

Finding a needle in a haystack: an interactive video archive explorer for professional video searchers

Peer-reviewed author version

HAESEN, Mieke; MESKENS, Jan; LUYTEN, Kris; CONINX, Karin; Becker, Jan Hendrik; Tuytelaars, Tinne; Poullisse, Gert-Jan; Phi, The Pham & Moens, Marie-Francine (2013) Finding a needle in a haystack: an interactive video archive explorer for professional video searchers. In: MULTIMEDIA TOOLS AND APPLICATIONS, 63(2), p.331-356..

DOI: 10.1007/s11042-011-0809-y

Handle: <http://hdl.handle.net/1942/12151>

Finding a Needle in a Haystack: an Interactive Video Archive Explorer for Professional Video Searchers

Mieke Haesen · Jan Meskens ·
Kris Luyten · Karin Coninx · Jan
Hendrik Becker · Tinne Tuytelaars ·
Gert-Jan Poulisse · Phi The Pham ·
Marie-Francine Moens

Received: date / Accepted: date

Abstract ¹ Professional video searchers typically have to search for particular video fragments in a vast video archive that contains many hours of video data. Without having the right video archive exploration tools, this is a difficult and time consuming task that induces hours of video skimming. We propose the video archive explorer, a video exploration tool that provides visual representations of automatically detected concepts to facilitate individual and collaborative video search tasks. This video archive explorer is developed by employing a user-centred methodology, which ensures that the tool is more likely to fit to the end user needs. A qualitative evaluation with professional video searchers shows that the combination of automatic video indexing, interactive visualisations and user-centred design can result in an increased usability, user satisfaction and productivity.

Keywords Searching and Browsing Video Archives · Information Filtering · User-Centred Software Engineering · Interactive Visualisations

Mieke Haesen, Jan Meskens, Kris Luyten, Karin Coninx
Hasselt University - tUL - IBBT, Expertise Centre for Digital Media,
Wetenschapspark 2, B-3590 Diepenbeek, Belgium
E-mail: {mieke.haesen, jan.meskens, kris.luyten, karin.coninx}@uhasselt.be

Jan Hendrik Becker, Tinne Tuytelaars
ESAT/PSI, K.U.Leuven
Kasteelpark Arenberg 10 - bus 02441, B-3001 Heverlee, Belgium
E-mail: {janhendrik.becker, tinne.tuytelaars}@esat.kuleuven.be

Gert-Jan Poulisse, Phi The Pham, Marie-Francine Moens
Department of Computer Science, K.U.Leuven
Celestijnenlaan 200A, B-3001 Heverlee, Belgium
E-mail: {gert-jan.poulisse, phithe.pham, sien.moens}@cs.kuleuven.be

¹ Published by Springer, accessible on <http://www.springerlink.com/content/n7v5517q736v9484/>

1 Introduction

Today, there is a need for robust video searching tools to master the continuously growing digital video archives that are available in the TV broadcasting industry. Such archives typically contain terrabytes of video fragments of broadcasted TV programs and news bulletins. Finding the intended fragment in such an archive is the responsibility of professionally trained *video searchers*. This is often a difficult task, because the videos in these archives are mostly not well described. One query typically results in dozens of videos. Video searchers may have to skim video fragments for hours in order to find the desired segment.

On the one hand, the composition of a relevant search query will influence the search results that are shown to a searcher. On the other hand, the growing amount of available digital videos causes additional difficulties when a particular video fragment is sought. This work concentrates on the exploration of search results that are presented after entering a query. Not only the query, but also the representation of the search results may influence the search experience and efficiency.

To ease video searching, algorithms have been developed to detect features and concepts in videos [21]. These automatically detected concepts can be exploited to build interactive video archive visualisations. Some of these visualisations help users to quickly explore a large amount of videos without having to watch every video in detail. This is often accomplished by video summaries such as storyboards [4, 42, 38], video skims [8] or layered timelines [15]. Other interactive video visualisations such as DRAGON [22], slit-tear [36] and the smart-player [6] allow users to locate a small video fragment in a larger video.

Because of the novel technologies applied in video retrieval, this domain concentrates mainly on technological aspects, with only limited notice of the end users [33]. However, the efficiency of retrieving videos also depends on the user interface that presents the search results. Interactive visualisations that take into account the needs and goals of professional video searchers are most likely to increase the search experience. A good way to design an application that takes all these aspects into account, is to employ a user-centred approach. Such an approach usually starts with a user needs analysis in order to identify who needs to use the tool (e.g. professional video searchers), how the videos are explored, what tasks need to be supported, what devices can be used (e.g. desktop pc, large multi-touch display, mobile device), etc. Iterative design, development, evaluation and observation activities gradually contribute to desired video exploration tool.

This paper presents a video archive explorer for the TV broadcasting industry that combines interactive visualisations and automatic detection of video fragments. Although, we do not underestimate the technical aspects, the video archive explorer and design decisions for visualising search results will be presented from a user-centred perspective because these considerations also influence the search experience. Using a user-centred approach, we first express the needs of a video retrieval tool for professionally trained video

searchers. Based on these needs, we define a typical video search scenario. This visual scenario was used to develop our system and to evaluate its use in the professionals' day-to-day activities. The resulting prototype is running on a multi-touch table. This setting supports collaborative searching and provides a better overview of search results.

In summary, this paper contributes:

- a user-centred methodology to create better video search tools (Section 2).
- a multi-touch prototype for exploring video archives, suitable for professional video searchers (Section 3). Video exploration is facilitated using cross-media annotation algorithms such as story segmentation and naming of faces in video (Section 4).
- a qualitative evaluation, which showed that the combination of the visual representation and automatically generated annotations improves the search experience (Section 5).

2 User-Centred Process

Visualisation and interaction techniques are more accurately adapted to the target group by following a user-centred design (UCD) approach. By involving end users from the beginning of the development process, it is more likely that the visualisation of the final user interface corresponds to their needs and goals [40]. The development process that is applied, is based on a framework for user-centred software engineering [12]. Figure 1 illustrates the UCD process for the video archive explorer, including extracts of the artefacts that were created and used.

The center of Figure 1 shows all stages of the process that are carried out in one iteration. This process is started by investigating the *new and legacy system* during several stakeholder meetings and a user study. Following, a *structured interaction analysis* is carried out to model the results of the first stage as a foundation for prototyping and development purposes. The *low- and high-fidelity prototyping* stages concern the gradually evolving design of the user interface. Lastly, the *final user interface* is obtained by combining the high-fidelity prototype and application logic.

The work presented in this paper concerns one iteration, while several results will be used as input for a next iteration. The most important techniques and artefacts of the process, that contributed to the video archive explorer, concern the *Contextual Inquiry* (see Figure 1, top left), the *Visual Scenario of Use* (see Figure 1, right) and *Iterative Prototyping* (see Figure 1, bottom left). These techniques and artefacts will be described in the following.

2.1 Contextual Inquiry

The end users involved in the development process are professional video searchers. In the first stage of the process, the characteristics of the *legacy*

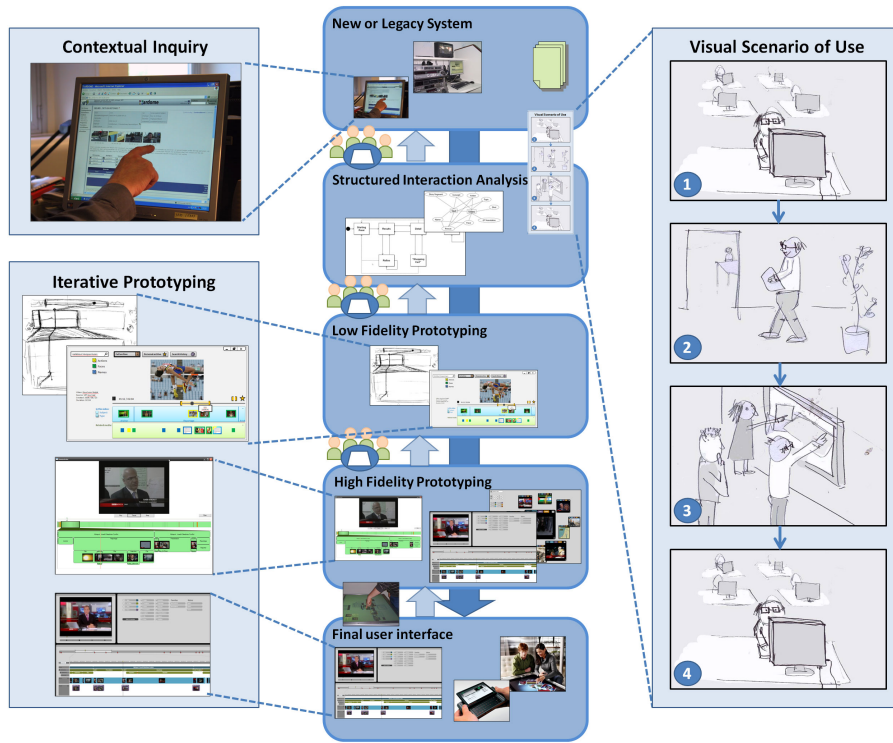


Fig. 1 Center: The user-centred process that was adopted for the development of the user interface. Left - Right: Detail of the most important techniques and artefacts.

system, and the requirements of the *new system* - a video archive explorer - are examined by conducting a Contextual Inquiry (CI) [3] in cooperation with domain experts of the Flemish public broadcasting company (see Figure 1, top left). A CI involves observing and interviewing end users while they are performing their daily activities.

This user study taught us that professional video searchers need to browse large amounts of data in order to find a suitable video fragment of a few minutes. Besides video searching, most video searchers were also responsible for creating manual annotations of the most recent TV broadcasts. The task and underlying considerations for manually annotating videos were also observed. Although in the video archive explorer annotations are automatically generated, some search strategies of video searchers are closely related to annotating videos. For example, the searchers carefully annotated several people occurring in one video fragment because a realistic search concerns finding a video fragment in which two particular people appear. During the CI, we observed that the main job of the end user is to search video fragments of TV

broadcasts presenting particular people or situations that will be used in a TV programme or news broadcast.

Video searchers start the search by consulting online search engines (e.g. Google, Yahoo, Wikipedia) to get familiar with the topic. From this online information, video searchers can deduce some relevant keywords. These keywords can then be used together with other search criteria (e.g. date or programme title) to find suitable video fragments in the vast archive of the TV broadcasting company.

After entering keywords and search criteria, a textual list of video search results and accompanying manual video annotations can be browsed to select videos that may be suitable. Once a video is selected from the archive, the video searcher has to browse the entire video manually in order to find and select a suitable fragment. The absence of a good visualisation of the content makes this a daunting and time consuming job. Efficiency decreases because the users need to combine several separate applications.

A typical search consists of finding a fragment that shows a particular person in specific circumstances. The resulting fragments from this job often have to meet certain qualitative criteria such as funny, unique and attractive. This means that the video searchers continue their search until they discover a fragment that matches the search query and meets these criteria.

An example search task that was carried out by the observed video searchers consists of finding a fragment representing a traffic jam. This fragment was needed for an item in the daily news broadcast. In this situation, the video searcher wants to ensure that this fragment looks like a fragment that was recorded that day. Therefore, she determines whether the found video fragment corresponds to the weather conditions of the day (e.g. rain, sun, snow), the season (e.g. bare trees or blossom), etc. Since this information is typically not included in the annotations, the exploration and selection of qualitative video fragments depends on human judgement. Depending on the purpose of the video fragment (e.g. TV show or news broadcast) and the obscurity of a fragment, a search can take five minutes to an entire day.

2.2 Visual Scenario of Use

The CI resulted into a scenario of use and an accompanying visual representation of it (see Figure 1, right). The scenario exemplifies how one integrated future application can be used for searching archives, browsing an archive video and adding video fragments to a favorites folder. In this visual scenario, a video searcher has to search and browse video fragments of a sportsperson in order to assemble a news item concerning the career of this person. First, the video searcher retrieves interesting videos in his office, behind his own desk (Figure 1-1). He saves the files on his mobile device. Following, he takes the mobile device and hurries to a meeting room where he will discuss the video fragments with some colleagues (Figure 1-2). In this meeting, a large multi-touch display is used to browse through the available video fragments

and to talk about the most suitable fragments (Figure 1-3). Afterwards, the video searcher goes back to his office and starts assembling the resulting video (Figure 1-4).

Browsing physical as well as digital libraries has been widely investigated in earlier work. Next to individual search, collaborative search is a search strategy that is commonly used [25,37]. Twidale et al. distinguished co-located vs. remote search, and synchronous vs. asynchronous search [37]. Two interesting categorisations for collaborative search are “joint search” involving a small group of people that share one computer and “coordinated search” that involves a small group of people using computers individually. Our visual scenario includes individual search (Figure 1-1) and joint search (Figure 1-2).

The visual scenario was used to discuss the application with the stakeholders. By annotating the visual scenario, it provided input for the later phases in the UCD process. While keeping in mind the visual scenario and its annotations, the task model (see the center rectangle in Figure 2), was created in the *structured interaction analysis* stage [12] (see the UCD process in the center of Figure 1). This task model describes the tasks of the proposed end user on a low level. Both the visual scenario and the task model were used during the *iterative prototyping*, which is discussed in the next section.

2.3 Iterative Prototyping

Prototypes created in the iterative prototyping process gradually evolved from low-fidelity prototypes to high-fidelity prototypes. The bottom left part of Figure 1, bottom left, shows the different prototypes that were created. Once a particular prototype was available, it was evaluated.

The first *low-fidelity prototypes* for a desktop and multi-touch user interface were created using pencil and paper and Powerpoint. These prototypes were evaluated in meetings that involved stakeholders and domain experts.

The *high-fidelity prototypes* were built in .NET, using C# and XAML. In this phase, a graphic designer delivered detailed UI designs. These first high-fidelity prototypes were evaluated in stakeholder meetings. Accordingly, the high-fidelity prototypes were adjusted. In this stage, we further specified the multi-touch interactions for the prototype. This extended version of the prototype was evaluated in a user study that involved professional video searchers. Besides evaluating the visualisations, this prototype allowed us to introduce the new technology and approach described in the visual scenario to the video searchers and to explore how they would use it.

The UI designs, visualisations, and interaction techniques included in the high-fidelity prototype for the multi-touch table are discussed in the following section. Section 4 explains the underlying annotations for this video archive explorer. The user evaluation of the video archive explorer is described in section 5.



Fig. 2 Taxonomy of the user tasks, and the related visualisations and interactions for the video archive explorer.

3 The Video Archive Explorer

This section discusses the most important tasks that the video archive explorer supports. This discussion is based on Figure 2, which outlines all tasks together with the visualisations and interaction techniques that the news archive explorer offers to complete these tasks.

3.1 Enter a Search Query

In order to find the right video, users start by entering a search query (Figure 2, task 1). A search query typically contains a keyword and/or date range (Figure 3-A) and is entered using a software keyboard input panel. After pressing the search button, the system will provide a set of videos that are relevant to the search query (Figure 3-B).

3.2 Explore Results

After retrieving the relevant search results, users can start exploring the search results (Figure 2, task 2). Typically, users will first do a high level exploration of the search results (Figure 2, task 3) and then take a closer look at the interesting videos in the advanced video player (Figure 2, task 4).

3.2.1 High Level Search Result Exploration

As shown in Figure 2, the video search explorer provides three techniques to explore search results on a high level: a keyframe slideshow, a video clock and a video timeline.

Keyframe Slideshow The video archive explorer represents each video as an animated slideshow of key-frames (Figure 3-C) that helps users to get a short overview of the most important and relevant facts in the underlying video. These key-frames are either frames that represent the different *stories* in a video or the key-frames which contain *faces* that are relevant to the search query. Both the stories and the faces slide shows are built using automatic



Fig. 3 The video archive explorer’s user interface after entering a search query for *Robert Mugabe* (A), showing the retrieved results (B). Every result (C) is animated and can show the relevant stories (D) or faces (E).

video annotations, which will be discussed in Section 4. A user can toggle between story or face *mode* by pressing on the story (Figure 3-D) or face (Figure 3-E) button at the top of a video search result.

The animated video slide shows can be explored visually using two storyboard visualisations: a video clock or a video timeline. Depending on the end user preferences, one of these two visualisations will be used.

Video Clock Visualisation The video clock shows video key-frames in a circle around a video search result. This circle appears when a user presses a search result (see Figure 4-1). When pressing a key-frame in this circle with a second finger (Figure 4-2), the part of the video corresponding to this key-frame will be played in a video player.

In the video clock, story or face key-frames are displayed depending if face or story mode has been selected. These key-frames are organised chronologically in a clockwise order: the frame that first appears in the video is located at the blue line (see Figure 4-3 A), the last one at the left side of this line. Key-frames of shots or faces that are most relevant to the entered search query are surrounded by a yellow border (see Figure 4-3 C and D). The clock’s green line (see Figure 4-3 B) acts as a *clock hand* that automatically goes over all key-frames. The middle of the clock shows an enlarged version of the key-frame the hand is currently pointing at.

Video Timeline Visualisation The video timeline displays a video as a horizontal line of keyframes under a video search result. Similar to the clock

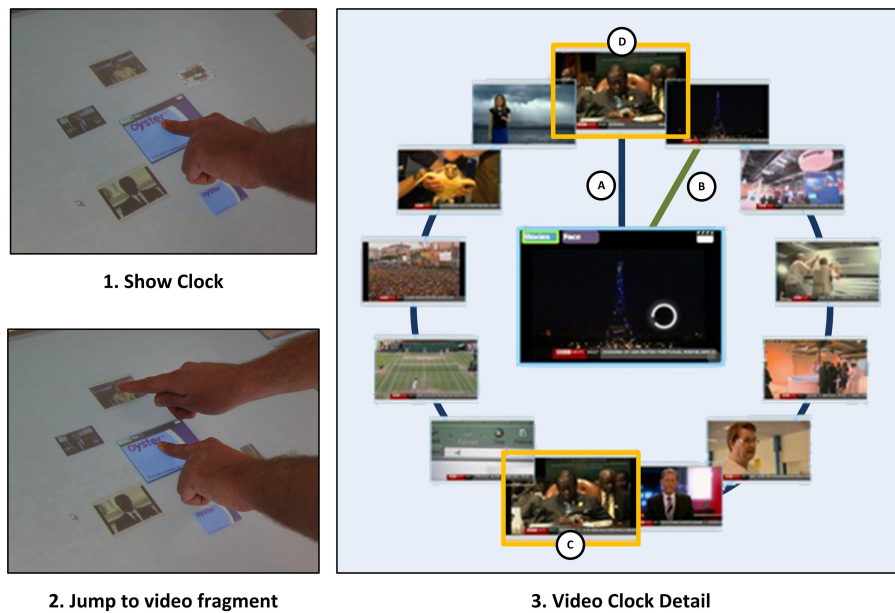


Fig. 4 The video clock visualisation aligns story or face key-frames around the video search result

visualisation, this line is shown when a user presses a search result with one finger (Figure 5-1) and when pressing a key-frame with a second finger (see Figure 5-2), the corresponding video fragment is shown in a video player. Depending on the currently selected mode, the video timeline will show story or face key-frames. These key-frames are shown chronologically from left to right. Frames that correspond to very relevant fragments are surrounded by a yellow border (Figure 5-3 A and B).

3.2.2 Watch Video in the Advanced Video Player

In order to have a more detailed look at a retrieved video result, users can watch a video using an advanced video player. This video player shows the video together with the automatically detected video annotations. When a user presses on a key-frame during video exploration (as described in section 3.2.1), the advanced video player opens and starts playing the video fragment that contains the pressed frame.

The advanced video player combines a time slider, based on the time sliders in commercial video players such as Apple Quicktime Player and Microsoft Windows Media Player, with a timeline video visualisation [15]. A time slider is employed to manipulate the current time of the played video fragment (Figure 6-A) and to specify an area of interest around this time (Figure 6-B). The timeline (Figure 6-C) gives a detailed view on the content in this area of interest.

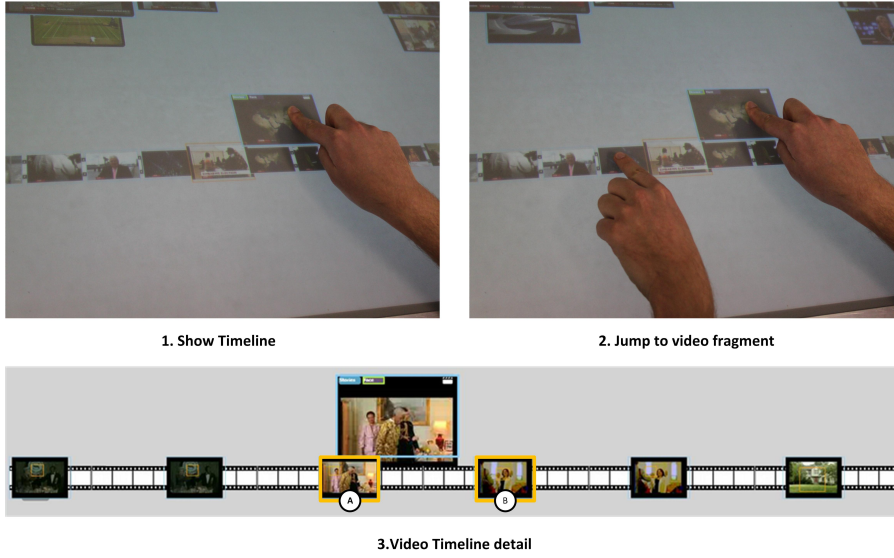


Fig. 5 The video timeline visualisation aligns story or face key-frames chronologically on a horizontal line under the video search result. 1 and 2 show a timeline with many keyframe results, 3 shows a timeline with only few results.

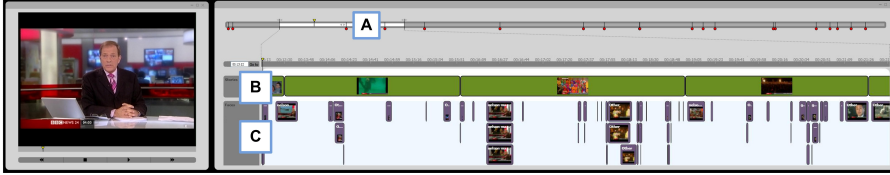


Fig. 6 The advanced video player aligns a video with the automatically detected annotations such as stories and faces.

The combination of the area of interest and the timeline contributes to the basic idea of *focus+context* in information visualisation [5]. *Context*, on the one hand, is visualised using the video time slider, where red dots indicate the parts of the video relevant to the search query. *Focus*, on the other hand, is visualised in the timeline. Similar to other timeline based approaches [30], we use semantic zooming to specify the level of detail in the timeline. By resizing the focus area in the time slider (Figure 6-A), users can zoom in or out on the video timeline. A small focus area increases the level of detail in the timeline, a wider focus area decreases this level of detail as shown in Figure 7-1. Resizing the focus area is done using a *pinch to zoom* gesture, which is shown in Figure 7-2.

The timeline shows a layered view of the video as computed by the automatic video annotation algorithms. At the first layer (Figure 6-B), the different stories that occur in a video are shown. Every story is represented by a key-frame thumbnail image. The second layer (Figure 6-C) shows all persons

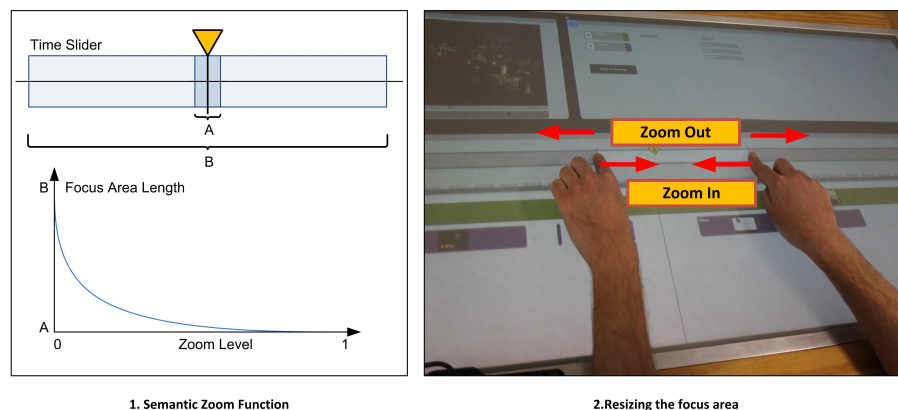


Fig. 7 Semantic zooming is applied using a pinch to zoom gesture.

that are detected in the videos. These persons are visualised by a key-frame thumbnail image in combination with the person’s name.

3.3 Organise Videos

While exploring the search results, users get a better understanding of the retrieved videos which they can use to organise the results (Figure 2, task 5). Using standard touch and multi-touch operations, search results can be moved, rotated and resized. For example, users can create piles of good and bad search results or enlarge their favorite results.

3.4 Background Information

Background information shows contextual information about a search query using other media like images and text. This information helps users getting familiar with the query topic. While entering a search query, the background information appears for the first time and remains visible throughout the whole search process (Figure 2, right rectangle). For example, when searching videos that contain “*Robert Mugabe*”, as well pictures as a textual description of this person are shown (Figure 3-F).

4 Automatic Video Indexing

Automatic Video Indexing is an essential technique to create a video archive that can be efficiently searched. A full discussion of the automatic video indexing techniques fall out of the scope of this paper. For an overview of our work on this we refer the reader to [28,29,31]. Since efficiently searching a large video archive requires appropriate indices to be available, our approach

relies on the transcripts (actually time aligned subtitles) on which standard text-based retrieval methods give accurate results. Additionally, because people are the most common and often also the most relevant subject in the visual datastream, we propose a method to automatically link faces in the video with names in these transcripts (see section 4.1). This allows to search based on the visual content, and to select appropriate keyframes for the advanced video player. Moreover, the original news videos need to be split into smaller, semantically coherent stories. This is achieved using a multimodal analysis, as described in section 4.2.

4.1 Assigning Names to Faces

Labeling persons appearing in video frames with names detected from the video transcript helps improving the video content identification and the search task. We have developed an unsupervised technique that exploits the cross-modal nature of names in text and faces in video to identify important persona in the news broadcasts. We have explored different alignment schemes for assigning names to the faces, assigning faces to the names, and establishing name-face link pairs. On top of that, we use textual and visual structural information to predict the presence of the corresponding entity in the other modality. The techniques are described in detail by Pham et al. [28].

To further improve the performance of the name-face alignments in news video, a face naming method that learns from labeled and unlabeled examples using iterative label propagation in a graph of faces connected by their visual similarity, is developed. The advantage of this method is that it can use very few labeled data points and incorporate the unlabeled data points during the learning process. We have shown that the algorithm is successful and better copes with noisy face detections than using a traditional supervised classifier [29]. Anchor detection and metric learning for computing the similarity between faces are incorporated into the label propagation process to help boosting the face naming performance.

4.2 Story Segmentation

The story segmentation algorithm we are using is described in detail by Poullisse et al. [31]. In text, a story segment is a coherent grouping of sentences, discussing related topics and names. The multimedia equivalent, such as found in video, would be a temporal segment containing imagery accompanied by a spoken description of the single event. Three different channels: text (transcribed speech available via subtitles), video and audio are at our disposal to accomplish this segmentation task. In order to identify potential story boundaries and thus identify the individual news items, we examine a number of features that model the likelihood of such a story boundary occurring. A number of features measure the lexical cohesion of two adjacent text passages, while others model this by examining the repetition of named entities or the occurrence

of typical cue words and phrases that mark the opening or closing of a news story. We use the output of a shot cut detector as a feature, as it is likely that visual changes closely relate to the transition from one news story to another. Finally, we consider the speaker pause duration of the news anchor; a news anchor is more likely to pause for a longer duration when he is transitioning from one story to another.

All these features were used to train a single classifier which can be used to accurately determine story segments. The segmentation algorithm works by initially seeding an unseen text with random story boundaries, and progressively moving these boundaries around to find better story segments, as determined by the classifier. After a number of iterations and a number of restarts, the resultant segmentation is the average of all the trials. Interestingly, while some features were more effective than others, the best results came when many features were combined in synergy.

5 Evaluation

We conducted an informal qualitative evaluation of the video archive explorer in order to gain feedback about the usefulness of various features in the tool and the usability of the basic interactions. This lab evaluation is part of the UCD process presented in section 2. The results will serve as input for new iterations of the design and development of the video archive explorer.

Ten domain experts, five male and five female, participated in the study. Searching videos constitutes at least twenty percent of the participants' workload. Eight participants spend even more than fifty percent of their job on searching the TV video archive.

The next sections present the evaluation setup and the test results, followed by a discussion. When discussing test results related to a particular user task, we refer to the tasks that are shown in Figure 2.

5.1 Evaluation Setup

The system that we used for the evaluation consisted of a custom made horizontal multi-touch table with rear screen-projection. This table tracked touch-points using the Frustrated Total Internal Reflection (FTIR) method [14]. The participants only interacted with the table and did not use any mouse or keyboard.

For the evaluation of the video archive explorer a sample archive of BBC broadcasts about the elections in Zimbabwe was preprocessed to obtain the automatically generated annotations. This archive allowed us to conduct an evaluation of the features of the video archive explorer. In order to obtain specific feedback of the visualisation, the subjects were observed while performing some video exploration tasks in a lab environment.

In each evaluation session, the participant first had the opportunity to get familiar with the interactions on the multi-touch table by resizing and



Fig. 8 Some photographs that were taken during the evaluation.

rotating some pictures. The second part consisted of an introduction of the video archive explorer and the interactions. We gave the participant a written manual, asked her to read the manual and to freely explore the different parts of the video archive explorer’s GUI.

The main part of the evaluation session concerned four video search tasks, where the participant had to search for two fragments that contained a given person and two fragments that handled about a given topic. Since the video archive explorer concentrates on the exploration of search results, the influence of the query composition was excluded by providing the exact queries to the subjects beforehand. Example queries include: “Robert Mugabe” and “War in Africa”. Following, for each search task, the subject was asked to select a video fragment that was the most suitable for her situation. The goal of the task was finding a fragment that meets the quality that our subjects pursue in their everyday jobs and not quickly finding any video fragment for a query.

The search tasks were assigned using a latin square experimental design where we ensured that each participant used the video clock visualisation for a topic and a person, and the video timeline visualisation for searching the other topic and person. Since the search criteria for finding a suitable video fragment are often defined based on personal experience and differ for each subject, we did not measure the execution time of search tasks. In this evaluation, we were interested in observing how each participant completed the realistic search tasks and how she used the visualisations and interactions in the video archive explorer.

Figure 8 shows some photographs that were taken during the experiment. During the observations, a think aloud protocol was employed to understand the actions and decisions of the participant. Additional questions were asked by the facilitator if necessary. After the participant finished her task, she filled

out a questionnaire asking about the video archive explorer’s core features in several situations. Most questions asked for the subject’s opinion using a five-point Likert scale on which 1 indicates a negative appreciation and 5 indicates a positive appreciation. Furthermore, the visual scenario described in section 2 was used to explain the situation in which this application could be used and to consult the participant about this scenario of use.

5.2 Test Results

Several interesting observations in combination with the participants’ answers to the questionnaire teach us more about the use of the video archive explorer. After the first part of the test, most of the subjects got familiar with the interactions of the multi-touch table. Although some of them were rather hesitant and careful in the first part of the test, they got used to touching the table in a correct way while carrying out the test. They all could easily use the virtual keyboard to enter search queries.

After entering the search query (Figure 2, task 1), most participants started browsing the search results immediately (Figure 2, task 2). The subjects were very impressed by the fact that they could visually browse all video previews (Figure 2, task 3), instead of using a textual list. After browsing the previews, they selected a particular video to watch it in the advanced video player (Figure 2, task 4). When they wanted to explore more videos, they navigated back to the set of video previews (Figure 2, task 3). Observations and comments of most participants show that they expect videos that were already watched to be marked or repositioned. Some participants mentioned that they were missing textual information, containing the number of search results, time codes, descriptions of stories and other textual annotations next to the previews.

When searching a face that was not familiar to the subjects, we observed that they all used the background information shown next to the search results and in the advanced video player (Figure 2, bottom rectangle). Mainly the images shown in this background information helped them to find a video that contained the person they were looking for. The subjects commented: *“This is easy! I did not know what the person I was searching for looks like.”* *“This is interesting! In the current system, I have to precede my search of people that I do not know by consulting search engines and Wikipedia.”* In the questionnaire, they judged this feature as a useful feature ($Mean = 4.5$, $Median = 4.5$, $\sigma = 0.53$).

The subjects did not have any preference concerning the video clock and video timeline visualisations to explore search results on a high level (Figure 2, task 3). Both visualisations were judged as easy to use. Some subjects noticed that their preference for the visualisation may depend on the number of available search results. Consequently, they prefer the possibility to switch between both visualisations.

Most subjects were very positive that in the advanced video player, a video automatically starts playing the fragment that they selected in the set of video

previews (Figure 2, task 4). According to the results of the questionnaire, the automatic recognition of stories and faces in a video would increase efficiency when searching video fragments ($Mean = 5$, $Median = 4.7$, $\sigma = 0.48$). Some of them suggested that automatic recognition of faces and topics is interesting in combination with the recognition of actions. One subject was surprised that people on the background are also recognized, while another subject noticed that this system can avoid incorrect search results because of spelling mistakes in manual annotations.

In the questionnaire we asked whether the subjects would use a desktop version of this application. All subjects answered positive on this question ($Mean = 5$, $Median = 4.6$, $\sigma = 0.51$). Not all subjects agreed that this desktop application would be better than the current system, but in general they were positive ($Mean = 4$, $Median = 3.6$, $\sigma = 1.07$). The subjects that were not completely convinced about the benefits of the video archive explorer, mentioned that they would like to use this application as an additional system, and maybe would switch to it after a while.

Since the focus of this evaluation was on the visualisations and automatic detection of search results, we did not include another factor in this experiment to evaluate the collaborative aspect of the video archive explorer. Instead we used the visual scenario of use to discuss collaborative searching using the video archive explorer. This type of representation is very suitable to consider multiple solutions and to share and discuss conceptual ideas with stakeholders and end users [13, 27, 9]. We presented the visual scenario discussed in section 2 to the subjects and asked in the questionnaire whether the application on the multi-touch table was useful in the context of use presented by this scenario. Not all subjects select fragments in a meeting, but some subjects commented that the current application can be useful during brainstorm meetings for quickly searching the availability of some fragments about a particular topic or person. Furthermore, the application can be used in meetings where video fragments are edited to prepare a news item or documentary. For instance the application on multi-touch can be used to quickly search a suitable fragment that fits to some audio.

5.3 Discussion

The following describes the lessons we learned of the test results. Interesting issues and challenges for next iterations of the system will be discussed.

Entering a search query (Figure 2, task 1) in the system was straightforward, no subjects had problems with this part of the user interface. Once a set of videos was shown, most subjects started exploring these results (Figure 2, task 2). Nevertheless, some of the subjects tried to map our prototype to the system they currently use in their job and expected some kind of textual information about the search results. This issue can be met by adding a small popup or tooltip which gives more information about a video or a particular

key-frame in the video (such as time codes, descriptions of stories and other textual annotations).

The visualisations to explore search results (Figure 2, task 3) were considered as an improvement by the subjects. A visual representation of the search results, shown immediately after the query is entered, allows users to select suitable videos before consulting any textual list of search results. When we asked which visualisation they would favour in their job, there was no obvious preference between both visualisations: the video timeline and the video clock. An interesting issue here is what visualisation would be most appropriate when one search result contains a lot of interesting key-frames.

The use of the advanced video player (Figure 2, task 4) depends on the personal search strategies, preferences and search queries of end users. In the current evaluation, the end users were impressed by the fact that the video player automatically navigates to the video fragment that seemed relevant in the preview of the search results. Some subjects used the annotated timeline to watch and navigate the video, while other subjects only used the time slider. For end users, it can be interesting to include aligned subtitles to the timeline view.

We observed that some subjects sorted the search results (Figure 2, task 5) based on the search history. For some other subjects it was not always clear whether they already watched a particular search result or not. This problem can be solved by adding a search history and marking the search results that have been watched.

All subjects would like to use this application on their desktop pc. An additional version of the application for a multi-touch table, can be used during meetings.

6 Related Work

In this section we review the work which relates to our development of a news archive explorer for professional video searchers. In particular we provide an overview of the research done on user-centred design, multimedia retrieval and multimedia interaction techniques.

6.1 User-Centred Design

A User-Centred Design (UCD) approach is recommended by ISO 13407 to design and develop systems that have an increased user satisfaction and productivity and are easier to understand [20]. The activities presented by ISO 13407 are present in several UCD methodologies such as Rapid Contextual Design [3], GUIDE [32] and Effective Prototyping [2]. The UCD process we used to design and develop the video archive explorer for video search professionals is derived from the MuiCSer process framework [12], which supports the aforementioned UCD methodologies but combines several techniques from UCD and Software Engineering.

Most existing interactive visualisations of large video archives mainly focus on efficiency and reliability of automatically generated search results and technological aspects. However, for the design and development of IGroup [39], an interactive visualisation for web image search results, several UCD techniques are applied. Smeaton et al. [33] present a tabletop system for collaborative video search, which is developed based on some techniques of UCD. They compare two designs for collaborative interaction in a lab experiment.

Wassink et al. [40] present a whole UCD approach for interactive visualisation design. In their work, the use of paper-based prototypes and visual scenarios is recommended to describe the initial ideas about a system. The MedioVis system [11, 18] is illustrated using a visual scenario [26], to present the project description and how the system works to those who are interested. Consequently, the use of a sketched visual scenario as presented in section 2 is not new. However, we enrich these visual scenarios by highlighting annotations and reuse them in later stages of the process [13]. This approach and accompanying tool support the UCD team to use visual scenarios for the creation and verification of several design results, the preparation of usability evaluations and the discussion of design solutions with end users.

6.2 Multimedia Retrieval

A good impression of the latest state-of-the-art in multimedia retrieval systems can be obtained from the yearly organised challenges TrecVid [34] and VideoOlympics [35]. With its video corpus of up to 400 hours of video material, TrecVid focusses especially on large scale video retrieval performance. One of its core tasks is interactive video search. Here users are allowed to rephrase their queries in response to initial results, putting more stress on the user interface. The VideoOlympics challenge takes this idea one step further and lets systems compete in realtime in front of a live audience, both with expert users (usually the system developers) as well as with novice users unfamiliar with the system. This way, the influence of interaction mechanisms and advanced visualisations in the interface becomes clear: apart from efficient, the interface should also be intuitive and user friendly. Our experiments work under a slightly different setting, in that we assume that, apart from the raw video input, also textual transcripts of the videos are available, as is usually the case in a professional news archive. In our case, these were extracted from time-aligned subtitles.

One of the best known video retrieval systems is the Informedia project of CMU [17]. This system has evolved over more than a decade, and exists in different versions and flavours. They were the first to advocate the use of large lexicons of visual concept detectors for retrieval, in an attempt to bridge the semantic gap between the low-level cues obtained from a color, shape, and texture-based image analysis and the high-level information needs of users, often expressed with text descriptions [16]. Apart from the standard storyboards, they also experiment with visualisations using timelines (rank video

thumbnails chronologically on a time axis), map views (show the geographic distribution of locations mentioned in a set of videos), and many others [7].

Another very successful system is Mediamill developed by the University of Amsterdam [41]. This tool exploits the relation between video keyframes to support video browsing. For example, a user can easily explore the keyframes that contain similar concepts or explore video keyframes in a linear way. Our video archive explorer builds on this idea, allowing users to explore a video based on detected concepts such as stories or faces. While Mediamill supports the sequential exploration of videos, we use multi-touch technology to explore multiple videos at the same time. This allows users to compare multiple videos by showing their relevant keyframes in a timeline or clock visualisation. Moreover, the automatic linking of names and faces makes sure the system is automatically up-to-date without the need for a separate annotation phase.

Finally, the MediaMagic system developed by researchers of FX Palo Alto Laboratory [1] is worth mentioning. This system for interactive video search focusses on a rapid assessment of query results and easy pivot from those results to form new queries, maximizing the benefit of the user in the loop. This is accomplished by a streamlined interface and redundant visual cues throughout. Also an extension to two searchers collaborating in real time is provided. Such a collaborative setting seems especially well suited for new interface devices such as the multi-touch table used in our work.

6.3 Video Interaction Techniques

Multimedia interaction techniques are advanced techniques for navigating through a video in order to find the intended fragment. In most traditional video players like VLC² or Windows Media Player³, a video can be skimmed by manipulating a time slider. The approaches presented in this section go beyond this traditional time slider.

A first category of works like the Smart Player [6] and elastic skimming [19] help users to scale some parts of the video time slider different from specified. This allows focussing on certain parts of the video without losing the context of the whole video. The Smart Player [6] proposes the “scenic car driving” metaphor, where the video is played slowly near interesting events and speeds through unexciting parts of the video. This allows users to see the whole video and to pay attention to the particularly most interesting fragments. Huřst et al. [19] propose elastic skimming, a combination of the traditional video time slider and elastic graphical user interfaces [23]. Elastic skimming relates the skimming speed to the distance between the mouse pointer and the thumb on the time slider. The news video archive explorer also allows users to focus on certain parts of a video. Therefore, we provide a focus area that users can stretch and move. The part of the video within this area is rescaled on

² <http://www.videolan.org/vlc/>

³ <http://windows.microsoft.com/nl-NL/windows/products/windows-media-player>

a larger timeline which gives more information about the events and persons that appear in this part.

Other works like DRAGON [22] and relative flow dragging [10] use direct manipulation to skim videos. These techniques make it possible to click on moving objects in the video. A clicked object can then be moved along its trajectory line. This is particularly useful in certain domains like football or traffic surveillance where it matters if objects were at a certain place or not. In these cases, direct manipulation skimming systems help to answer questions like “did the ball cross the line?” or “did the car hit that object?”. This is different from our video archive explorer approach, which supports more open search tasks where video fragments have to meet certain qualitative criteria such as being funny and attractive. In certain cases, e.g. when searching for a fragment in a football match, direct manipulation video skimming might complement this process.

A third category of video interaction techniques uses still-image abstraction to construct video summaries. These abstractions are mostly shown by means of storyboards [4, 42, 7], which are sequences of key frames that represent the most important scenes in a video. A special case of storyboarding tools is video manga [38], where more important scenes have a larger size. Still-image abstraction techniques are important for giving a high-level overview of the contents in a video. Therefore we also incorporated this technique in the video archive explorer by means of the video clock and video timeline. This helps users to gain an idea of the most important contents of a video fragment even before they have opened it.

7 Conclusions and Future Work

In this paper, we presented a video archive explorer for professional video searchers. This tool was developed using a user-centred (UCD) design process in order to take into account the goals and needs of the future end users. The resulting prototype combined automatic video indexing methods and interactive visualisations. A qualitative evaluation with professional video searchers showed that the combination of these visualisations and automatic detection of videos is likely to fit to the end user needs and can result in an increased search experience and efficiency. Compared to the current system, the visual exploration of search results in one integrated system eliminates the step in which the user searches a textual list of search results and their annotations before watching the video. The automatically generated annotations can complement the manual annotations and increase their accuracy. For instance, since the face recognition algorithm detects faces in the background of a video frame, many more annotations can be considered in the archive.

It was difficult for end users to imagine a future system that combines multimedia information and new technologies during the Contextual Inquiry. Therefore, it was necessary to consider user needs throughout the entire design and development process. The resulting video archive explorer that was

evaluated, seemed to meet the user requirements and clearly gave the participants of the evaluation a better understanding of the opportunities of a future system that supports their job. In combination with the visual scenario, the prototype invites users to comment about particular features and encourages them to think about possible improvements of the system.

Frequently involving end users in an iterative development process, such as the process used for the video archive explorer, enables to fine-tune the application. In the next iteration, we will consider the feedback from the participants of the evaluation to adjust the application. For instance, we will bridge the gap between the mainly textual search strategies of the current search tools and the visual exploration of video archive by adding more textual descriptions to the search results. Other collaborative search strategies for the multi-touch application will be explored and evaluated. Furthermore, we will examine possible visualisations and interactions for filtering a large number of search results.

In future work, we also plan to investigate how users can benefit from our video archive explorer on other computing platforms such as mobile devices and tablet PCs. To efficiently maintain and coordinate the multi-platform development of our application, additional tool support will be needed. An interesting source of inspiration for managing such complex development approaches is D-Macs [24], a tool that allows to automate design and development effort across platform-specific GUIs.

Acknowledgements

This research was supported by the IWT Project AMASS++ (SBO-060051). We would like to thank all members of the AMASS++ user committee for their feedback about the results discussed during the half-yearly user committee meetings. Special thanks to the Flemish broadcasting company VRT to contribute to the user studies. Grateful acknowledgement is given to Koen Deschacht, Wim De Smet, Scott Martens, Ineke Schuurman, Vincent Vandeghinste, Frank Van Eynde and Luc Van Gool. Furthermore, we would like to thank Jimmy Cleuren for his contribution to the high-fidelity prototypes and Karel Robert for the graphic design of the user interface.

References

1. John Adcock, Matthew Cooper, and Jeremy Pickens. Experiments in interactive video search by addition and subtraction. In *Proceedings of the 2008 international conference on Content-based image and video retrieval*, CIVR '08, pages 465–474, New York, NY, USA, 2008. ACM.
2. Jonathan Arnowitz, Michael Arent, and Nevin Berger. *Effective Prototyping for Software Makers (The Morgan Kaufmann Series in Interactive Technologies)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2006.
3. Hugh Beyer and Karen Holtzblatt. *Contextual Design: Defining Customer-Centered Systems*. Morgan Kaufmann Publishers Inc., ISBN: 1-55680-411-1, 1998.

4. John Boreczky, Andreas Girgensohn, Gene Golovchinsky, and Shingo Uchihashi. An interactive comic book presentation for exploring video. In *CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 185–192, New York, NY, USA, 2000. ACM.
5. S. K. Card, J. D. Mackinlay, and B. Shneiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.
6. Kai-Yin Cheng, Sheng-Jie Luo, Bing-Yu Chen, and Hao-Hua Chu. Smartplayer: user-centric video fast-forwarding. In *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*, pages 789–798, New York, NY, USA, 2009. ACM.
7. Michael G. Christel. Supporting video library exploratory search: when storyboards are not enough. In *CIVR '08: Proceedings of the 2008 international conference on Content-based image and video retrieval*, pages 447–456, New York, NY, USA, 2008. ACM.
8. Michael G. Christel, Michael A. Smith, C. Roy Taylor, and David B. Winkler. Evolving video skims into useful multimedia abstractions. In *CHI '98: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 171–178, New York, NY, USA, 1998. ACM Press/Addison-Wesley Publishing Co.
9. Sal Cilella. Did you ever know that you're my hero?: the power of storytelling. *interactions*, 18:62–66, January 2011.
10. Pierre Dragicevic, Gonzalo Ramos, Jacobo Bibliowicz, Derek Nowrouzezahrai, Ravin Balakrishnan, and Karan Singh. Video browsing by direct manipulation. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, CHI '08, pages 237–246, New York, NY, USA, 2008. ACM.
11. Christian Grün, Jens Gerken, Hans-Christian Jetter, Werner A. Knig, and Harald Reiterer. Medioviz - a user-centred library metadata browser. In *ECDL 2005: Proceedings of the 9th European Conference on Research and Advanced Technology for Digital Libraries 2005*, pages 174–185. Springer Verlag, Sep 2005.
12. Mieke Haesen, Karin Coninx, Jan Van den Bergh, and Kris Luyten. Muicser: A process framework for multi-disciplinary user-centred software engineering processes. In *HCSE '08: Second Conference on Human-Centered Software Engineering*, pages 150–165, 2008.
13. Mieke Haesen, Jan Meskens, Kris Luyten, and Karin Coninx. Draw me a storyboard: Incorporating principles and techniques of comics to ease communication and artefact creation in user-centred design. In *In Proc. of BCS Conference on Human Computer Interaction*, September 2010.
14. Jefferson Y. Han. Low-cost multi-touch sensing through frustrated total internal reflection. In *UIST '05: Proceedings of the 18th annual ACM symposium on User interface software and technology*, pages 115–118, New York, NY, USA, 2005. ACM.
15. Alexander Haubold and John R. Kender. Vast mm: multimedia browser for presentation video. In *CIVR '07: Proceedings of the 6th ACM international conference on Image and video retrieval*, pages 41–48, New York, NY, USA, 2007. ACM.
16. Alexander G. Hauptmann, Michael G. Christel, and Rong Yan. Video Retrieval Based on Semantic Concepts. *Proceedings of the IEEE*, 96(4):602–622, April 2008.
17. Er Hauptmann, Robert V. Baron, Ming yu Chen, Michael Christel, Lily Mummert, Steve Schlosser, Xinghua Sun, Victor Valdes, and Jun Yang. Informedia trecvid2008: Exploring new frontiers. In *Trecvid 2008*, 2008.
18. Mathias Heilig, Mischa Demarmels, Werner A. König, Jens Gerken, Sebastian Rexhausen, Hans-Christian Jetter, and Harald Reiterer. Medioviz: visual information seeking in digital libraries. In *Proceedings of the working conference on Advanced visual interfaces*, AVI '08, pages 490–491, New York, NY, USA, 2008. ACM.
19. Wolfgang Hürst, Georg Götz, and Philipp Jarvers. Advanced user interfaces for dynamic video browsing. In *Proceedings of the 12th annual ACM international conference on Multimedia*, MULTIMEDIA '04, pages 742–743, New York, NY, USA, 2004. ACM.
20. International Standards Organization. *ISO 13407. Human Centred Design Process for Interactive Systems*. Geneva, Swiss, 1999.
21. Kankanhalli, M. S. and Rui, Y. Application Potential of Multimedia Information Retrieval. *Proceedings of the IEEE*, 96(4):712–720, March 2008.

22. Thorsten Karrer, Malte Weiss, Eric Lee, and Jan Borchers. Dragon: a direct manipulation interface for frame-accurate in-scene video navigation. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 247–250, New York, NY, USA, 2008. ACM.
23. Toshiyuki Masui, Kouichi Kashiwagi, and George R. Borden, IV. Elastic graphical interfaces to precise data manipulation. In *Conference companion on Human factors in computing systems*, CHI '95, pages 143–144, New York, NY, USA, 1995. ACM.
24. Jan Meskens, Kris Luyten, and Karin Coninx. D-macs: building multi-device user interfaces by demonstrating, sharing and replaying design actions. In *Proceedings of the 23rd annual ACM symposium on User interface software and technology*, UIST '10, pages 129–138, New York, NY, USA, 2010. ACM.
25. Meredith Ringel Morris. Interfaces for collaborative exploratory web search: Motivations and directions for multi-user design. In *ACM SIGCHI 2007 Workshop on Exploratory Search and HCI: Designing and Evaluating Interfaces to Support Exploratory Search Interaction*, pages 9–12, 2007.
26. University of Konstanz HCI Group. Medioviz - project description. <http://hci.uni-konstanz.de/index.php?a=research&b=projects&c=16314278&lang=en#description>.
27. Jeff Patton. Consider multiple solutions. *IEEE Software*, 25:72–73, 2008.
28. Phi The Pham, M.-F. Moens, and T. Tuytelaars. Cross-media alignment of names and faces. *Multimedia, IEEE Transactions on*, 12(1):13–27, 2010.
29. Phi The Pham, Marie-Francine Moens, and Tinne Tuytelaars. Naming persons in news video with label propagation. In *International Workshop on Visual Content Identification and Search (VCIDS)*, pages 1528–1533, 2010.
30. Catherine Plaisant, Brett Milash, Anne Rose, Seth Widoff, and Ben Shneiderman. Lifelines: visualizing personal histories. In *CHI '96: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 221–ff., New York, NY, USA, 1996. ACM.
31. Gert-Jan Poulisse, Marie-Francine Moens, Tomas Dekens, and Koen Deschacht. News story segmentation in multiple modalities. *Multimedia Tools Appl.*, 48:3–22, May 2010.
32. D. Redmond-Pyle and A. Moore. *Graphical User Interface Design and Evaluation*. Prentice Hall, London, 1995.
33. Alan Smeaton, Hyowon Lee, Colum Foley, and Sin'ead McGivney. Collaborative video searching on a tabletop. *Multimedia Systems*, 12:375–391, 2007. 10.1007/s00530-006-0064-7.
34. Alan F. Smeaton, Paul Over, and Wessel Kraaij. Evaluation campaigns and trecvid. In *Proceedings of the 8th ACM international workshop on Multimedia information retrieval*, MIR '06, pages 321–330, New York, NY, USA, 2006. ACM.
35. Cees G. M. Snoek, Marcel Worring, Ork de Rooij, Koen E. A. van de Sande, Rong Yan, and Alexander G. Hauptmann. VideOlympics: Real-time evaluation of multimedia retrieval systems. *IEEE MultiMedia*, 15(1):86–91, January/March 2008.
36. Anthony Tang, Saul Greenberg, and Sidney Fels. Exploring video streams using slit-tear visualizations. In *AVI '08: Proceedings of the working conference on Advanced visual interfaces*, pages 191–198, New York, NY, USA, 2008. ACM.
37. Michael B. Twidale, David M. Nichols, and Chris D. Paice. Browsing is a collaborative process. *Inf. Process. Manage.*, 33:761–783, December 1997.
38. Shingo Uchihashi, Jonathan Foote, Andreas Girgensohn, and John Boreczky. Video manga: generating semantically meaningful video summaries. In *MULTIMEDIA '99: Proceedings of the seventh ACM international conference on Multimedia (Part 1)*, pages 383–392, New York, NY, USA, 1999. ACM.
39. Shuo Wang, Feng Jing, Jibo He, Qixing Du, and Lei Zhang. Igroup: presenting web image search results in semantic clusters. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '07, pages 587–596, New York, NY, USA, 2007. ACM.
40. I. Wassink, O. Kulyk, E.M.A.G. van Dijk, G.C. van der Veer, and P.E. van der Vet. Applying a user-centred approach to interactive visualization design. In *Trends in Interactive Visualization*. Springer Verlag, London, 2008. ISBN=978-1-84800-268-5.
41. Marcel Worring, Cees G. M. Snoek, Ork de Rooij, Giang P. Nguyen, and Arnold W. M. Smeulders. The MediaMill semantic video search engine. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages –, Honolulu, Hawaii, USA, April 2007. *Invited paper*.

-
42. HongJiang Zhang, Shuang Yeo Tan, Stephen W. Smoliar, and Gong Yihong. Automatic parsing and indexing of news video. *Multimedia Systems*, 2:256–266, 1995. 10.1007/BF01225243.