Human-Computer Interaction -- INTERACT'03
M. Rauterberg et al. (Eds.)
Published by IOS Press, (c) IFIP, 2003, pp. 932-935

Patch-based video browsing

Ynze van Houten, Mark van Setten & Jan-Gerrit Schuurman

Telematica Instituut
P.O.Box 589, 7500 AN Enschede, The Netherlands

Abstract: Interacting with video content can be very time-consuming, and solutions are needed to support people in the process of browsing a video so that they can efficiently interact with the video content. Based on information foraging theory, we have developed a video content browsing tool. Video content is accessed by means of "patches", which provide interaction with the video document that partly surpasses the burden of video's time-based nature. Patches can be of various types and can be combined using logical operators. Within-patch and between-patches browsing behaviour is well supported. The browsing process is considered to be the (fast) alternation between metadata-driven and data-driven interaction. The current application's priority is to support this browsing process. Preliminary test results indicate that a "patch–approach" is very fruitful. Important research issues are identified related to video content descriptions and browsing behaviour.

Keywords: multimedia, foraging, video patch, metadata, content-based access, browsing heuristics

1 Introduction

Video plays an important role in our culture, and we are bombarded with it all the time. The current trend is that more video is made more easily available to more people: digital video cameras are becoming more affordable, analog video recordings are often digitised, processing power is increasing, storage capacity is increasing, compression techniques are improving, the internet keeps growing (more video is put on-line), the bandwidth of internet connections is increasing, etcetera.

Interacting with video, however, is not straightforward. Video is a rich, but also very complex multimedia type communicating sight, sound, motion, time and space, often completed with synthetic images, sounds, and texts. Video represents a lot of information relevant to different situations. However, perhaps the biggest problem with video is its time-based nature, making the process of interacting with video very time-consuming. One can very easily get in the situation of encountering 1, 10, 100 or more hours of video material that "may be relevant" or "potentially interesting". The main issue regarding video interaction is *efficiency*.

The practical problem that is addressed here is how to support people in the process of browsing through a video, given its complexity. Typical usage includes: find out what a video is about, find out whether X or Y appears in a video, or have fun.

This paper describes a prototype of a video content browser of which the design is based on information foraging theory. Video content is accessed by means of "patches", which provide interaction with the video document that partly surpasses the burden of video's time-based nature. The design is aimed at supporting within-patch and between-patches browsing behaviour. The browsing process is considered to be the (fast) alternation between metadata-driven and data-driven interaction with video content. The current priority is to support this type of browsing. An iterative design approach was taken including a heuristic evaluation and a walkthrough. The developed prototype will be used to study browsing behaviour.

2 Design starting points

A lot of research is aimed at automating the process of defining meaningful elements in a video, extracting a meaningful structure of the video, and providing fast access to the video content using these data (Yeo & Yeung, 1997). A typical scheme of video-content analysis and indexing involves four primary processes: feature extraction, structure analysis, abstraction, and indexing (Dimitrova et al, 2002). In Lee & Smeaton (2002), an orderly overview of video browsing applications and related issues can be found.

We applied ideas from information foraging theory (Pirolli & Card, 1999) to develop a video browsing application. Information foraging theory is an approach to understanding how strategies and technologies for information seeking, gathering, and consumption are adapted to the flux of information in the environment The theory states that a person "forages" through an information space in search of a piece of information in a manner similar to that employed by an animal on the forage for food. The information environment is said to have a "patchy" structure, analogous to the way berries can be found in patches on berry bushes. Information foraging often involves navigating through spaces (physical or virtual) to find high-yield patches. A patch may be a collection of documents or an individual document viewed as a collection of content. Just like a bird foraging for berries, an information forager can stay within a patch, or has to face the decision to leave the patch and seek a new one. There are tradeoffs between 1) time spent identifying patches, 2) time spent moving between patches, 3) time spent enriching a patch, and 4) time spent foraging within a patch. Once in a patch, the forager engages in within-patch foraging and faces the decision of continuing to forage the patch or leaving to seek a new one. Information foragers use cognitive strategies to allocate their time to between-patches versus within-patch foraging activities. A cognitive strategy will be superior to another if it yields more useful information per unit cost.

A number of user characteristics must be taken into account. To start with, people are visual virtuosos. As Donald Hoffman stated (1998): "You can buy a chess machine that beats a Master, but can't yet buy a vision machine that beats a toddler's vision". In visual searching, humans are very good in rapidly finding patterns, recognising objects, generalising or inferring information from limited data, and making relevance decisions. Next, users often do not know exactly what they are looking for; users have difficulty with articulating their needs verbally; users can have problems with direct, precise searching (e.g. keyword search); user's information needs often change during the process of searching; and the search process itself is often as important as the results (Rice et al, 2001).

When interacting with video, users on one hand will benefit a lot from a patch-based, metadata-driven access to video content, and on the other hand from direct access to the data (that is, the video content) for making fast relevance decisions.

3 The application

The application is the outcome of the design process described in the next section. Figure 1 displays the current design of the patch-based video browser. The browser has a number of components, each with its own distinctive colour code.

First of all, "straightforward" interaction with the video (in this case, a newscast) is possible. At the bottom of the application, standard controls for interaction (play, pause, etcetera) are available. The user can rapidly browse the video content by either dragging the slider or by fast-forwarding or fast-backwarding. The results of these data-driven interactions are seen in the player (top-left).

The content of the video is decomposed in patches, which can be considered to be a kind of metadata describing the video content (the actual data). The top-right part of the application provides means for patch selection. A patch can be all video fragments containing news item introductions, all fragments containing interviews, all fragments containing the newsreader, etcetera (whatever is manually created or automatically detected). All these content-based patches are presented in a menu. Patches can also contain the viewing history of the user, time-based samples of the video (e.g., the first 5 seconds of every minute), user-defined favourite fragments, or the results of a query. In the top-most right part, it is possible to query for information related to patch *items* (the video fragments within a patch). This information can be descriptions of the item, and/or the speech within the item (either derived from speech recognition, optical character recognition, or captions).

When the patch list is very long, users can create a menu with a selection of their own "favourite" patches. This menu can also be related to specific content types or user profiles, and be presented at start-up. Regarding user interaction, it does not matter from which menu a patch is selected. The interface is designed to support fast identification of patches - also facilitated by providing descriptions when pointing the mouse at the patch name - and fast movement between patches.

Figure 1. Design of the patch-based video browsing prototype

Currently, patches can be combined using the logical operators OR and AND. This provides a way to enrich a patch. For example, when looking at the patch of actor A, this patch can be enriched by combining it with another patch containing fragments of actor B or a certain location.

There a several ways that the user is supported in browsing within a patch. As can be seen near the bottom, above the name of the patch (combination), patches are visually presented as a collection of bars (the patch items) on a time-line. Information about frequency, distribution, and duration of the items within the patch is directly visible to the user.

An important aspect of the design is that the real data (that is, the video content) can be inspected directly. The patch can be played (using the controls left to the patch name), which means that only those parts of the video that are in the patch (or patch combination) are presented in the top-left player. Additionally, a selection of the patch items is zoomed in on using a slidebar with a flexible size. These selected patch items are displayed above the slidebar (the "zooming window"), and of each item a representative video frame is displayed. Moving the zooming window provides a fast way to visually inspect frames from this patch.

More within-patch browsing is supported by zooming in on patch items. When a patch item is selected (by clicking its related bar or video frame), the top-left player displays a visual summary of the item by presenting nine (time-sampled) frames. Clicking any of the nine frames enlarges it to the size of the player screen. The patch item can also be played, or scanned by dragging a slider (see controls below the player). When available, a description and the speech transcript related to the item is displayed. Using the controls next to the displayed patch name, the user can select the previous or next patch item and very rapidly scan the visual, speech and metadata information connected to each item. A selected patch item is highlighted in all rows by an orange frame. The top-left-middle component of the application is also orange, except when the user views a part of the video that is not within the patch (by using the bottom controls). The user can easily switch back to the “patch mode” by selecting any patch item or by pressing a dedicated button.

An important feature of the interface is that it displays which patches the active patch item (partly) overlaps. This is indicated in the patch menu, where overlapping patches are highlighted. This way, the user is provided with ideas to specify, or “enrich” the current patch by combining it with another.

4 Design approach

An iterative design approach was taken to develop the application. Based on the starting points stated in section 2, an initial video browsing prototype was developed as part of a large, component-based video application, which provides a powerful environment for handling video data and for studying video interaction behaviour. The application presented here is the result of several design changes that resulted from two usability evaluation techniques: heuristic evaluation (Nielsen, 1993) and a cognitive walkthrough (Wharton et al, 1994).

For the tests, a 25-minute newscast was prepared by creating patches related to syntax (e.g., news items, introductions, interviews) and item subjects (e.g., politics, agriculture, sport). The heuristic evaluation was performed by two usability experts, leading to a long list of design adjustments. For the walkthrough, three subjects performed tasks with the application. They clearly indicated they liked the idea of patches. The patches provided useful guidance to browsing the video, and allowed for several browsing strategies to emerge. Combining patches was considered very useful (for example, to find out which sportsman was interviewed, the “Interview” patch could be combined with the “Sport” patch to provide the relevant video fragments).

5 Future research

First results regarding the usability and usefulness of application show that the “patch approach” - as is indicated by information foraging theory - proves to be very promising. Users appear to have a quick grasp of the patch idea. A usability test will be the next step in improving the current application.

The research described here will address two important issues. First, an experiment will be executed to establish what are “good” patches, and how they are related to video content (e.g., genre), user tasks and user characteristics. An important question is whether robust or specialised metadata should be developed. Another experiment will be dedicated to browsing behaviour itself. The concepts and nature of browsing have not yet been systematically studied and are thus not well understood (Rice et al, 2001). The question is which cognitive strategies people apply when browsing. In real-world situations, time, knowledge, and computational capacities are often limited. To successfully deal with these kinds of situations, it is hypothesised that humans apply *fast and frugal heuristics* for decision making (Gigerenzer et al, 1999). The question is which browsing heuristics people apply when interacting with video content.

References

Dimitrova, N., Zhang, H. J., Shahraray, B., Sezan, I., Huang, T., & Zakhor, A. (2002). Applications of video-content analysis and retrieval. *IEEE MultiMedia, 9(3)*, 42-55.

Gigerenzer, G., Todd, P. M., & the ABC Research Group (1999). *Simple heuristics that make us smart.* New York: Oxford University Press.

Hoffman, D. D. (1998). *Visual intelligence.* New York, NY USA: W.W. Norton.

Lee, H. & Smeaton, A.F. (2002). Designing the User Interface for the Físchlár Digital Video Library. *Journal of Digital Information, 2(4).*

Nielsen, J. (1993). *Usability engineering.* San Francisco, CA USA: Morgan Kaufmann Publishers.

Pirolli, P. & Card, S. K. (1999). Information foraging. *Psychological Review, 106,* 643-675.

Rice, R. E., McCreadie, M., & Chang, S.-J. L. (2001). *Accessing and browsing information and communication.* Cambridge, USA: MIT Press.

Wharton, C., Rieman, J., Lewis, C., & Polson, P. (1994). The cognitive walkthrough method: a practitioner’s guide. In Nielsen, J. and Mack, R.L. (Eds.), *Usability Inspection Methods*. John Wiley & Sons, New York, NY.

Yeo, B.-L. & Yeung, M. M. (1997). Retrieving and visualizing video. *Communications of the ACM, 40,* 43-52.