achievable with a simple video player, which was investigated through an additional "baseline study" [44].

As the paper contains descriptions of the video browsing tools that were used for VBS 2013, it was written in collaboration with participants of the VBS 2013. The majority of the paper, however, was authored by me.

Own contribution: concept/idea: 60%, implementation: 0%, study design: 60%, execution: 0%, evaluation: 60%, manuscript: 55% (est.)

EVA-3.3 Klaus Schoeffmann and Claudiu Cobarzan, **An Evaluation of Interactive Search with Modern Video Players**, in *Proceedings of the 2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW 2013)*, San Jose, CA, USA, 2013, pp. 1-4.

In this paper we present the evaluation results of a "baseline study"[6] for known-item search in single videos, which was performed to assess the achieved performance of the professional video browsing tools competing in the VBS (see above, EVA-3.2).

Own contribution: concept/idea: 100%, implementation: 0%, study design: 70%, execution: 0%, evaluation: 60%, manuscript: 40% (est.)

[6]More precisely, we evaluate the performance that can be achieved with a basic – i.e., very simple – tool for video browsing, namely a common video player.

EVA-3.4 Claudiu Cobarzan and Klaus Schoeffmann, **How Do Users Search with Basic HTML5 Video Players?**, in *Proceedings of the 20th International Conference on Multimedia Modeling (MMM 2014)*, Dublin, Ireland, 2014, pp. 109-120.

In the paper above (EVA-3.3) we evaluated how users search in video with simple video players for visual known-item tasks (i.e., when the user knows the video content). The main rationale was that we wanted to compare the "baseline performance" (achievable with a simple video player) to the professional tools used in the Video Browser Showdown. However, the third Video Browser Showdown (VBS 2014) did also perform description-based search tasks in video, in addition to visual known-item search tasks that were performed in the first two years. Therefore, we conducted another study with description-based search tasks to be performed with simple video players. This study and the findings are described in this paper.

This paper received the "*Best Poster Paper*" award at the 20th International Conference on Multimedia Modeling (MMM 2014), in Dublin, Ireland, on January 7th, 2014.

Own contribution: concept/idea: 60%, implementation: 0%, study design: 60%, execution: 0%, evaluation: 50%, manuscript: 20% (est.)

EVA-3.5 Werner Bailer, Klaus Schoeffmann, David Ahlström, Wolfgang Weiss, and Manfred Del Fabro, **Interactive Evaluation of Video Browsing Tools**, in *Proceedings of the 19th International Conference on Multimedia Modeling (MMM) 2013*, Huangshan, China, pp. 81-91.

This paper summarizes the first Video Browser Showdown (VBS 2012) competition and evaluates the different tools of the participating teams in terms of performance.

Own contribution: concept/idea: 50%, implementation: 0%, study design: 50%, execution: 0%, evaluation: 20%, manuscript: 20% (est.)

1.6.4 Video Processing in Medical Endoscopy

The final research item of my habilitation thesis is video processing. Interactive video search tools use video processing methods for several purposes. First, it is used for segmentation of the video content into basic segments/units. For most video domains these segments are known as *video shots*, which are continuous recordings of a single camera. These basic units are often used as input for further processing algorithms, such as keyframe extraction, where a representative frame is extracted from each segment and used to present the content of a video. However, video processing is also used to summarize video content [1], to filter specific content, to enable content-based similarity search, and to support interaction with video content, e.g., through enhanced seeker-bars with abstract content visualization [43].

The publications in this part focus on the special field of endoscopic video processing. This is processing of video content recorded during medical endoscopic interventions, also known as minimally invasive surgery (or keyhole surgery). Over the last few years more and more surgeons switched over to record entire endoscopic interventions and save them in a long-term video storage archive, since it is an important source of information for retrospective analyses and for training of young surgeons. However, special tools are needed to enable interactive search and retrieval in such an archive of endoscopic videos.

Endoscopic video processing is a very challenging field due to the special characteristics of the video content. For example, there are no shot boundaries in the video content, since the recorded footage is usually unedited. Therefore, the papers listed below address video segmentation through different ways of content analysis (e.g., based on motion as well as local features). Moreover, there are many segments that provide no semantic information, for example because the content is too blurry or because it contains content recorded outside of the patient. These content segments need to be filtered out. Finally, since an endoscopic video archive is quickly growing, special coding and storage techniques need to be used to keep the data manageable. The listed publications in Section **??** originated from two research projects, which I am co-leading:

- *CODE-MM: Community of Domain Experts in Medical Multimedia.* This is a Lakeside Labs research project, partially funded by KARL STORZ GmbH & Co KG[7], Tuttlingen, Germany, with two PhD students (one of them, Manfred Jürgen Primus, under my supervision) and one post-doc researcher.

- *EndoViP: Endoscopic Video Processing.* This is a contract research project entirely funded by KARL STORZ GmbH & Co KG, with one PhD student (Bernd Münzer) under my supervision.

EVP-4.1 Klaus Schoeffmann, Manfred Del Fabro, Tibor Szkaliczki, Laszlo Böszörmenyi, and Jörg Keckstein, **Keyframe Extraction in Endoscopic Video**, *Multimedia Tools and Applications*, Journal, Springer, 2014, pp. 1-18, *to appear.*

In this journal paper we evaluate a few different methods for keyframe extraction in endoscopic video. The evaluation was performed in collaboration with the Department of Gynaecology and Obstetrics at the hospital LKH Villach in Austria. The evaluation of the presented

[7]KARL STORZ GmbH & Co KG is an international leading company in the area of endoscopes and medical instruments manufacturing.

keyframe extraction method was performed by Prof. Dr. Jörg Keckstein (co-author of the paper), who is a surgeon with international reputation in the field of endometriosis. The results show that the proposed method performs best in our evaluation (and in comparison to other proposed works in the field) but still leaves avenues for further improvement in order to better meet with the requirements of surgeons.

Own contribution: concept/idea: 80%, implementation: 60%, study design: 50%, execution: 50%, evaluation: 70%, manuscript: 60% (est.)

EVP-4.2 Bernd Münzer, Klaus Schoeffmann, and Laszlo Böszörmenyi, **Detection of Circular Content Area in Endoscopic Videos**, in *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems (CMBS 13)*, Porto, Portugal, 2013, pp. 534-536.

Many endoscopic videos show the content in a centered circle surrounded by a dark border. This dark border often varies in color and in noise, such that the compression efficiency of endoscopic videos is not optimal. Moreover, it is important to know where the actual content circle – which is not static but spatially moves over time – is located in the frames, such that video content analysis can be limited to the actual content. Therefore, in this paper we propose a reliable algorithm to detect the circular content area in endoscopic videos.

Own contribution: concept/idea: 70%, implementation: 0%, study design: 60%, execution: 0%, evaluation: 20%, manuscript: 20% (est.)

EVP-4.3 Bernd Münzer, Klaus Schoeffmann, Laszlo Böszörmenyi, **Improving Encoding Efficiency of Endoscopic Videos by Using Circle Detection Based Border Overlays**, in *Proceedings of the 2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW 2013)*, San Jose, CA, USA, 2013, pp. 1-4.

Here we evaluate the improved encoding efficiency that can be achieved by coloring the border of all frames in endoscopic video with uniform black. This removes color variance and noise in the border area and hence produces higher redundancies, which finally allows for better compression of the endoscopic video content. The algorithm proposed in EVP-4.2 (see above) is used to detect the actual border area.

Own contribution: concept/idea: 80%, implementation: 0%, study design: 60%, execution: 0%, evaluation: 20%, manuscript: 20% (est.)

EVP-4.4 Bernd Münzer, Klaus Schoeffmann, Laszlo Böszörmenyi, Johannes F. Smulders, and Jack J. Jakimowicz, **Investigation of the Impact of Compression on the Perceptional Quality of Laparoscopic Videos**, in *Proceedings of the 27th IEEE International Symposium on Computer-Based Medical Systems (CBMS 2014)*, New York, USA, 2014, pp. 1-6.

This is a very important paper, which describes a study about the necessary perceptional quality of laparoscopic videos, which is a special field in medical endoscopy. The study was performed in cooperation with the Catharina Hospital Eindhoven, as well as Delft University of Technology, from the Netherlands. The results of the study show that the "technical quality" (i.e., bitrate of the the compressed video) can be significantly reduced without a significant change on the perceived semantic quality.

Own contribution: concept/idea: 80%, implementation: 0%, study design: 50%, execution: 0%, evaluation: 20%, manuscript: 20% (est.)

EVP-4.5 Manfred Jürgen Primus, Klaus Schoeffmann, and Laszlo Böszörmenyi, **Segmentation of Recorded Endoscopic Videos by Detecting Significant Motion Changes**, in *Proceedings of the 11th IEEE International Workshop on Content-Based Multimedia Indexing (CBMI)*, Veszprem, Hungary, 2013, pp. 223-228.

In this paper we propose and evaluate a segmentation method for endoscopic videos. The algorithm is based on (i) motion tracking in video with the Kanade-Lucas-Tomasi (KLT) feature tracker, and (ii) a detection of significant changes of motion in single spatial areas of the frame.

Own contribution: concept/idea: 80%, implementation: 0%, study design: 50%, execution: 0%, evaluation: 20%, manuscript: 30% (est.)

EVP-4.6 Bernd Münzer, Klaus Schoeffmann, Laszlo Böszörmenyi, **Relevance Segmentation of Laparoscopic Videos**, in *Proceedings of the 2013 IEEE International Symposium on Multimedia (ISM 2013)*, Anaheim, CA, USA, 2013, pp. 84-91.

Here we propose and evaluate a method to automatically detect irrelevant segments in the special field of laparoscopic video. In particular, our method detects (i) scenes recorded out of the patient, (ii) too blurry segments, and (iii) segments that are completely black, which also often occur in the corpus of laparoscopic videos.

Own contribution: concept/idea: 60%, implementation: 20%, study design: 50%, execution: 0%, evaluation: 20%, manuscript: 20% (est.)

1.7 Summary and Outlook

In this habilitation thesis I have motivated the need for interactive video search. I have defined the four key areas with (i) video presentation, (ii) video interaction, (iii) evaluation, and (iv) video processing, and described how these areas interrelate with each other. I have shown that in my research I made contributions in all four areas. Many publications originated from funded research projects, which underlines the relevance and timeliness of my research. However, there are still several issues to be investigated. In particular, in the fields of video presentation (Section **??**) and interaction (Section **??**) further studies are required to fully investigate the potential of modern mobile multimedia devices. The presented work is only a first step towards this challenging and interesting domain. In our current work we are evaluating interactive video search tools designed for collaborative use, where several users work together and collaboratively search for desired content in parallel (each user with her own device). This will enable increased search performance due to shared load among different users in a distributed search scenario. This could be a valuable tool in situations where quick information retrieval is needed but video retrieval tools are not applicable (e.g., to collaboratively search in surveillance video for suspicious content in cases of attacks or disasters). The Video Browser Showdown competition (Section **??**) will be continued with its fourth occurrence in January 2015 and further iterations are planned, in order to push research on interactive video search. Also in the field of endoscopic video processing (Section **??**) several further studies are planned to provide better storage techniques, filtering methods, and similarity search in an endoscopic video archive.

1.8 Acknowledgement

I would like to give thanks to my long-lasting research colleagues, for their help and support: David Ahlström, Werner Bailer, Claudiu Cobarzan, Manfred Del Fabro, Marco Andrea Hudelist, Bernd Münzer, and Manfred Jürgen Primus. Many thanks go to my boss and continuing mentor, Prof. Laszlo Böszörmenyi, for his enduring support, his excellent mentoring, and his inspiring discussions. Last but not least, I would like to pay tribute to my family for their continuous support. To my wife, who has been there for me throughout the years and showed tremendous understanding, and to my mother, without whose sacrifice my academic career would not have been possible. I cannot thank them enough.

Several papers of this habilitation thesis were funded by the Austrian Federal Ministry for Transport, Innovation and Technology (bmvit) and Austrian Science Fund (FWF): TRP 273-N15. Several papers were also funded by the European Regional Development Fund and the Carinthian Economic Promotion Fund (KWF), under grant KWF - 20214 22573 33955, supported by Lakeside Labs GmbH, Klagenfurt, Austria.

1.9 Bibliography

References

[1] B. Adams, S. Greenhill, and S. Venkatesh. Towards a video browser for the digital native. In *Multimedia and Expo Workshops (ICMEW), 2012 IEEE International Conference on*, pages 127–132, July 2012.

[2] F. Arman, R. Depommier, A. Hsu, and M.-Y. Chiu. Content-based browsing of video sequences. In *Proceedings of the Second ACM International Conference on Multimedia*, MULTIMEDIA '94, pages 97–103, New York, NY, USA, 1994. ACM.

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

[4] W. Bailer, K. Schoeffmann, D. Ahlström, W. Weiss, and M. Del Fabro. Interactive evaluation of video browsing tools. In S. Li, A. Saddik, M. Wang, T. Mei, N. Sebe, S. Yan, R. Hong, and C. Gurrin, editors, *Advances in Multimedia Modeling*, volume 7732 of *Lecture Notes in Computer Science*, pages 81–91. Springer Berlin Heidelberg, 2013.

[5] 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.

[6] S.-F. Chang, W. Chen, H. J. Meng, H. Sundaram, and D. Zhong. Videoq: An automated content based video search system using visual cues. In *Proceedings of the Fifth ACM International Conference on Multimedia*, MULTIMEDIA '97, pages 313–324, New York, NY, USA, 1997. ACM.

[7] S.-F. Chang, J. R. Smith, M. Beigi, and A. Benitez. Visual information retrieval from large distributed online repositories. *Communications of the ACM*, 40(12):63–71, Dec. 1997.

[8] M. Christel, C. Huang, N. Moraveji, and N. Papernick. Exploiting multiple modalities for interactive video retrieval. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, volume 3, pages iii–1032, May 2004.

[9] M. G. Christel and R. Yan. Merging storyboard strategies and automatic retrieval for improving interactive video search. In *Proceedings of the 6th ACM International Conference on Image and Video Retrieval*, CIVR '07, pages 486–493, New York, NY, USA, 2007. ACM.

[10] T.-S. Chua and L.-Q. Ruan. A video retrieval and sequencing system. *ACM Transactions on Information Systems*, 13(4):373–407, Oct. 1995.

[11] 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.

[12] M. Del Fabro, B. Münzer, and L. Böszörmenyi. Smart video browsing with augmented navigation bars. In S. Li, A. Saddik, M. Wang, T. Mei, N. Sebe, S. Yan, R. Hong, and C. Gurrin, editors, *Advances in Multimedia Modeling*, volume 7733 of *Lecture Notes in Computer Science*, pages 88–98. Springer Berlin Heidelberg, 2013.

[13] M. Del Fabro, K. Schoeffmann, and L. Böszörmenyi. Instant video browsing: A tool for fast non-sequential hierarchical video browsing. In G. Leitner, M. Hitz, and A. Holzinger, editors, *HCI in Work and Learning, Life and Leisure*, volume 6389 of *Lecture Notes in Computer Science*, pages 443–446. Springer Berlin Heidelberg, 2010.

[14] S. M. Drucker, A. Glatzer, S. De Mar, and C. Wong. Smartskip: Consumer level browsing and skipping of digital video content. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '02, pages 219–226, New York, NY, USA, 2002. ACM.

[15] M. Eskevich, G. J. Jones, R. Aly, R. J. Ordelman, S. Chen, D. Nadeem, C. Guinaudeau, G. Gravier, P. Sébillot, T. de Nies, P. Debevere, R. Van de Walle, P. Galuscakova, P. Pecina, and M. Larson. Multimedia information seeking through search and hyperlinking. In *Proceedings of the 3rd ACM Conference on International Conference on Multimedia Retrieval*, ICMR '13, pages 287–294, New York, NY, USA, 2013. ACM.

[16] G. Gaughan, A. F. Smeaton, C. Gurrin, H. Lee, and K. McDonald. Design, implementation and testing of an interactive video retrieval system. In *Proceedings of the 5th ACM SIGMM International Workshop on Multimedia Information Retrieval*, MIR '03, pages 23–30, New York, NY, USA, 2003. ACM.

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

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

[19] A. Hauptmann, R. Yan, and W.-H. Lin. How many high-level concepts will fill the semantic gap in news video retrieval? In *Proceedings of*

the 6th ACM International Conference on Image and Video Retrieval, CIVR '07, pages 627–634, New York, NY, USA, 2007. ACM.

[20] A. G. Hauptmann and M. G. Christel. Successful approaches in the TREC video retrieval evaluations. In *Proceedings of the 12th Annual ACM International Conference on Multimedia*, MULTIMEDIA '04, pages 668–675, New York, NY, USA, 2004. ACM.

[21] A. G. Hauptmann, W.-H. Lin, R. Yan, J. Yang, and M.-Y. Chen. Extreme video retrieval: Joint maximization of human and computer performance. In *Proceedings of the 14th Annual ACM International Conference on Multimedia*, MULTIMEDIA '06, pages 385–394, New York, NY, USA, 2006. ACM.

[22] D. Heesch, P. Howarth, J. Magalhaes, A. May, M. Pickering, A. Yavlinsky, and S. Rüger. Video retrieval using search and browsing. In *TREC Video Retrieval Evaluation Online Proceedings*, 2004.

[23] F. Hopfgartner and J. Jose. Evaluating the implicit feedback models for adaptive video retrieval. In *Proceedings of the International Workshop on Multimedia Information Retrieval*, MIR '07, pages 323–331, New York, NY, USA, 2007. ACM.

[24] J. Huber, J. Steimle, R. Lissermann, S. Olberding, and M. Mühlhäuser. Wipe'n'watch: spatial interaction techniques for interrelated video collections on mobile devices. In *Proceedings of the 24th BCS Interaction Specialist Group Conference*, BCS '10, pages 423–427, Swinton, UK, UK, 2010. British Computer Society.

[25] 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 *Proceedings of the International Conference on Multimedia*, MM '10, pages 341–350, New York, NY, USA, 2010. ACM.

[26] M. Hudelist, K. Schoeffmann, and D. Ahlström. Evaluation of image browsing interfaces for smartphones and tablets. In *Multimedia (ISM), 2013 IEEE International Symposium on*, pages 1–8, Dec 2013.

[27] W. Hürst. Interactive audio-visual video browsing. In *Proceedings of the 14th annual ACM International Conference on Multimedia*, pages 675–678. ACM, 2006.

[28] W. Hürst, G. Götz, and M. Welte. Interactive video browsing on mobile devices. In *Proceedings of the 15th International Conference on Multimedia*, MULTIMEDIA '07, pages 247–256, New York, NY, USA, 2007. ACM.

[29] Y.-G. Jiang, C.-W. Ngo, and J. 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.

[30] M. Larson, E. Newman, and G. Jones. Overview of VideoCLEF 2009: New perspectives on speech-based multimedia content enrichment. In C. Peters, B. Caputo, J. Gonzalo, G. Jones, J. Kalpathy-Cramer, H. Müller, and T. Tsikrika, editors, *Multilingual Information Access Evaluation II. Multimedia Experiments*, volume 6242 of *Lecture Notes in Computer Science*, pages 354–368. Springer Berlin Heidelberg, 2010.

[31] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain. Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2(1):1–19, Feb. 2006.

[32] F. C. Li, A. Gupta, E. Sanocki, L.-w. He, and Y. Rui. Browsing digital video. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, volume 1, pages 169–176, 2000.

[33] K. Manske, M. Mühlhäuser, S. Vogl, and M. Goldberg. Obvi hierarchical 3d video browsing. In *Proceedings of the ACM International Conference on Multimedia*, pages 1–6, 1998.

[34] B. Münzer, K. Schoeffmann, and L. Boszormenyi. Improving encoding efficiency of endoscopic videos by using circle detection based border overlays. In *Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on*, pages 1–4, July 2013.

[35] B. Münzer, K. Schoeffmann, and L. Böszörmenyi. Relevance segmentation of laparoscopic videos. In *Proceedings of The 2013 IEEE International Symposium on Multimedia (ISM 2013)*, pages 1–8, 2013.

[36] P. Over, G. Awad, M. Michel, J. Fiscus, G. Sanders, B. Shaw, W. Kraaij, A. F. Smeaton, and G. Quénot. TRECVID 2012 – an overview of the goals, tasks, data, evaluation mechanisms and metrics. In *Proceedings of TRECVID 2012*, 2012.

[37] M. J. Primus, K. Schoeffmann, and L. Böszörmenyi. Segmentation of recorded endoscopic videos by detecting significant motion changes. In *Content-Based Multimedia Indexing (CBMI), 2013 11th International Workshop on*, pages 223–228, June 2013.

[38] K. Rodden, W. Basalaj, D. Sinclair, and K. Wood. Evaluating a visualization of image similarity as a tool for image browsing. *IEEE Symp. on Information Visual.*, page 36, 1999.

[39] K. Rodden, W. Basalaj, D. Sinclair, and K. Wood. Does organisation by similarity assist image browsing? In *Proc. of SIGCHI Conference on Human Factors in Computing Systems*, pages 190–197, 2001.

[40] G. Schaefer. A next generation browsing environment for large image repositories. *Multimedia Tools and Appl.*, 47:105–120, 2010.

[41] 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. *International Journal of Multimedia Information Retrieval (MMIR)*, 3(2):113–127, 2014.

[42] K. Schoeffmann, D. Ahlström, and L. Böszörmenyi. Video browsing with a 3d thumbnail ring arranged by color similarity. In K. Schoeffmann, B. Merialdo, A. Hauptmann, C.-W. Ngo, Y. Andreopoulos, and C. Breiteneder, editors, *Advances in Multimedia Modeling*, page 660–662. Springer Berlin Heidelberg New York, Jan 2012.

[43] K. Schoeffmann and L. Boeszoermenyi. Video browsing using interactive navigation summaries. In *Proceedings of the 7th International Workshop on Content-Based Multimedia Indexing*, pages 243–248, Chania, Crete, June 2009. IEEE.

[44] K. Schoeffmann and C. Cobârzan. An evaluation of interactive search with modern video players. In *Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on*, pages 1–4, July 2013.

[45] 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.

[46] 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 *Proceedings of the First Annual ACM SIGMM Conference on Multimedia Systems*, MMSys '10, pages 247–258, New York, NY, USA, 2010. ACM.

[47] K. Schöffmann and W. Bailer. Video browser showdown. *ACM SIGMultimedia Records*, 4(2):1–2, 2012.

[48] N. Sebe, M. Lew, X. Zhou, T. Huang, and E. Bakker. The state of the art in image and video retrieval. In E. Bakker, M. Lew, T. Huang, N. Sebe, and X. Zhou, editors, *Image and Video Retrieval*, volume 2728 of *Lecture Notes in Computer Science*, pages 1–8. Springer Berlin Heidelberg, 2003.

[49] A. F. Smeaton, P. Over, and W. Kraaij. Evaluation campaigns and trecvid. In *Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*, pages 321–330, 2006.

[50] A. F. Smeaton, P. Wilkins, M. Worring, O. de Rooij, T.-S. Chua, and H. Luan. Content-based video retrieval: Three example systems from trecvid. *International Journal of Imaging Systems and Technology*, 18(2-3):195–201, 2008.

[51] 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.

[52] S. W. Smoliar and H. Zhang. Content-based video indexing and retrieval. *IEEE MultiMedia*, 1(2):62–72, June 1994.

[53] C. Snoek, K. Sande, O. Rooij, B. Huurnink, J. Uijlings, v. M. Liempt, M. Bugalhoy, I. Trancosoy, F. Yan, M. Tahir, et al. The mediamill trecvid 2009 semantic video search engine. In *TRECVID workshop*, 2009.

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

[55] C. G. M. Snoek, M. Worring, O. de Rooij, K. E. A. Van de Sande, R. Yan, and A. Hauptmann. Videolympics: Real-time evaluation of multimedia retrieval systems. *MultiMedia, IEEE*, 15(1):86–91, Jan 2008.

[56] S. Srinivasan, D. Ponceleon, A. Amir, and D. Petkovic. "what is in that video anyway?": in search of better browsing. In *Multimedia Computing and Systems, 1999. IEEE International Conference on*, volume 1, pages 388–393, Jul 1999.

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

[58] H. J. Zhang, J. Wu, D. Zhong, and S. W. Smoliar. An integrated system for content-based video retrieval and browsing. *Pattern Recognition*, 30(4):643–658, 1997.