User Interface Issues for Browsing Digital Video

Hyowon Lee

School of Computer Applications, Dublin City University
Glasnevin, Dublin 9, Ireland
hlee@compapp.dcu.ie

Alan F. Smeaton

School of Computer Applications, Dublin City University
Glasnevin, Dublin 9, Ireland
asmeaton@compapp.dcu.ie

Jonathan Furner

Department of Library & Information Science, University of California Los Angeles
Los Angeles, CA 90095-1520
jfurner@ucla.edu

Abstract

In this paper we examine a suite of systems for content-based indexing and browsing of digital video and we identify a superset of features and functions which are provided by these systems. From our classification of these we have identified that common to all is the fact of being predominantly technology-based, with little attention paid to actual user requirements. As part of our work we are developing an application for content-based browsing of digital video which will incorporate the most desirable but achievable of the functions of other systems. This will be achieved via a series of continuously refined demonstrator systems from Spring 1999 onwards which will be subjected to analysis of performance in terms of user.

1   Introduction

The user interface to an IR system is the part of the system that bridges the system's functionality with the users' requirements of the system. However strong an IR system's internal retrieval mechanism is, if it is not possible for users to make use of this then communication between the user and the system does not effectively work out and the overall system fails. Until recently consideration of the UI in the design of an IR system has often been regarded as unimportant, a surface look to make the system attractive. But now more emphasis is beginning to be given to UI design and there is more realisation that UI concerns should be integrated well into the system development process, and that when designing UI we have to give sufficient thought on the users to which the system is targeted and also thought on the nature of the particular medium in concern.

The medium of video is characterised by its multipleness of information (visual + audio + textual) and its temporal basis. However, there is as yet no comprehensive work on a study of interface designs specifically for content-based access to the video medium taken in the context of an entire video IR system. Work available is generally the more isolated, narrower concerns such as how to provide efficient browsing/viewing of video sequences such as Videoline [1], Video Streamer [2], hierarchical video magnifier [3], and the Siemens browser [4]. These studies mention the usual problem statements that generally an entire video sequence is too big in size for the network environment and that the traditional VCR-style playback of video sequences takes too long and too much effort for the user to locate wanted scenes.

At present, there is no good theoretical basis established on general UI design for any medium, and the situation is worse with the UI design for the particular medium called video. The question we are concerned with in our work is can we design a user-interface that provides an efficient and easy way for searching for a video clip ? What we would like in an ideal world is study of applying general UI design concerns to a video IR interface. In practice, our approach is to survey and categorise existing and emerging UI features for searching video and from these to engineer a system which implements the most appropriate of these for an application. Through a series of user studies on this operational system we will validate and continuously refine our choices. In the following section of this paper we discuss some general principles of user interfaces to video IR systems and in section 3 we outline some details of commercial and research systems which are currently available. Section 4 contains a categorisation

of the features provided by the currently-available systems and this is the chief contribution of the paper. In section 5 we sketch our plans for progressing our own work and in section 6 we conclude the paper.

2   User-Interfaces for Video IR System

For any development of a completely new kind of system, a user-study is required to find out the target system users' needs and develop the system accordingly. Knowing about users and their requirements becomes more important for the UI design part, as this is what the users will actually see and interact with the system. In the text retrieval area there have been many user-studies, though not sufficient and not comprehensive enough, to reveal their bibliographic and full-text information search/browsing behaviours in different subject area and in different interface styles. For example, highlighting users' query keywords in full-text viewing has been empirically validated for its user preference and we now see many text search interfaces adopting this feature as a norm. In image retrieval, such user studies can be found in only small numbers. An example would be Enser's query analysis [5] which revealed the query behaviour for a non-domain-specific users.

In the area of digital video there is no such systematic study of users and thus no established ideas on its users, their needs and information seeking behaviours. Although depending on for what purpose the video collection is used, generally we do not have sufficient ideas on user requirements that can thus guide our UI design for video systems. This is actually the situation for other systems mentioned above as well as design of UI has often been in a very much ad hoc manner, with most design decisions largely based on simple assumptions and on previous systems seen. Information about the users and their preferences is directly fed from observation into future system development and its UI design.

2.1   Query-Formulation Tools for Video IR

In the usual model of an IR search session, one starts by expressing an information need in a form understood by the system. Although in many older IR systems a user has to continue coming back to this query-formulation screen to modify query statement based on observation of the previous retrieval-set, nowadays more and more emphasis is placed on search by browsing e.g. Marchionini [6]. Most especially in the case of databases with image and video data where subjectivity can be a big problem in indexing/identification, search by browsing becomes an even more important way of locating wanted items [7]. It also seems to be a good idea to let the system make more use of the human user's efficient visual image recognition system [8], thus loading some burden off computer. This probably explains why some of the more advanced multimedia systems such as FRANK [9] and MediaKey [10] concentrate much more on browsing & information visualisation than on search by querying.

Despite its more limited use, search by query-formulating can be a good way of 'pruning' a search space to initially get eliminate a large set of definitely unrelated items from those to be browsed. This kind of rough filtering as a starting point then doing more fine-tuned searching is said to be efficient, and a good example of this is IBM's QBIC (Query-By-Image-Content) system [11], with its initial, though not very specific, keyword search as a start, and from then on by continuing search by more sensitive visual querying/browsing the reduced set of images.

In video interfaces, one type of tool that seems promising is 'query-by-motion'. The idea here is that users of video systems want to initiate their search by querying a video clip's moving characteristic – 'I want to find a shot with an airplane flying slowly from left to right' or 'a shot with camera zooming in'. Though we still have to determine what kind of user group (if at all) and in which domain/profession such queries are likely to be asked in what format exactly, it seems to make sense to provide a query tool to allow specification of the moving aspect of video for the system especially since motion is the one of the key attributes of video medium. At present there are not many video systems that provide any form of search tool where the user can specify desired elements of motion in a video. A good example of this is the VideoQ system [12] which provides a sketch-drawing tool where the user draws an object by defining its shape, colour and texture, then specifying the motion trajectory of that object within the drawing canvas. In this tool, camera motion such as zoom-in and zoom–out can be also specified. The retrieved set of video sequences will have all similar objects moving in the similar direction, with similar camera movements. The MovEase [13] system also provides a similar object & camera motion definition tool with icons representing each object.

Another example of a search tool by motion query is the NeTra-V system [14] which takes a QBE (Query-By-Example) method. Example video sequences are played on the screen and the user selects one of those videos. As a result the system will retrieve another set of videos that shows similar movement to the selected video sequence.

Although technically interesting, we still do not know whether these query tools will actually be a major help to users, and also these tools' retrieval effectiveness in themselves may not be sufficient or precise enough to be used as the sole query method for a video system. But as mentioned above they could provide a starting point, complementing later stage browsing tools.

2.2   Browsing Tools for Video IR

User-interface concerns concentrating on browsing are closely related to the presentation of different levels/types of abstraction of a particular medium and their trade-offs. If the abstraction was done very efficiently so that each item can be represented in a very simple form, a user will be able to browse a large number of such items on one screen quickly with little effort, meaning no need to remember much. Such a condensed abstraction often may not provide sufficient information on each item. On the other hand if the abstraction form is very detailed – as loose as in the original form – then user can fully appreciate and figure out the item which satisfies him but takes more time and effort while also unable to have a wide glance in relation to other items.

With conventional text IR interfaces, two usual abstraction forms available to the users would be the one lineful of text surrogate for initial query result display, and full-text for detailed inspection. Between these two extremes of abstraction, we can think of an interface that provides different levels of detail such as showing subject descriptors or abstracts or combinations of these. Thus our concerns become what would be the best form of abstraction, how many different forms are needed and how to provide changing browsing modes from one form to another, for the best performance of users' browsing. Exactly the same idea applies to video IR interfaces, though what kind of abstraction form can be provided seems to be a more open question and have more possibilities (or more difficulties) than text abstraction due to the nature of video medium itself.

Currently available or claimed browsing tools for video IR are quite limited in number and variations. On the visual side there are at least the following possibilities for abstraction:

- Thumbnails taken from keyframes or a synthesised still-image
- Dynamic thumbnail, showing all the keyframes from one slot
- Keyframe list, or storyboard, ordered chronologically or hierarchically
- Skims obtained by sub-sampling or some more intelligent method
- Playback

In terms of text related to video, the levels of abstraction can vary from a short textual description of shots and scenes to a detailed viewing of the entire transcript with a synchronisation of the different representations. Where there are alternative representations available, presenting a mixture of these representations is a good idea. This is especially since the systems that use spoken words within the video (either automatically producing transcripts by speech recognition, or using available caption/subtitles) tend to enhance the browsing facility by displaying a transcript along with player or other visual abstraction methods and allow synchronisation between them while the user browses different positions within the video sequence.

The FRANK [9] and MediaKey [10] systems use such synchronised browsing where keywords in the transcript can be quickly located and when this happens the user switches to visualisation of the images and thus we see integration of different components in overall browsing.

2.3   Transition between Different Abstraction Viewings

One would normally design an IR interface with an initial screen showing items with most condensed form (thus accommodating a large number of items at once and allowing a user to have a look at the larger set, its characteristics, etc.) and subsequently getting into a more detailed view based on users' requests. This is often quoted as 'overview first, zoom and filter, then details-on-demand' [15]. If there are many abstraction views possible, then generally the more availability of such abstractions, the better, allowing a user to select a particular abstraction level for overview browsing of a number of items on screen, as well as allowing for a particular item's full set of available abstractions to be ready for viewing at any moment in the session at a user's request.

3   Some Video Browsing Tools

After an extensive survey of literature, web pages and online information, we have gathered information on fifteen different systems which perform content access or browsing on digital video. Some of these are commercial products, others prototypes. These systems will be used later in our feature-system matrix in section 4. We now

present a brief pointer to each of them and more deatails of what features each has are included in section 4.

- The SWIM (Show What I Mean) system [7] at the National University of Singapore is cumulation of various tools developed for general video archiving purpose including broadcasting companies, film industries, security agencies, libraries and educational purposes.

- MediaKey [10] is a commercial product from Informedia project (URL: http://www.informedia. cs.cmu.edu/) at Carnegie Mellon University, one of the six Digital Library Initiatives projects in US. Technologies developed by Informedia project are fed into MediaKey system, which targets at broadcasting, film, advertising, training industries for real-time logging, cataloguing, archiving and searching of their video materials.

- The University of Kansas's VISION (Video Indexing for SearchIng Over Networks) [16] system aims to provide Internet access to mainly news broadcast for school and other Internet users.

- Columbia University's WebSEEk system (URL: http://disney.ctr.columbia.edu/webseek/) is a Web-based image/video cataloguing system that automatically spiders the Web to catalogue and provide content-based access to image/video files on the Web.

- Internet CNN NEWSROOM (URL: http://www.nmis.org/NewsInteractive/CNN/Newsroom /contents.html) is a Web version of CNN NEWSROOM, 15 minute-per-day free video program by Turner Broadcasting. The system automatically digitises each program with closed caption, for educational use at primary and secondary school classrooms.

- VideoSTAR (Video Storage And Retrieval) [17] from the Norwegian University of Science and Technology is a generic database platform on which various indexing/search tools are developed for professional librarians/archivists documenting and searching mainly for TV broadcasting news and films.

- FRANK (TV/Film Researchers Archival Navigation Kit) [9] is a project at DIMMIS (Distributed Interactive Multimedia Information Services) group, RDN-CRC (Research Data Network Cooperative Research Centre), Australia, to provide online browsing tool for mainly research part of TV/film production, to help evaluate potential materials for inclusion in their production.

- CAETI Internet Multimedia Library (URL: http://www.videolib.princeton.edu/) at Princeton University is focusing on providing access to video materials of political advertisement, news reports and NASA documentaries for school education on the Web.

- Pop-Eye (URL: http://pop-eye.tros.com/) and OLIVE (URL: http://twentyone.tpd.tno.nl/olive/) projects are Telematics Application Programme in Language Engineering Sector, lead by Dutch broadcasting company TROS and research institute TNO-TPD respectively, and other European partners from universities and software houses. These two projects are continued projects aiming to develop multilingual, automatic video archiving system concentrating on making subtitle/speech recognition for their TV production/broadcasting partners within the project, helping them reuse video materials by providing detailed content-based access.

- Columbia University's VideoQ [12] is an experimentation on providing query tools based on visual features of video content including colour, texture, shape and motion.

- MovEase (Motion Video Attribute Selector) [13] is an experimental system developed by Siemens Corporate Research Inc., to mainly consider motion as a primary attribute in query formulation and retrieval of video.

- NeTra-V ('netra' means eye in Sanskrit) [14] is developed at University of California Santa Barbara, and also experimenting on query tool for low-level content-based search based on colour, texture and motion.

- MediaArchive is a commercial video archival system developed from an Esprit project, EUROMEDIA (URL: http://www.foyer.de/euromedia/home.html), developing tools for large-scale digital archives for media producers such as TV broadcasting, publishers and multimedia producers.

- Screening Room (URL: http://www.excalib.com/products/video/screen.html) is Excalibur Technologies's automatic video archiving/retrieval system, targeting broadcasters, video producers, advertising agencies and entertainment companies to help manage their video assets.

- VideoLogger (URL: http://www.virage.com/market/cataloger.html) is an automatic video logging/cataloguing tool developed by Virage Inc., main function being real-time automatic cataloguing of video. For browsing task, the company provides developer toolkit to customise interface tools, used in video archives for TV/film production, education, etc.

4   Review of System Features

In this section we examine some of the features of video browsing tools we introduced in section 3. Before we present the feature-system matrix detailing which systems have which features, a brief explanation of these features is now given.

- 'Cataloguing tools' is a distinct interface designed for the cataloguers or indexers of video material. 'Manual cataloguing tool' provides a human cataloguer or indexer with an easy way to navigate video sequences frame-by-frame and allows making segmentation and text annotation. 'Semi-automatic tool' is basically the same kind of tool as the manual one, but is particularly designed to help human indexer to make any changes or additional annotations after the system's automatic indexing has been done. This tool usually provides a playback screen and segmented shot list with keyframes along with text annotation fields, etc. 'Threshold adjusting before automatic segmentation' is to set shot detecting sensitivity so that more appropriate segmentation of video can be done. For example, Excalibur's Screening Room allows the indexer to pre-select particular genre (animation, drama/comedy, documentary, etc.) before automatic segmentation is conducted to produce more precise result.

- 'Keyframe-based sketch-drawing' is a content-based query tool where the user defines static visual features (colour percentage, texture, shape, etc.) in a drawing tool and the system matches this against keyframes within the database. Basically this is a tool found in those popular content-based still-image retrieval systems such as QBIC system [11], but quite frequently adopted in video retrieval systems to retrieve keyframes. Though often criticised for not being capable of addressing motion or audio attributes of video information (such as in Dimitrova [18] and Iyengar [19]), this could be one of the complementary search tools based on a video's visual characteristic.

- 'Histogram manipulation' is a technique where the user can modify the histogram of a keyframe's visual features such as colour and then request other keyframes with a similar histogram. This could be useful for users who know what a histogram is, by helping specify low-level visual features very specifically. Query-formulation is often the most problematic stage for novice users as it can be difficult for them to figure out, for example, how to use the sketch-drawing tool or histogram tool. One solution offered is the QBE (Query-By-Example) method where example items are presented to the user and he can simply query asking all the similar items to the one he has selected. 'Keyframe-based QBE (Query-By-Example)' is such a method where the user can browse through a set of keyframes from a video sequences and request a new set of keyframes that are all similar to any particular one.

- 'Motion-based sketch-drawing' was mentioned earlier and extends the keyframe-based sketch-drawing tool by including user specifying object/camera motion (in addition to objects' static characteristics) to search for video sequences. This motion query tool comes from the idea that a good IR system should provide search tools based on the medium's attributes (in the video medium's case, audio, image and motion attributes). One of the difficulties in providing this feature is that composing motion-based picture could be a complicated task for the user, as he has to be able to be quite specific in defining not only objects and its characteristics, but also the movement of those objects and different camera motions as well. Saving a composed query for later reuse or for other users can be a way to reduce this problem (as in MovEase system [13]), as well as 'Motion-based QBE (Query-By-Example)' which allows the user to browse through a number of scenes with similar visual and motion characteristics.

- Also listed are various Video abstractions, roughly in the order of their condensedness. Representing a video sequence as a single line of textual description (usually by manual annotation) or as a single thumbnail extracted from a representative keyframe are two common video abstractions used in most video systems. Transcript displaying is also one way of representing a video's content, though in many systems such a transcript may not be readily available unless they apply some means of automatically generating transcript from video contents.

- 'Keyframe list in chronological order' is the common abstraction method that displays a set of keyframes within a shot/scene/programme, often called 'storyboard.'

- 'Option for different density of keyframe list viewing' is often used in a networked environment to help the user have a look at the content of video before having to download large size data. Users usually have options for storyboards sampled at different rates.

- 'Interactive hierarchical keyframe browser' is a particular video browser tool that shows all the keyframes in a video hierarchically – following a particular portion of interest brings up more detailed keyframes for the user.

- 'Timed playback of keyframes' is one form of video abstraction that shows a set of keyframes in one location sequentially. Screen 'real estate' can be saved with this method, because only single rectangular space for all keyframes is required, at the expense of the user having to keep on looking at these 'dynamic' thumbnails longer

to view a number of keyframes in a slide-show style.

- 'Highlight playing' is a moving extension of 'timed playback of keyframes': a condensed version of video sequences much the same as those movie trailers, showing only the most interesting (or important) sequences of a video. Though the resultant sequence is still something which has to be played (i.e. is time-dependent), the length of time the user has to watch is reduced thus saving browsing time. This abstraction comes from the Informedia project, where it is called 'skim [20].'

- 'Playback' is the normal VCR-like tool that plays video sequences, as seen in most video systems. In some cases, user's ultimate goal of using a digital video system might be watching a long programme using such a playback tool, or it might be just previewing a small, low quality playback to locate a right video material, with analogue full-version sequence later requested off-line. In a playback tool usually there are buttons for play, fast forward, fast backward, frame-by-frame, etc.

- Also listed are features that provide synchronised presentation of more than one video abstraction (transcript + playback, keyframe list + playback, etc.) and text/scene search capability in this synchronised mode.

- 'Intelligent keyframe selection' indicates any content-based method that selects one or more frames from a video sequence, so that those selected frames can be used as the 'representative' of the whole shot, as one form of video abstraction. These selected representative frames are supposed to provide a good idea of what the shot as a whole is about. An automatically keyframe selection method often used is to select the first frame from the shot, or the middle frame, or any other arbitrary frame (the 10th frame is used in the Siemens browser [4]). Here there is the question of whether the first frame necessarily represents the whole shot, and whether we could find some more 'intelligent' way to automatically select a more representative frame from each shot. For example, the SWIM system [7] uses colour & motion information within the shot to select keyframes.

- This issue of keyframe selection is included here because which frames within a shot are displayed on the screen will surely affect user's search performance – it is technical issue but also a UI concern, too. Within our research group we are also working on this problem.

We now present our overall feature-system matrix (Table 1). Interface features are roughly grouped by their nature/similarity so that it is easy to see at a glance which aspect of interface each system concentrates more than other features.

Feature \ System	SWIM	MediaKey (Informedia*)	VISION	WebSEEk	Internet CNN NEWSROOM	VideoSTAR	FRANK	CAETI*
Cataloguing tool								
Semi – automatic cataloguing tool		O						
Manual cataloguing tool						O		
Threshold adjustable before automatic segmentation								
Textual query								
Natural language (or Keyword) input	O	O	O	O	O	O	O	O
Category or Keyword list browsing	O	O	O	O	O	O		O
Use audio information for indexing/searching		O	O		O		O	O
Automatic full transcript generation**		O						
Visual query								
Keyframe-based sketch-drawing	O							
Histogram manipulation				O				
Keyframe-based QBE	O			O				
Motion-based sketch-drawing								
Motion-based QBE	O							
Video Abstractions								
Textual description	O	O	O	O	O	O		O
Transcript		O			O		O	O
Single keyframe	O	O	O		O			
Keyframe list in chronological order (storyboard)		O					O	O
Option for different density of keyframe list viewing								O
Interactive hierarchical keyframe browser	O							
Timed playback of keyframes				O				O
Highlight playing ('skim')		O						
Playback	O	O	O	O	O	O	O	O
Transcript + Playback synchronisation		O					O	
Keyframes + Playback synchronisation		O					O	
Text search with Playback and/or Keyframes sync						O	O	
Intelligent keyframe selection	O	O		O				O

*This project is not yet completed at the time of writing, thus could facilitate more features in the near future.

**Partial transcript generation such as word spotting is included in previous 'Use audio information for indexing/searching.'

Table 1 Feature-System matrix

Feature \ System	Pop-Eye / OLIVE*	VideoQ	MovEase	NeTra-V	EUROMEDIA*	Screening Room	VideoLogger
Cataloguing tool							
Semi – automatic cataloguing tool					O	O	O
Manual cataloguing tool							
Threshold adjustable before automatic segmentation						O	O
Textual query							
Natural language (or Keyword) input	O	O			O	O	O
Category or Keyword list browsing	O	O			O		O
Use audio information for indexing/searching	O					O	O
Automatic full transcript generation**	O						
Visual query							
Keyframe-based sketch-drawing			O				
Histogram manipulation							
Keyframe-based QBE							
Motion-based sketch-drawing		O	O				
Motion-based QBE		O		O			
Video Abstractions							
Textual description		O	O		O	O	O
Transcript	O					O	O
Single keyframe	O	O	O	O	O	O	O
Keyframe list in chronological order (storyboard)	O				O	O	O
Option for different density of keyframe list viewing					O	O	
Interactive hierarchical keyframe browser							
Timed playback of keyframes							
Highlight playing ('skim')							
Playback	O	O	O	O	O	O	O
Transcript + Playback synchronisation							
Keyframes + Playback synchronisation	O						O
Text search with Playback and/or Keyframes sync							O
Intelligent keyframe selection					O		

*This project is not yet completed at the time of writing, thus could facilitate more features in the near future.

**Partial transcript generation such as word spotting is included in previous 'Use audio information for indexing/searching.'

Table 1 (Continued) Feature-System matrix

5   Selection of Good Features

With so many potentially useful tools for video browsing listed above and indeed many others we have not covered, the most appropriate way to make progress may be to identify different classes of users who will require different sets of features and functions, some of which may as yet be unknown which best help accomplish their tasks. This is the "user-driven" approach to system development and evaluation rather than the technically-driven approach most often used in conventional information retrieval.

The most obvious class of users are indexers or cataloguers who will need 'cataloguing tools' (top row in the matrix) to manually or semi-automatically index and annotate video but whose role will become less important as video indexing technology becomes more mature and collections become larger. It is this class of user that many of the current crop of systems are targeted at but we have much to learn about other potential end-users of our video systems. We are completely ignorant about what the domestic consumer is likely to want from digital TV broadcasts which have now started throughout Europe. As we move from the current passive viewing of broadcast TV and on to digital TV which we also now receive and into interactive TV, personalised TV and access to archives of digital TV/movies, the great unknown is what do untrained, domestic consumer users want from all this technology and what tools can they be provided with? To progress this somewhat we are planning to do the following

- We will develop a system to digitally record broadcast digital TV programs, as selected by users using videoplus requests, onto our server where they will remain archived for a fixed period of time.
- We will analyse this digital video, performing operations such as our own shot boundary detection [21] and key frame identification.
- We will develop an initial browsing tool which will provide the usual VCR controls on this archive of video as well as key frame browsing and some of the features mentioned in the previous section.
- We will deploy this system for use within our department with a Web interface to allow users to indicate their preferred programs to be recorded, which they can subsequently play/browse/search/save, from their desktops.
- We will monitor the use of this system, which effectively becomes our rolling testbed demonstrator for interface ideas and we will use continuous user feedback to refine and enhance.

This approach gives us a closed user population of over 50 people who have a real motivation to use the system (the major topics of conversation in our department are traffic congestion, the weather, and what is, was or will be on TV). Actual trials of this system will reveal much about the end-users' needs, behaviour and preferences regarding various browsing tools and with the testbed system we will conduct user studies with various interface tools, whose results in turn will feed to various modifications of interface design.

6   Conclusion

UI design must not be an ad hoc process at the end of system development, with ideas largely coming from how previous systems looked. It should be based on proper theories (if not enough at the moment, as least based on available heuristics and guidelines), and every design decision should have its rationale behind it, in regard to the medium it carries and the users who will be using it.

Our work is trying to develop an interface for a video IR system by:

(1) identifying video browsing tools which are available and implementable by us;
(2) incorporating them into a working system with proper UI principles applied;
(3) monitoring their use, then
(4) changing and refining according to the analysis of user monitoring.

This process could be a good case of a systematic development of an interface for a very new medium, involving market study, reasoned UI design work, and iterative design based on actual user studies. We expect to have our first testbed version operational by Spring 1999.

7   References

1.   Aigrain P, Joly P, Lepain P, Longueville, V. Representation-based user interfaces for the audiovisual library of year 2000. Proceedings of SPIE, 2417, Multimedia Computing and Networking, San Jose, CA, February 5-10, 1995, 35-45.

Project WWW site at URL: http://www.irit.fr/SSI/ACTIVITES/EQ_AMI/ami.en.html

2. Elliott E. Thinking with motion images via streams and collages. Master thesis, MIT, 1993.
Video Streamer demonstration at URL: http://tower.lcs.mit.edu/~richard/eddie/VideoStreamer.html

3. Mills M, Cohen J, Wong Y. A magnifier tool for video data. CHI'92. Proceedings on Human factors in computing systems, Monterey, CA, May 3-7, 1992, 93-98.

4. Arman F, Depommier R, Hsu A, Chiu M-Y. Content-based browsing of video sequences. Proceedings of ACM International Conference on Multimedia '94, San Francisco, CA, October 15-20, 1994, 97-103.

5. Enser P. Query analysis in a visual information retrieval context. Journal of Document & Text Management, 1993, 1(1), 25-52.

6. Marchionini G. Information seeking in electronic environment. Cambridge Series in Human-Computer Interaction, No.9. Cambridge, U.K: Cambridge University Press, 1995.

7. Zhang H, Low C, Smoliar S, Wu J. Video parsing, retrieval and browsing: an integrated and content-based solution. Proceedings of ACM International Conference on Multimedia '95, San Francisco, CA, November 7-9, 1995, 503-512.

8. Enser P. Pictorial information retrieval. Journal of Documentation, 1995, 51(2), 126-170.

9. Simpson-Young B, Yap K. FRANK: trialing a system for remote navigation of film archives. SPIE International Symposium on Voice, Video and Data Communications, Boston MA, November 18-22, 1996.
Project WWW site at URL: http://www.syd.dit.csiro.au/projects/dimmis/frank/

10. MediaKey Digital Video Library System. ISLIP Media.
Available online at URL: http://www.islip.com/fprod.htm

11. Ashley J, Barber R, Flickner M, et al. Automatic and semiautomatic methods for image annotation and retrieval in query by image content (QBIC). Proceedings of SPIE, 2420, Storage and Retrieval for Image and Video Databases III, San Jose, CA, February 5-10, 1995, 24-35.
QBIC system demonstration at URL: http://wwwqbic.almaden.ibm.com/cgi-bin/photo-demo

12. Chang S, Chen W, Meng H, Sundaram H, Zhong D. VideoQ: an automated content based video search system using visual cues. Proceedings of ACM International Conference on Multimedia '97, Seattle, WA, November 9-13, 1997, 313-324.
Project WWW site at URL: http://www.ctr.columbia.edu/VideoQ/

13. Ahanger G, Benson D, Little TDC. Video query formulation. Proceedings of SPIE, 2420, Storage and Retrieval for Image and Video Databases III, San Jose, CA, February 5-10, 1995, 280-291.

14. Deng Y, Mukherjee D, Manjunath BS. NeTra-V: towards an object-based video representation. Proceedings of SPIE, 3312, Storage and Retrieval for Image and Video Databases VI, San Jose, CA, January 24-30, 1998, 202-213.
Project WWW site at URL: http://copland.ece.ucsb.edu/Demo/video/

15. Shneiderman B. The eyes have it: a task by data type taxonomy for information visualisation. Proceedings of IEEE Symposium on Visual Languages, Denver, CO, 1996, 336-343.

16. Li W, Gauch S, Gauch J, Pua K. VISION: a digital video library. DL '96. Proceedings of ACM International Conference on Digital Libraries, Bethesda, MD, March 20-23, 1996, 19-27.

17. Hjelsvold R, Lagorgen S, Midtstraum R, Sandsta O. Integrated video archive tools. Proceedings of ACM International Conference on Multimedia '95, San Francisco, CA, November 5-9, 1995, 283-293.

18. Dimitrova N. The myth of semantic video retrieval. ACM Computing Surveys,1995, 27(4), 584-586.

19. Iyengar G, Lippman AB. Semantically controlled content-based retrieval of video sequences. To appear in presentation at Storage and Retrieval for Image and Video Databases VII, San Jose, January 1999.

20. Christel M, Winkler D, Tayler C. Multimedia abstractions for a digital video library. DL '97. Proceedings of ACM International Conference on Digital Libraries, Philadelphia, PA, July 23-26, 1997, 21-29.
    Project WWW site at URL: http://www.informedia.cs.cmu.edu/

21. O'Toole C, Smeaton F, Murphy N, Marlow S. Evaluation of automatic shot boundary detection on a large video test suite. To appear in The Challenge of Image Retrieval, Newcastle, February 15-16, 1999.