IDEA GROUP PUBLISHING

701 E. Chocolate Avenue, Hershey PA 17033, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

ITP4387

Principles for supporting and enhancing user navigation of digital video in video browsers

HARRY W. AGIUS
Department of Information Systems and Computing, Brunel University,
Uxbridge, Middlesex UB8 3PH, UK; Voice: +44 (0)1895 816222, Fax: +44 (0)1895 251686,
E-mail: harryagius@acm.org, Web: http://www.brunel.ac.uk/~cssthwa/

ABSTRACT

Video browsers provide an environment in which a user may navigate digital video content. They are therefore an important application for improving access to digital video resources. We apply a general framework for navigation, proposed by Spence [1], and consider existing proposals for video browsers made within the research literature within the context of the framework. We then use this analysis to derive a number of principles that reflect the requirements for supporting and enhancing the user in the digital video navigation task.

1. INTRODUCTION

Digital video is no longer a novelty data type and its uses are becoming plentiful, diverse, and central to current and future computing environments. Although originally available primarily in restricted low-resolution formats on CD-ROMs, digital video now forms the basis for a host of applications including DVDs, video-on-demand, media streaming applications over the Web, digital television, and applications running on newly-emerging mobile devices, such as Pocket PCs. However, despite this abundance, effective user access to digital video resources is proving problematic.

Video browsing is one area of research that offers promise in improving access to digital video [2-11]. Video browsers assist the user by enabling them to navigate through digital video, e.g. by limiting the amount of content that is presented and by providing overviews of entire digital videos. However, not only have few video browsing applications been proposed, but moreover there has been little research on how user navigation, rather than browsing, of digital video should be supported in such systems.

This paper proposes a number of principles for supporting and enhancing user navigation of digital video in video browsers. The principles are derived through the application of a navigation framework proposed in the research literature by Spence [1]. The structure of the remainder of this paper is as follows. Section 2 provides a concise review of video browsers. Section 3 presents Spence's general framework for navigation. Section 4 applies this framework to digital video navigation and considers existing video browsers within the context of the framework. Section 5 then presents a number of principles for supporting and enhancing user navigation of digital video that stem from the application of the framework. Section 6 concludes the paper.

2. VIDEO BROWSERS

The objective of a video browser is to enable a user to navigate through a digital video efficiently and effectively. In other words, a video browser may be considered a media-specific type of browser. Video browsers may exist in two forms:

· **For explicit query support**: Given a user-formulated query (e.g. expressed in a database query language), a video search and retrieval system presents candidate digital video segments to the user, i.e. video segments that meet the user's query criteria. The user should then be able to navigate through these candidate video segments to ascertain their suitability and/or to determine whether the query criteria need refinement [2].

· **For implicit query support**: In this case, navigation occurs as a result of implicit querying. The user has an internal query that is never made explicit (e.g. it is not expressed via a database query language). Instead, the user seeks to satisfy their query by searching for a digital video segment of interest by navigating 'at leisure'.

The conventional approach to digital video navigation is that provided by standard media players, such as Real Networks' RealPlayer [12], Microsoft's Windows Media Player [13], and Apple's QuickTime Player [14]. Such players provide conventional VCR-like controls (fast forward and rewind) and also step forward and step backward controls (to jump a uniform number of frames per 'step' of the video). The main controls in Microsoft's Windows Media Player are shown in Figure 1. As will be elaborated upon in Section 4, these controls limit greatly the ability of the user to navigate efficiently and effectively.

Figure 1. Microsoft's Windows Media Player (version 7, default skin).

Because of such limitations, a number of approaches have been proposed with the aim of improving user navigation of digital video. These include:

· Video abstraction: This approach generates a sequence of representative key frames within a digital video segment, which are selected using a variety of automatic processing techniques [3, 9, 10]. For example, analysing camera work and selecting a representative video frame for each shot within the video. The generated sequence represents a video abstract which summarises the video and thereby assists the user in browsing and locating pertinent random-access points.

· Spatial and temporal zoomers: Here, multiple low-resolution versions of a digital video segment are used, which the user can then selectively zoom into in space and time, thereby increasing the resolution for specific segments [5, 15]. Thus, a digital video may be navigated at numerous hierarchical levels.

· Video mosaics: A video mosaic [11] synthesises into a single image a representation of the visual contents of an entire digital video segment. One example is the *micon* (movie icon) [8], which places each video frame directly behind the previous frame to form a single 3D visual cube. Browsing then takes place by the user examining vertical or horizontal 'slices' through the micon. *Salient stills* [7] and *video space icons* [9] are similar variations of this.

In the following section, we consider the concept of navigation itself with a view to better understanding and analysing digital video navigation.

1. SPENCE'S FRAMEWORK FOR NAVIGATION

Spence [1] proposes a general framework for navigation that is applicable to navigation activities in general, i.e. both computing- and non-computing-related. The framework is illustrated in Figure 2.

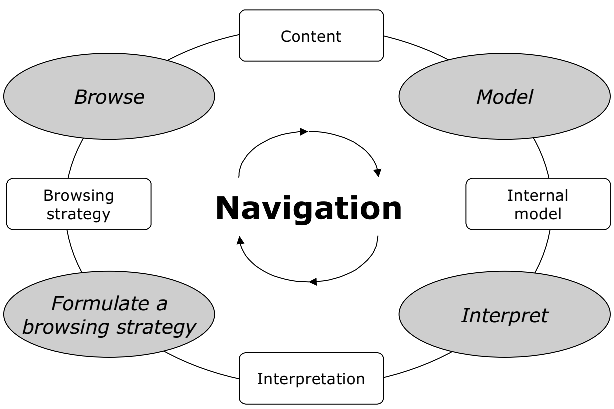

Figure 2. The general framework for navigation proposed by Spence [1].

As can be seen, navigation within this framework is an iterative process that consists of four interconnected activities [1]:

· Browse: This is the registration of content by the user and takes place according to a browsing strategy. Browsing depends upon an externalisation of content being present. More expert browsers tend to exhibit *weighted browsing*, whereby the nature of the content to be registered is selected based upon, for example, anticipated use or value.

· Model: This is where the user constructs an internal model (cognitive map) of the browsed content. This model is highly influenced both by the nature in which the content is externalised and by the user interface *affordances* that permit the externalised content to be rearranged according to the user's needs. Note that an internal model may decay due to lack of use, be refreshed by regular viewing of externalised content, or be extended.

· Interpret: Here, the internal model, together with other (external) data, are interpreted with the view of deciding if and how navigation should proceed and/or if and how the original task should be modified.

· Formulate a browsing strategy: Here, the user (consciously or unconsciously) decides on how browsing should take place. Browsing strategies may be *planned* (chosen and followed, perhaps according to weights) or *opportunistic* (based on stimuli in content, perhaps without weights). The strategies will be influenced by both *cognitive* determinants (interpretation or new ideas) and *perceptual* determinants (externalised content). The formulated browsing strategy must be supportable by the user interface affordances that permit user navigation with the content.

In this framework, browsing is one activity within the entire navigation process. The framework is therefore useful for considering video navigation within a wider context. However, we will continue to use the term 'video browser' when referring to navigational video applications since this term is now used commonly within the research literature. At the same time, however, we will maintain a distinction between the activities of browsing and navigation.

In the following section, we apply the above framework to digital video navigation to reveal how the user may be supported and enhanced in this task by video browsers.

1. APPLICATION OF THE FRAMEWORK TO DIGITAL VIDEO NAVIGATION

By considering digital video navigation as a content navigation activity like any other, we are able to apply Spence's framework for analysis. Our discussion here will be used in Section 5 to derive a number of principles for supporting and enhancing user navigation of digital video in video browsers.

1.1. Browsing of video content

The browsing activity within navigation requires content to be externalised so that it may be registered by the user. For video content play back, the externalisation traditionally takes place through a media player, i.e. through a 2D video window with any associated audio played back in synchrony. This approach provides no overview or summary of the video content and therefore forces the user to view the entire video sequentially. Thus, when browsing, the user is not able to register the content of the entire digital video segment rapidly and little (if any) support exists for weighted browsing. This is particularly the case where the user is navigating to find a digital video that they have never viewed before or when they are browsing opportunistically. In these cases, the user may be lacking completely (and unable to develop) any internal, consciously selected, weighting criteria to assist in their browsing activity. The lack of support therefore has a damaging impact not only on the browsing activity but also on the other navigation activities.

Consequently, it is vital that superior forms of externalisation are sought and used to support and enhance the user's browsing activity. Video abstraction, spatial and temporal zoomers, and video mosaics (discussed in Section 2) provide increased opportunities for rapid registration of content by users. Video abstracts and video mosaics both provide means by which an entire video may be rapidly 'grasped' in its entirety, almost in a single instant. Similarly, zoomers enable the depth of the view to be increased or decreased as necessary, providing users with some opportunity to grasp the 'big picture'.

Moreover, further enhancing the externalisation so that more semantically pertinent segments within the video are highlighted provides increased support for weighted browsing. This might take place by increasing the relative resolution of key frames from these segments, so that they appear larger on screen, for example.

1.1. Modelling of video content

The internal model that the user constructs of the browsed video content will be highly influenced by the types of externalisations provided and the user interface affordances that enable the externalisations to be manipulated by the user. It is therefore crucial that both of these are aligned with the internal model that is most relevant to the navigation task being performed. For example, one distinction can be made between navigation based on implicit and explicit queries. Clearly, these are two navigation tasks that have sufficiently different objectives to warrant different modelling and affordance criteria.

The topological nature, or 'connectedness', of the content is particularly important here, since externalisations that reflect this topology serve to assist the user in the creation of their internal model [1]. With video content, this topology arises from the interaction of many complex semantics, such as events, actions, and objects [16]. These interactions give rise to further levels of complex semantics such as plot lines [17]. It is these semantics that form much of the user's internal model and therefore must be expressed within the externalisation.

Within conventional media players, frames are presented sequentially and all previous frames and segments must be remembered by the user. Therefore, there is no topological expression of the video content other than the relationship between consecutive frames. Thus, the only topology expressed to the user in media players is that provided by the standard *video slider* that depicts the current frame and its position relative to the entire video duration (see Figure 1).

However, at the same time, within the video browser approaches presented earlier (Section 2), the overriding emphasis is on non-semantic, low-level content features, such as camera angles, colours, and segment start/finish points (shots and scenes). While this undoubtedly forms some part of the internal model, and is therefore required to some extent, it forms a much smaller part relative to the semantic features.

It is also necessary to be cautious with the amount of externalisation of video content given to the user. Too little and the user will be unsupported; too much and the user will be overwhelmed. This is illustrated in Figure 3. The most appropriate amount of externalisation is illustrated by the *point of ideal externalisation* within the figure. Unfortunately, this can be difficult to determine, and will vary between users.

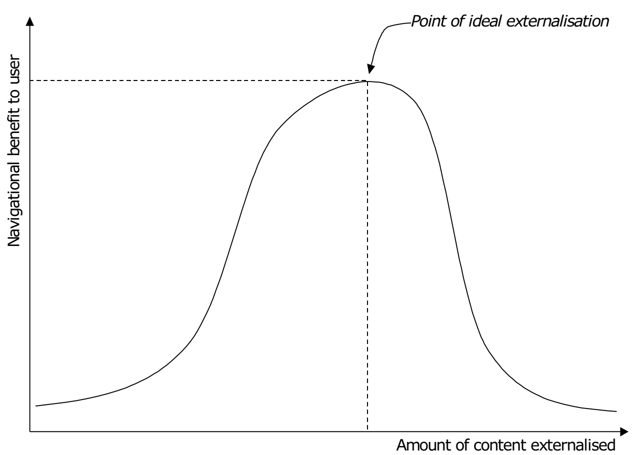

Figure 3. Relationship between amount of externalised content and user navigational benefits.

As mentioned in Section 3, the user's model may decay over time if it is not refreshed. Support is therefore also required for *maintaining* the internal model [1]. Tonomura et al. [9] have developed a three-stage model that reflects this process, illustrated in Figure 4. Video abstracts are particularly suitable for this purpose since they enable speedy refreshment of the user's existing knowledge of the video content and help to push them someway from the rough idea stage towards the full idea stage (much in the same way as traditional document abstracts do). Since much of the user's model will be semantic, the more semantic such video abstracts are, the more beneficial the user will find them. When these video abstracts are integrated with user interface affordances such that they are also used to display feedback on current location, they can serve as a very rich method by which to maintain the user's internal model.

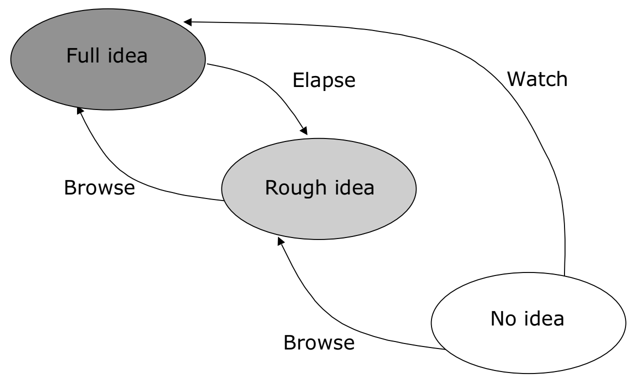

Figure 4 Tonomura et al.'s [9] model depicting changes in the user's internal model.

In distributed environments, such as the Web, the risk of decay is more prominent because of common delays that the user will experience that will interfere with their modelling task as it is being carried out. Thus, the user must either wait for each digital video to be downloaded before they can navigate them, or, if media streaming is used, they must wait for each digital video to be buffered and then re-buffered as they browse through the video [5]. The use of a video abstract or spatial/temporal zoomer can assist here since these types of browsers place less demands on the network than moving through the entire digital video in a conventional media player.

1.1. Interpretation of video content

In this activity, the user must make a decision on how to proceed, based on their internal model and current externalisations. Thus, it is important that the user is able to ascertain their current location and is aware of other navigational paths that may be deemed appropriate.

The use of video abstracts that are integrated with user interface affordances, which were mentioned in the previous section, can be particularly beneficial since not only do they seek to help maintain the user's internal model but they also provide a useful means by which the user is assisted in their interpretation. This is because these types of video abstracts enable traces of remote but potential navigational 'targets' to be depicted. The user may then decide to navigate to these targets and explore them further.

By the same criteria, video mosaics and spatial and temporal zoomers also help support interpretation. With the video mosaic, the synthesis of the entire digital video within a single mosaic provides traces by depicting motion or other patterns within the video. For example, the sides of the micon depict camera and object movement, shot and scene changes, colour changes, and other motion. These can serve to provide important visual clues to the user as to where they should navigate next. With spatial and temporal zoomers, lower resolution (in space and/or time) also serves to provide important traces of potential targets which the user is then able to zoom into and explore further if interpretation concludes this.

However, as was the case previously, during interpretation the user will be seeking more semantic rather than lower level information. Therefore, the provision of semantically relevant traces of potential targets greatly assists the user here.

1.2. Formulation of a video browsing strategy

Browsing strategies are supported by user interface affordances and externalisations of content. Together they define (restrict) how navigation may take place and therefore what browsing strategies may legitimately be followed.

The standard VCR-like controls provided by media players are inappropriate because the user may only forward and reverse linearly, without reference to the topology of the video content. Since little support is offered, this can be a lengthy process for the user. Moreover, the fast forward and rewind functions are more limited than they are on a conventional VCR. On a VCR, during fast forward and rewind the video is still displayed on screen but at a faster playback speed. Thus, the user at least has some opportunity to register the content as it speeds past. However, in media players, fast forwarding and rewinding take place at tremendous speed (frame jumping is almost instantaneous) giving the user limited opportunity to register the video content as it flies past. Similarly, the step forward and step backward functions within media players do not display the frames occurring between each step.

In DVDs, browsing is enhanced a little through the use of data structures that permit the user to 'jump' to new digital video segments via a graphical menu [18]. While this enables some form of topology to be supported, this still does little to enhance the user's navigation task, particularly in the case of explicit queries where it is likely that many candidate digital video segments will be returned whose individual and aggregate duration will be lengthy (particularly when entire programmes or movies are returned).

Nevertheless, providing alternative browsing paradigms has proved difficult thus far. The main proposals that have been made in the research literature were discussed in Section 2. The use of *hypervideo* [19], which combines hyperlinking of related digital video segments, is perhaps one technology that offers promise in supporting browsing strategies, particularly when it is combined with effective externalisations and user interface affordances.

2. PRINCIPLES FOR SUPPORTING AND ENHANCING USER NAVIGATION OF DIGITAL VIDEO

Based on our analysis of digital video navigation within the context of Spence's framework, we now propose a number of principles that reflect the user's requirements in this task:

1. **The Alignment Principle**: The video browser should be aligned with the nature of the task, i.e. for implicit or explicit querying. This must be reflected in the type of externalisation and user interface affordances used.
2. **The Topology Principle**: Both the externalisation and the user interface affordances must reflect the topology of the video content. Furthermore, the two must be tightly integrated. The use of information visualisation techniques [20] can be used to fulfill this requirement, e.g. visualising the digital video content in a tree-like structure.
3. **The Ideal Externalisation Principle**: The interface should be customisable for each individual user so that the ideal amount of externalisation of content is utilised and is adaptable dynamically. This includes the use of tools for changing and presenting appropriate levels of detail, such as video abstractions and spatial and temporal zoomers.
4. **The Traces Principle**: The user should be assisted in their interpretation activity through the provision of pertinent traces of remote but potentially beneficial targets.
5. **The Semantics Principle**: Semantic content should always take precedence over lower-level content, since it is semantic content that will form the major basis of the user's internal model. The use of content-modelling techniques [16, 21] can be used to meet this requirement.
6. **The Model Maintenance Principle**: The user should be provided with the means to quickly and effectively refresh the knowledge they already have of the content.
7. **The Location Principle**: The user should be provided with feedback on their relative and absolute locations.
8. **The Limited Delays Principle**: The user should be provided with timely feedback and delays in feedback must not affect the user's internal modelling task.
9. **The Metaphor Principle**: All of the above principles should be embodied in and supported by a user interface metaphor other than the VCR metaphor. This may bring the user 'closer' to the video content and enable them to navigate more naturally and with fewer difficulties via the video browser [22].

A video browser that accommodates all of the above principles would therefore not only support the user in their navigation of digital video but also greatly enhance their task. We view them therefore as 'guiding principles' for the development of future video browser applications.

3. CONCLUDING REMARKS

While video browsers have emerged as a promising technology for improving navigation of digital video and offer more support than conventional media players, they are still limited in the support that they provide and do little to enhance the navigational task for the user. By applying a general framework for navigation, as proposed by Spence [1], we have considered the support that existing video browsers provide to the user and have derived a number of principles that reflect how the user may be supported and enhanced in their digital video navigation task. Accommodation of these principles is therefore significant for improving video browsing applications in the future.

REFERENCES

1. R. Spence, 'A framework for navigation', *International Journal of Human-Computer Studies*, vol. 51, 1999, pp. 919-945.
2. R.M. Bolle, B.-L. Yeo, and M.M. Yeung, 'Video query: research directions', *IBM Journal of Research & Development*, vol. 42, no. 2, 1998, pp. 233-252.
3. A.M. Ferman and A.M. Tekalp, 'Efficient filtering and clustering methods for temporal video segmentation and visual summarization', *Journal of Visual Communication and Image Representation*, vol. 9, no. 4, 1998, pp. 336-351.
4. R. Lienhart, S. Pfeiffer, and W. Effelsberg, 'Video abstracting', *Communications of the ACM*, vol. 40, no. 12, 1997, pp. 55-62.
5. J.R. Smith, 'VideoZoom spatio-temporal video browser', *IEEE Transactions on Multimedia*, vol. 1, no. 2, 1999, pp. 157-171.
6. S.W. Smoliar and H.J. Zhang, 'Content-based video indexing and retrieval', *IEEE MultiMedia*, vol. 1, no. 2, 1994, pp. 62-72.
7. L. Teodosio and W. Bender, 'Salient video stills: content

and context preserved', in *Proceedings of ACM Multimedia '93*, 1993, pp. 39-46.

8. Y. Tonomura, 'Video handling based on structured information for hypermedia systems', in *Proceedings of ACM Multimedia '91*, 1991, ACM Press / Addison-Wesley, New York, NY, pp. 333-344.

9. Y. Tonomura, A. Akutsu, Y. Taniguchi, and G. Suzuki, 'Structured video computing', *IEEE MultiMedia*, vol. 1, no. 3, 1994, pp. 34-43.

10. S. Uchihashi, J. Foote, A. Girgensohn, and J. Boreczky, 'Video Manga: generating semantically meaningful video summaries', in *Proceedings of ACM Multimedia '99*, 1999, ACM Press / Addison-Wesley, New York, NY, pp. 383-392.

11. H.J. Zhang, 'Content-based video browsing and retrieval', in *Handbook of Multimedia Computing*, B. Furht (ed.), CRC Press, Boca Raton, FL, 1999, pp. 255-280

12. Real Networks Web site, <http://www.real.com/>

13. Windows Media Player Web site, <http://www.microsoft.com/mediaplayer/>

14. QuickTime Web site, <http://www.apple.com/quicktime/>

15. M. Mills, J. Cohen, and Y.Y. Wong, 'A magnifier tool for video data', in *Proceedings of ACM CHI '92*, 1992, ACM Press, New York, NY, pp. 93-98.

16. H.W. Agius and M.C. Angelides, 'A method for developing interactive multimedia from their semantic content', *Data & Knowledge Engineering*, vol. 34, 2000, pp. 165-187.

17. B. Laurel, *Computers as Theatre*, Addison-Wesley, Reading, MA, 1993.

18. J. Taylor, 'DVD-Video: multimedia for the masses', *IEEE MultiMedia*, vol. 6, no. 3, July-September 1999, pp. 86-92.

19. N. Sawhney, D. Balcom, and I. Smith, 'Authoring and navigating video in space and time', *IEEE MultiMedia*, vol. 4, no. 4, 1997, pp. 30-39.

20. J.D. Hollan, B.B. Bederson, and J.I. Helfman, 'Information visualisation', in Handbook of Human-Computer Interaction, *Second, completely revised edition, M.G. Helander, T.K. Landauer, and P.V. Prabhu (eds.), Elsevier, Amsterdam, 1997, pp. 33-48*

21. S. Marcus and V.S. Subrahmanian, 'Foundations of multimedia database systems', *Journal of the ACM*, vol. 43, no. 3, 1996, pp. 474-523.

22. D.C. Neale and J.M. Carroll, 'The role of metaphors in user interface design', in *Handbook of Human-Computer Interaction*, Second, completely revised edition, M.G. Helander, T.K. Landauer, and P.V. Prabhu (eds.), Elsevier, Amsterdam, 1997, pp. 441-462