Video Content Foraging

Ynze van Houten[1], Jan Gerrit Schuurman[1], and Pløn Verhagen[2]

[1] Telematica Instituut,
P.O.Box 589, 7500 AN Enschede, The Netherlands
{Ynze.vanHouten, JanGerrit.Schuurman}@telin.nl
[2] Educational Science and Technology, University of Twente,
P.O.Box 217, 7500 AE Enschede, The Netherlands
p.w.verhagen@utwente.nl

Abstract. With information systems, the real design problem is not increased access to information, but greater efficiency in finding useful information. In our approach to video content browsing, we try to match the browsing environment with human information processing structures by applying ideas from information foraging theory. In our prototype, video content is divided into video patches, which are collections of video fragments sharing a certain attribute. Browsing within a patch increases efficient interaction as other video content can be (temporarily) ignored. Links to other patches ("browsing cues") are constantly provided, facilitating users to switch to other patches or to combine patches. When a browsing cue matches a user's goals or interests, this cue carries a "scent" for that user. It is stated that people browse video material by following scent. The prototype is now sufficiently developed for subsequent research on this and other principles of information foraging theory.

1 Introduction

Humans are informavores: organisms that hunger for information about the world and about themselves [1]. The current trend is that more information is made more easily available to more people. However, a wealth of information creates a poverty of directed attention and a need to allocate sought-for information efficiently (Herbert Simon in [2]). The real design problem is not increased access to information, but greater efficiency in finding useful information. An important design objective should be the maximisation of the allocation of human attention to information that will be useful. On a 1997 CHI workshop on Navigation in Electronic Worlds, it was stated [3]: "Navigation is a situated task that frequently and rapidly alternates between discovery and plan-based problem-solving. As such, it is important to understand each of the components of the task – the navigator, the world that is navigated, and the content of that world, but equally important to understand the synergies between them." Information foraging theory [4] can be an important provider of knowledge in this regard, as it describes the information environment and how people purposefully interact with that environment.

In this paper, we describe the design of a video-interaction environment based upon ideas from information foraging theory. The environment is developed to test these ideas in subsequent research, as will be explained at the end of this paper.

P. Enser et al. (Eds.): CIVR 2004, LNCS 3115, pp. 15–23, 2004.

Finding video content for user-defined purposes is not an easy task: video is time-based, making interacting with video cumbersome and time-consuming. There is an urgent need to support the process of efficiently browsing video content. An orderly overview of existing video browsing applications and related issues can be found in [5]. Our approach adds a new perspective in that it applies a human-computer interaction theory to the problem of video content browsing.

Emphasis is on browsing - and not on querying - for a number of reasons. To start with, people are visual virtuosos [6]. In visual searching, humans are very good at rapidly finding patterns, recognising objects, generalising or inferring information from limited data, and making relevance decisions. The human visual system can process images more quickly than text. For instance, searching for a picture of a particular object is faster than searching for the *name* of that object among other words [7]. Given these visual abilities, for media with a strong visual component, users should be able to get quick access to the images. In the case of video, its richness and time-basedness can obstruct fast interaction with the images, so efficient filter and presentation techniques are required to get access to the images.

Except when the information need is well defined and easily articulated in a (keyword) query, browsing is an advantageous searching strategy because in many cases users do not know exactly what they are looking for. Well-defined search criteria often crystallise only in the process of browsing, or initial criteria are altered as new information becomes available. A great deal of information and context is obtained along the browsing path itself, not just at the final page. The search process itself is often as important as the results. Moreover, users can have difficulty with articulating their needs verbally, which especially applies in a multimedia environment, where certain criteria do not lend themselves well to keyword search. Furthermore, appropriate keywords for querying may not be available in the information source, and if they are available, the exact terminology can be unknown to the user [8].

Browsing is a search strategy closer related to “natural” human behaviour than querying [9]. As such, a theory describing natural behaviour in an information environment may be very useful when designing a video browsing environment. Information foraging theory addresses this topic.

2 Information Foraging Theory (IFT)

In this paper, we describe the design of a video browsing environment based upon ideas from information foraging theory [4]. IFT is a “human-information interaction” theory stating that people will try to interact with information in ways that maximise the gain of valuable information per unit cost. Core elements of the theory that we apply are:

- People forage through an information space in search of a piece of information that associates with their goals or interests like animals on the forage for food.
- For the user, the information environment has a “*patchy*” structure (compare patches of berries on berry bushes).
- Within a patch, a person can decide to forage the patch or switch to another patch.
- A strategy will be superior to another if it yields more useful information per unit cost (with cost typically measured in time and effort).

- Users make navigational decisions guided by *scent*, which is a function of the perception of value, cost, and access path of the information with respect to the goal and interest of the user.
- People adapt their scent-following strategies to the flux of information in the environment.

For applying IFT ideas to building an environment that supports efficient video browsing we need patches, and scent-providing links (browsing cues) to those patches. The patches provide structure to the information environment. Patches are expected to be most relevant when patches as defined in the database match with patches as the user would define them. To facilitate the use of patches, users need to be helped to make estimates of the gain they can expect from a specific information patch, and how much it will cost to discover and consume that information. These estimates are based on the user's experience, but also on the information provision of browsing cues.

The concept of scent provides directions to the design of information systems as it drives users' information-seeking behaviour. When people perceive no scent, they should be able to perform a random walk in order to spot a trace of scent. When people perceive a lot of scent, they should be able to follow the trail to the target. When the scent gets low, people should be able to switch to other patches. These types of browsing behaviours all need to be supported in the design. Typical design-related situations can be that browsing cues are misleading (the scent is high, but the target is not relevant/interesting) or badly presented (no or low scent, but a very relevant or interesting target).

2.1 Video Patches

A video can be looked at as a database containing individual video fragments [10]. The original narrative of the video is "only" one out of many ways of organising and relating the individual items. People often want to structure the information environment in their own way, where the "decodings are likely to be different from the encoder's intended meaning" [11].

Video patches are collections of video fragments sharing a certain attribute (see Figure 1). Attributes may vary along many dimensions, including complex human concepts and low-level visual features. Patches can form a hierarchy, and several combinations of attributes can be combined in a patch.

As video fragments can have a number of attributes (e.g., a fragment can contain certain people, certain locations, certain events etc.), the fragments will occur in a number of patches. When viewing a fragment in a patch, links to other patches can be presented to the user. As such, patches form a hyperlinked network.

Patches provide easy mechanisms to filter video content as users can browse a patch and ignore video fragments not belonging to the patch (see also Figure 3). What are good patches depends on issues like the user's task and the video genre. User experiments are needed to establish which patches are useful in which situation.

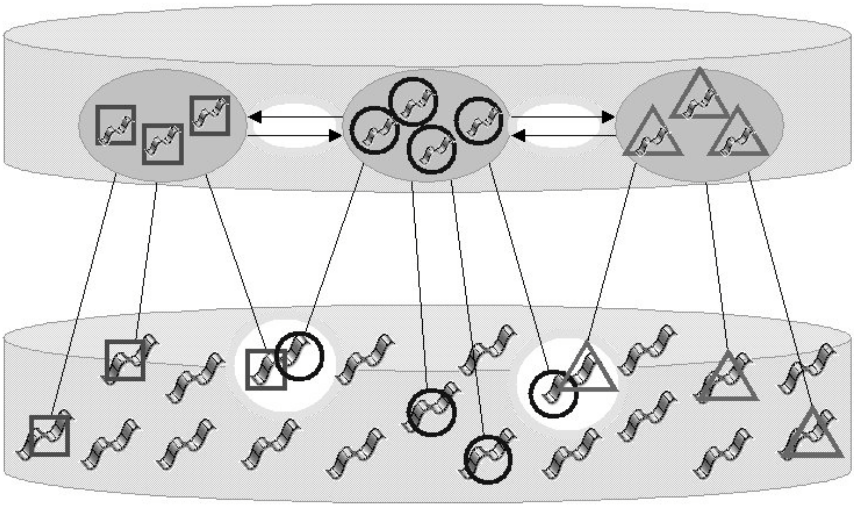


Fig. 1. Representation of video patches. The lower container is a database with video fragments (visualised as filmstrips), which can be the semantic units of a video. Fragments with the same attributes (here: squares, circles, triangles) are combined in video patches. The upper container is the browsing environment containing video patches (here: dark ellipses). When a fragment appears in two or more patches, links between these patches emerge (here: arrows between patches).

2.2 The Scent of a Video Patch

Certain video patches (or items within those patches) can have a semantic match with the user's goal or interests, and as such, give off scent. A user's goal or interest activates a set of chunks in a user's memory, and the perceived browsing cues leading to patches (or patch items) activate another set of chunks. When there is a match, these browsing cues will give off scent. Users can find the relevant patches by following the scent. Scent is wafted backward along hyperlinks – the reverse direction from browsing (see Figure 2). Scent can (or should) be perceivable in the links – or *browsing cues* - towards those targets. Users act on the browsing cues that they perceive as being most semantically similar to the content of their current goal or interest (see also [12]). For example, the scent of menus influences the decision whether users will use them or start a query instead [13].

In terms of IFT, The perception of information scent (via browsing cues) informs the decisions about which items to pursue so as to maximise the information diet of the forager. If the scent is sufficiently strong, the forager will be able to make the informed choice at each decision point. People switch when information scent gets low. If there is no scent, the forager will perform a random walk. What we are interested in is what exactly determines the amount of scent that is perceived.

We have built an experimental video-browsing application (Figure 3) to study a) what are good attributes to create patches with, and b) how to present browsing cues in such a way they can present scent to the user.

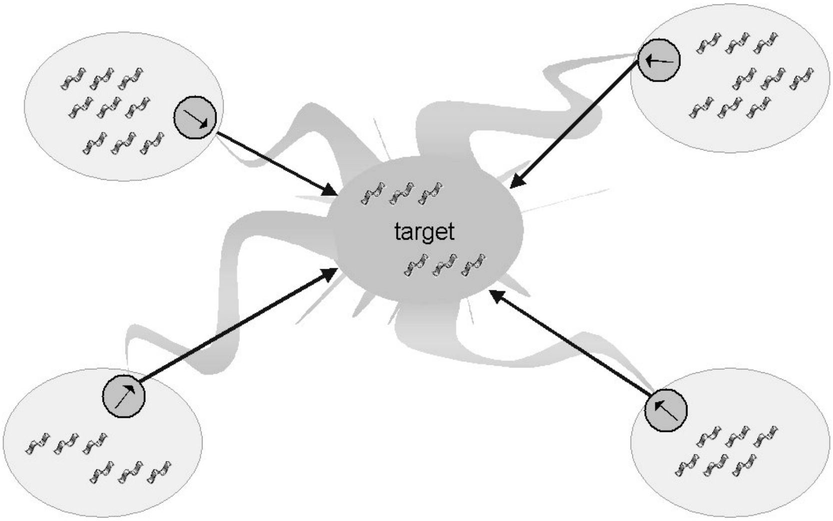
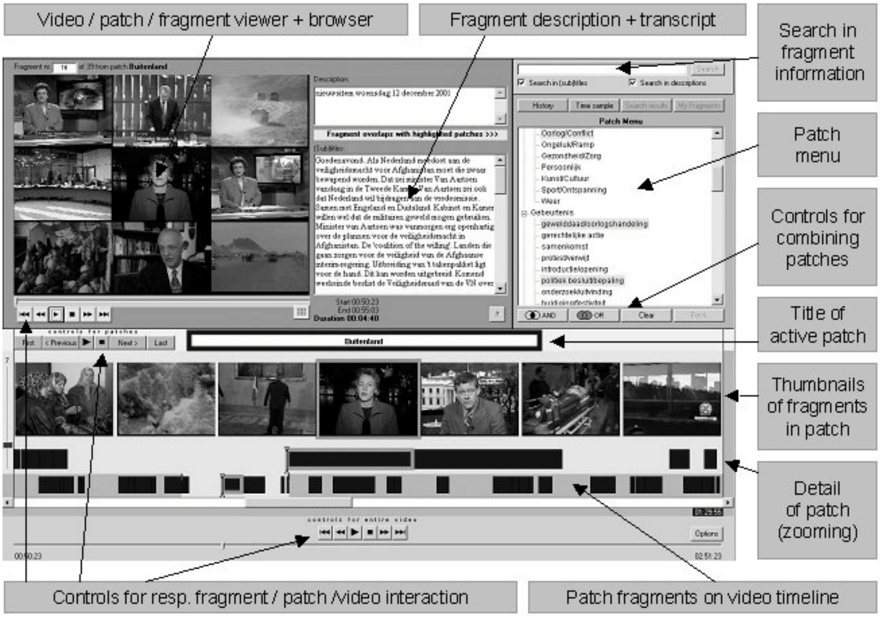


Fig. 2. Schematic illustration of the scent of a video patch. One or more video fragments in the central patch semantically matches with the user's goals or interests, and therefore can be considered a target. The browsing cues in surrounding patches that link to the target carry "scent", which is wafted backwards along the hyperlinks.

3 Design of a Video-Browsing Application

We designed a prototype of a video browsing tool applying the ideas from IFT (see Figure 3). The aim of the prototype is to validate the applicability of IFT for browser design. For that purpose, the video that is browsed needs to be prepared in advance as a set of patches and scent-carrying cues. The richness of semantic cues that have to be taken into account goes beyond the current state of the art in research about automated feature extraction, structure analysis, abstraction, and indexing of video content (see for instance [14], [15], [16], [17]). The automated processes that are technologically feasible cannot yet accomplish the video browsing structure that is cognitively desirable. For the experimental prototype we prepared the video mostly by hand.

In the prototype, scent-based video browsing will not be independent of querying. Smeaton [18] states that for video we need a browse-query-browse interaction. Querying helps to arrive at the level of video patches where the user needs to switch to browsing. Query-initiated browsing [19] is demonstrated to be very fruitful [20]. Golovchinsky [21] found that in a hypertext environment, (dynamic) query-mediated links are as effective as explicit queries, indicating that the difference between queries and links (as can be found – for example – in a menu) is not always significant. In the application described here, we see that the user can query the database and actually is browsing prefabricated queries.



nieuwsitem woensdag 12 december 2001

Goedenavond. Als Nederland meedoet aan de veiligheidsmacht voor Afghanistan moet die zwaar bewapend worden. Dat zei minister Van Aartsen vandaag in de Tweede Kamer. Van Aartsen zei ook dat Nederland wil bijdragen aan de vredesmissie. Samen met Engeland en Duitsland. Kabinet en Kamer willen wel dat de militairen geweld mogen gebruiken. Minister van Aartsen was vanmorgen erg openhartig over de plannen voor de veiligheidsmacht in Afghanistan. De 'coalition of the willing'. Landen die gaan zorgen voor de veiligheid van de Afghaanse interim-regering. Uitbreiding van 't takenpakket ligt voor de hand. Dit kan worden uitgebreid. Komend weekeinde besluit de Veiligheidsraad van de VN over

Fig. 3. Interface of the patch-based video browsing application. Users can start interacting with the video by a) selecting a patch from the patch menu, b) querying video fragment information after which results are presented as a patch, or c) simply playing the video. When a patch is selected, the patch fragments - represented by boxes - are displayed on the video timeline, thus displaying frequency, duration, and distribution information of the fragments in the patch. To get more focussed information, A part of the patch is zoomed in on, and for these items keyframes are presented. Always one fragment in the patch is "active". For this item, the top-left part of the interface presents 9 frames (when the item is not played or scrolled using the slidebar). All available information about the fragment is displayed (transcript, text on screen, descriptions). For the activated fragment it is shown in which other patches it appears by highlighting those patches in the patch menu.

3.1 Patch Creation and Presentation

When browsing starts and the user wants to select a first patch from the menu, the patch with the highest scent will probably get selected ("this is where I will probably find something relevant/of interest"). Which types of patches are created and the way they are presented (labelled) is here the central issue.

In order to create patches, the video is first segmented into semantically meaningful fragments, which will become the patch items in the video patches. What are relevant units is genre-dependent. For a newscast, the newsitems seem to be the best units. For an interview, it may be each new question or group of related questions. For a football game, it may be the time between pre-defined "major events" (e.g. goals, cards, etc.).

For each video unit, attributes are defined (in other words, what is *in* this fragment? What is this fragment about?). This step defines the type (or theme) of the video patches, and as such, the main entries in the patch menu that will be formed. User inquiries are needed to find out what are the most useful attributes. For a football match, it may be players, goals, shots on goal, cards etc. For a newscast, this may be the news category, persons, locations, events, etc.

Next, attribute values need to be defined (so, not just whether there is a person, but also who *is* that person). This step defines the specific patches that will be used in the browsing process. The main question here is which values will be most useful, which will depend on contexts of use. This is a typical example why much is done by hand: automatic feature extraction techniques currently have great difficulties performing this task. Data from closed captions, speech recognition, and OCR (optical character recognition), however, proved to be most helpful.

Fragments sharing the same attribute values are combined into patches, and these patches are indexed. All this information about the video is stored in MPEG7 [22].

3.2 Support for Different Types of Browsing Behaviour

As noted earlier, we distinguish three types of browsing behaviour: a random walk, within-patch browsing, and switching (between-patches browsing).

For a random walk, users can play or browse the video as a whole, or browse the patch containing all video fragments (thus getting information for each fragment).

For within-patch browsing, users can simply play the patch "hands-free": at the end of each fragment the video jumps to the start of the next fragment in the patch. Alternatively, users can move the zooming window (using a slidebar) to scan the thumbnails representing the fragments in the patch. By clicking a thumbnail (or the neutral boxes representing a fragment), the user "activates" a fragment, thus displaying a lot of detailed information about that fragment. Users can use "next" and "previous" buttons to easily see detailed information about other individual fragments. As such, users can quickly scan within a patch what is or is not relevant (that is, what does and does not associate with their goals or interests).

For every patch item, it is shown in which other patches it appears. Highlighting those patches in the patch menu indicates this. This also provides metadata about the current item. For example, when watching an item from the patch "politicians", the item "drugs" may be highlighted in the menu, indicating what issue a politician is referring to. When the user is interested in video fragments on drugs, the user can simply switch to that patch. When the user is interested in opinions of politicians about drugs, the user may combine the two patches using the logical operator AND (or OR, if the user is interested in both). This way, users can influence the structure of the information environment in a way IFT calls "enrichment" [4]. Of course, users can always switch to any other patch in the patch menu.

When the user wants to switch, the interface needs to present possibilities (that is, browsing cues) to switch to other patches. Reasons people want to switch may occur when: a) the current source has a scent below a certain threshold, or b) the user is inspired by what is seen within the current source ("I want to see more like this!"), or c) the browsing cues simply give off a lot of scent (cues point directly to sought-for information). If necessary, browsing cues can be presented dynamically, that is, query- or profile-dependent (see also [19]).

3.3 Scent Presented in the Interface

When the user is browsing video data, the current patch or fragment will give off a certain scent via so-called “scent carriers”. Assuming that people are scent-followers, the main question here is: How can we present scent in such a way that users can efficiently and effectively browse video material?

The title of the patch is the first scent carrier the user is confronted with. When the patch name semantically matches the user’s goals or interest, it will carry a certain amount of scent. The way patch names are organised and presented will influence the amount of perceived scent.

When a patch is activated, several indicators will provide more or less scent, indicating to the user that “I’m on the right track”, or “I need to switch to another patch”. First of all, the frequency, duration, and distribution information of the fragments in the patch can provide a lot of scent (for example, when looking for the main people involved in a story, frequency information may be very useful). Still frames representing the fragments in the patch are the next scent carrying cues. For the one active fragment, a lot of scent carriers are available: keyframes (in this case, nine), transcript (derived from closed captions [when available] and/or speech recognition), displayed text in the video (optical character recognition), and added description. Of course, the video images– that can be either viewed or browsed by fast-forwarding or using a slider - can also carry scent by themselves.

Regarding switching to other patches, the way indications about overlapping patches are presented will influence the scent people perceive.

4 Conclusions and Future Work

The practical problem we try to deal with is how people can interact with video content in such a way that they can efficiently pursue their goals. We translate this to the research problem of how we can match the information environment with human information processing structures. Information foraging theory is assumed to be a fitting theory to answer this question as it both describes how people perceive and structure the information environment, and how people navigate through this environment. Our prototypical browsing environment is based on the principles of this theory. Hypotheses we derive from the theory is that humans perceive the information environment as “patchy”, humans navigate through the information environment by following scent, and humans will interact with the environment in ways that maximise the gain of valuable information per unit cost. We applied these ideas to construct a solution for efficient interaction with video content, as described in this paper. The actual development took place in a few steps applying an iterative design apporach [23]. This is the starting point for our future research on the applicability of information foraging theory for the design of video interaction applications. As a first step we plan experiments to study the effectiveness of different presentations of browsing cues, how the perception of scent relates to different types of user tasks, how to support patch navigation, and which patches are useful in which situations.

References

1. Miller, G. A. (1983). Informavores. In F. Machlup & U. Mansfield (Eds.), *The Study of Information: Interdisciplinary Messages* (pp. 111-113). New York: Wiley.
2. Varian, H. R. (1995). The Information Economy - How much will two bits be worth in the digital marketplace? *Scientific American, September 1995*, 200-201.
3. Jul, S. & Furnas, G. W. (1997). Navigation in electronic worlds: a CHI'97 workshop. *SIGCHI* Bulletin, 29, 44-49.
4. Pirolli, P. & Card, S. K. (1999). Information foraging. *Psychological Review, 106*, 643-675.
5. Lee, H. & Smeaton, A. F. (2002). Designing the user interface for the Físchlár Digital Video Library. *Journal of Digital Information, 2.*
6. Hoffman, D. D. (1998). *Visual intelligence.* New York, NY USA: W.W. Norton.
7. Paivio, A. (1974). Pictures and Words in Visual Search. *Memory & Cognition, 2,* 515-521.
8. Rice, R. E., McCreadie, M., & Chang, S.-J. L. (2001). *Accessing and browsing information and communication.* Cambridge, USA: MIT Press.
9. Marchionini, G. (1995). *Information seeking in electronic environments.* Cambridge University Press.
10. Manovich, L. (2001). *The language of new media*. Cambridge, MA: The MIT Press.
11. Hall, S. (1980). Encoding/Decoding. In *Culture, Media, Language: Working Papers in Cultural Studies 1972-1979*. London: Hutchinson.
12. Blackmon, M. H., Polson, P. G., Kitajima, M., & Lewis, C. (2002). Cognitive Walkthrough for the Web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '02* (pp. 463-470). Minneapolis, Minnesota, USA: ACM.
13. Katz, M. A. & Byrne, M. D. (2003). Effects of Scent and Breadth on Use of Site-Specific Search on E-commerce Web Sites. *ACM Transactions on Computer-Human Interaction*, *10,* 198-220.
14. Yeo, B.-L. & Yeung, M. M. (1997). Retrieving and visualizing video. *Communications of the ACM, 40,* 43-52
15. 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.
16. Wactlar, H. D. (2000). Informedia - search and summarization in the video medium. In *Proceedings of the Imagina 2000 Conference.*
17. Snoek, C. G. M., & Worring, M. (2002). A Review on Multimodal Video Indexing. In *Proceedings of the IEEE Conference on Multimedia & Expo (ICME).*
18. Smeaton, A. F. (2001). Indexing, browsing and searching of digital video and digital audio information. In M. Agosti, F. Crestani, & G. Pasi (Eds.), *Lectures on information retrieval.* Springer Verlag.
19. Furnas, G. W. (1997). Effective view navigation. In *Proceedings of the conference on Human Factors in Computing Systems, CHI '97* (pp. 367-374). Atlanta, GA, USA: ACM.
20. Olston, C. & Chi, E. H. (2003). ScentTrails: Integrating Browsing and Searching on the Web. *ACM Transactions on Computer-Human Interaction, 10*, 177-197.
21. Golovchinsky, G. (1997). Queries? Links? Is there a difference? In *Proceedings of the Conference on Human Factors in Computing Systems, CHI '97* (pp. 407-414). Atlanta, GA, USA: ACM.
22. Nack, F. & Lindsay, A.T. (1999). Everything you wanted to know about MPEG-7: Part 1. *IEEE MultiMedia, 6(3)*, 65-77.
23. van Houten, Y., van Setten, M., & Schuurman, J. G. (2003). Patch-based video browsing. In M. Rauterberg, M. Menozzi, & J. Wesson (Eds.), *Human-Computer Interaction INTERACT '03* (pp. 932-935). IOS Press.