

For X11 VIDEO EXTENSION TECHNICAL MEETING, Bellcore, June 9-10, 1988

SOFTWARE CONSIDERATIONS FOR MULTIMEDIA VIDEO PROJECTS

Glorianna Davenport
Assistant Professor
Film/Video Group of the Media Laboratory
Massachusetts Institute of Technology

Prototype application software described in this paper was developed by Benjamin Rubin, MSVS '89, Louis Pepe '88, and Brian E. Bradley '89 on an MIT Project Athena Workstation

ABSTRACT

This paper describes a multi-media application, "A City in Transition: New Orleans, 1983-86." The project is being developed on Project Athena Visual Workstations as a curriculum resource for the study of architecture, urban planning, and political science. The project includes 3 hours of movie sequences and stills mastered on optical videodisc, as well as an extensive set of ASCII files containing support material.

The Project Athena Visual Workstation with a Parallax 1280 videographics board and a single 1280 x 1024 display has been configured with six videodisc players running through two 4 x 1 audio-follow-video switchers. The Project Athena network allows us to remotely access the relational database manager, INGRES; we currently attach to INGRES via GDB, a general purpose database handler, developed by Noah Mendelsohn, which allows more efficient processing of complex queries.

Prototype software, developed under X-10, includes a spring loaded mouse control interface to the videodisc player and a professional edit-list management package. Data can be entered and queries constructed using menu bars and postage stamp icons taken from the video footage. Users can also select their own postage stamp icons to represent particular shots during the editing process, or to create multimedia display structures. It is our hope under X-11 to create general tools, which can serve multiple applications wherever possible. For instance, one of the issues which has recently emerged is how icons are used to sequence and edit different media: is the intent with stills and film similar enough to develop program modules, and by what means does a user learn how to customize his/her environment?

INTRODUCTION

Observational recordings -- news footage, home movies, scientific and social documentaries -- communicate information about people, places, events, issues; they encourage personal reflection and interpretations about the how and why of the physical phenomena and human actions. For the past several years, the MIT Film/Video Group has been concerned with analyzing methods documentary moviemakers use to previsualize and edit their productions. How can we mirror the best of these methods in personalized multimedia environments? What is the function of the editors' log? How much more information are we holding in our heads? How do we as editors translate our previsualization into a cinematic story? How can our viewing practices and professional editing tools

be redesigned to serve a much wider group of users who will want to access movie material as an information resource as well as an artistic medium?

The smallest denominator of movie footage is the shot (one or more frames) and sound objects; shots and sound objects may or may not be of the same duration. Content and context information can be assigned to each shot or sound element; this kind of information can vary from a description of precise objects in a frame to actions -- expressions, gestures etc. -- in a shot, to the who, what, where, when, why of an event or sequence.

When we explore why (and in what detail) we might want to database movie content, we quickly distinguish two user groups: the editor and the viewer. As video becomes more common to workstation environments and as video source material becomes available for use in teleconferencing, education, social and scientific research, distinctions between viewer and editor quickly dissolve. All viewers will clearly need some level of sophistication in the editing tools!

AN EXPERIMENTAL CASE STUDY

Film/Video has always encouraged filmmakers to actively participate in the design of their own tools. From our earliest years we have constantly modified cameras to better serve the needs of the moviemaker's style. Likewise, a critical aspect of our evolving approach to multimedia as a form relies on successive production and implementation of new multimedia projects. [1]

In 1982, I happened to be in New Orleans and learned that, partially in response to a decision to hold the 1984 Louisiana World Exposition in downtown New Orleans, 2 1/2 miles of prime Mississippi River frontage would be developed virtually overnight. A viable story had appeared which was complex enough to merit multi-media delivery. The city was small (600,000), we were being given access, and the National Endowment for the Arts agreed to fund the project. In January 1983, Richard Leacock and I began a 3 year case study about urban change.

Simultaneously, Project Athena was starting up. As a moviemaking group, we wanted to experiment with new polylinear forms. In addition, a cinematic case study about urban change was a perfect resource for curriculum in the School of Architecture and Urban Planning (of which we were a part).

Over the three year period, Ricky and I visited New Orleans 18 times and shot close to 40 hours of 16mm film with synchronous sound. We observed many meetings, always recording sections which seemed, from what we knew at the time, particularly interesting to us. We also recorded a number of street scenes, demolition and construction scenes, and in the case of one development project, we recorded a long shot from the identical angle on almost every visit. As with any film, we collected a variety of support documentation as our filming progressed -- maps, still frames, reports, pro-formas, and character profiles, to mention a few.

The final 2 hour 45 minute edited movie was mastered on 6 write-once videodiscs in the fall of 1987. A development workstation with the Parallax board was delivered in late August. Project Athena provided the basic system software, including videodisc drivers and some application elements such as a prototype menu-maker. Throughout the summer, Brian Bradley continued to work on various modules for inputting and retrieving data from INGRES; Ben Rubin began prototyping editing software in September. Meanwhile, the New Orleans case study was used by 6 students for their final projects in an "Introduction to Urban Planning" course taught by Associate Professor Dennis Frenchman. This created a somewhat chaotic fall.

New programs and flat files appeared daily on the system but always lagged behind the needs of our real users. On the other hand, we were observing vociferous learners: how could we help them manage the information they had access to? Two students, Jon Teischer and Don Knerr, wrote their final paper for this course on the workstation using video excerpts as part of the text. The students then xeroxed the text for myself and Frenchman. The paper was a strong, well researched classical analysis; the reader has no idea what she is missing, when we see the bold headline "INSERT DISC 3, PRESS LAND-SALE." After reading the printout, Frenchman came in and reread the paper on line. What a difference! Although they never bothered to explain what you are missing when you read the text alone, the video segments which introduce you to the character, the style, and even, in some cases, the illicitness of an action, are critical to really absorbing the argument which the students make.

OUR HYPOTHETICAL USER

X-11 notwithstanding, we must be able to describe why and how a user might interact with a video-data bank in order to build a robust, flexible environment. If we can, we would like to avoid creating a scraggly experience in which the viewer feels forced to jump around so fast that s/he gains no insight at all into an experience which has been patiently explored, recorded, and reconstructed.

Using the New Orleans project, for want of any other, we can imagine the following scenario: Let us presume user Q is going to a conference on "Incentives for land development by the private sector" which is being held in New Orleans. We can assume that accessing a "movie bank" is similar to accessing text banks today. Upon opening a connection to the database s/he might query the system for a range of information: New Orleans, tourism, history, public policy/land use, music, restaurants etc. Any but the last item would probably lead our future visitor to "A City in Transition: New Orleans 1983-1986," assuming that this project was available through this system.

Bypassing the issue of whether the video source is local or remote, let us assume these discs are mounted. The return of multiple references to scenes from a single movie might encourage our viewer to sit back for a few minutes and enjoy the movie as it develops without interacting. Of course, if s/he is under time constraints, s/he might choose to fast forward through scenes which appear boring or off the subject. Alternatively, our viewer might prefer to browse interactively, using text, picture icons, or some other interface which suggests sub-plots or stories; the latter, of course, lays a heavy burden either on the authors or on the database and associated constraint programs.

No matter what mode the user has chosen, let us assume that when s/he views the "New Leviathan Orchestra" rehearsing, s/he thinks that they would be fun to go see live. While the clip is playing, our relaxed viewer might decide to tag the scene (on the fly) so that s/he can easily return to it later. A more impulsive user might want to stop and send the clip immediately, via "e-mail," to a fellow traveler with a note: "must see," or "New Leviathan Orchestra, a must see." To discover the name of the group, our user can make a standard query or use the postage stamp size icons containing pictures of characters, places and scenes which are dynamically displayed next to the main video window. Perhaps, when video mail becomes a reality, the database header will follow the video scene automatically, thereby obviating the need for the sender to type in the citation or other information.

Progressing further into the movie, our viewer is watching a scene in which 4 men are sitting around a table discussing who should pay for wharves which will be taken out of service if recommendations made in a current planning report are followed. Two of the men clearly represent the interests of the Port of New Orleans. A third man appears to be responsible for the report, and claims that if these costs are charged to either the private or public sector, he would have to recommend higher density for construction on land along the river's edge.

This discussion relates to the subject which our viewer will be talking about at the conference. S/he pauses, mouses the icons of the various players and obtains personal profiles of each of them. Complex queries follow: financial issues/legal actions, etc. Files containing the transcript, the report itself, the sponsors of the report, land values in the surrounding area are reviewed. Property locations are indicated by overlays on the maps. What do the Dock Board representatives really want? Revenue? Bargaining power? Besides the developers, who could come up with the required money to replace these wharves? How does this session fit into the city's financial plan?

During this session, our viewer has tagged several other scenes to which s/he was referred. S/he might now wish to review these or edit them together to create a short mini-movie as part of the final presentation.

DESIGN APPROACH

Our work can be divided into three categories: 1) the production of documentary programs; 2) developing new user interfaces for browsing, editing and formatting video into multimedia documents; 3) data structures which can support relational retrieval of multi-media material. Our preference is to make 2 and 3 as generic as possible in order to support a variety of users and programs.

User configuration, data-entry, and query menu bars

The first thing one becomes aware of as a video application designer is that the single screen format, while effectively allowing one to access different programs and information sets, can quickly become a jungle of windows. Even though windows (including video) can be iconized quickly, being able to smoothly switch from browsing to editing modes is critical.

Ultimately, the interface for user configuration should simplify one's ability to perform the following tasks quickly and efficiently: comment on a scene, save the scene to a list, check a new reference, access a reference, learn about the characters in the scene, extract 5 lines of dialog to one's personal notes, return to the paper in progress, bring up the icon editor, load an edit list, select and order the icons, play back the sequence, embed a tag in the text of the paper which will play this mini-movie at the correct point. At this point, one might want to take on the reader/viewers role and view what has been created.

The meta pull-down menu is an effective way of closing and opening programs as well as data lists. The menu could allow the viewer to indicate to a remote server (a video jukebox) what videodisc program one wants to mount. Menus offer a quick and reliable method for easy browsing, adding, deleting or selecting from a list of data objects. Consistency of naming and spelling is critical to effective data entry and selection; the menu bar helps to streamline our quest for knowledge about anything by simplifying our input to a mouse click.

A rather trivial handicap of using a hierarchy of pull-down menu-bars is that they take up space. When a user is already running, viewing, or editing a program, and is using word icons as well as visual icons to query the database, pop-up menus may be more effective in terms of screen management.

Videodisc control, selection, editing software

All movie viewers will want and need interfaces which will allow them to view, annotate, select, and reorder (or edit) video segments. In prototyping an edit controller, special consideration was given to the control interface for shuttling video. This is an issue which editors will be particularly sensitive to. The button box requires too much attention on the part of the viewer. The current design, combining a shuttle bar and spring loaded mouse, is extremely responsive and will allow us to effectively mark in and out points accurately in the midst of dialogue.

Our video edit control and list management software has been designed to offer a basic professional level editing facility, including mark in and out, replace, add, move, save list, load list. The button box is a relatively good interface for these functions. The requirement that users describe shots is helpful when one is developing a long list; this function could eventually be integrated with a data-entry program allowing users to join their own descriptive file to the INGRES list of shots.

In order to view edited sequences seamlessly, or to view material from a large number of videodiscs, editing and browsing programs must cue and play material running on multiple videodisc players. We currently have 6 videodisc players running into two ganged switchers. One decision we have yet to make is how to deal with the audio so that as an editor one can listen to the track in slow motion forward and reverse while marking in and out points, but will not hear the track at other times.

Data structures

The most common question raised by viewers of documentary movies is "why" did something happen. Unfortunately, *why* generally requires a subjective evaluation: even the filmmakers rarely have sufficient, let alone complete enough, information to answer this with any certainty. The *who*, *what*, *where*, *when* of a particular event can help us understand the event and relate it to other events. For instance, when we wonder why someone voted a particular way, we can look at how s/he has voted in the past, what property s/he owns, etc. This combined comments of the movie maker and one or more experts may often expand our understanding of "why."

The determination of descriptive content datatypes is particularly critical when the "movie bank" rather than the particular movie becomes the "information resource" or application. Leaving aside, for the purposes of this paper,

the idea of knowledge-based selection or editing of material, the simple examples of a map, an architect, or an architectural style support this principle.

Using R.T.I.'s relational database manager INGRES, we have built a set of representations for this project which we believe can be used for any other documentary or movie project. Our basic database table includes the following fields: object ID, object type, parent ID, sort order. This allows us, among other things, to describe only a sequence on our first pass through a project, and later break certain scenes into shots. The databasing process can thus be more or less detailed and time consuming depending on the needs of a particular project.

The datum type can be the shot/scene/movie distinction or can be used to describe what is happening within the shot or scene. In databasing the New Orleans project, we have found the following datum types to be useful: people, places, things, events, issues, comments, references. These are further broken down to separate (for instance, under "people") individual names, group names, and roles.

CONCLUSION

The theoretical progress we have made in our understanding of multimedia environments over the past year is enormous. For the first time I have seen the promise of multimedia movies. I have also seen the darker side, moments in which I question whether it wouldn't be easier and more honest to just go out and make another movie. It may be overly optimistic to think that with the combination of databased information and knowledge-based operators, we can provide students and researchers intense, intellectually sound, interruptible movie experiences. In my blackest moments, when it's no longer any fun and I feel that things are hopeless, I usually go in to our workstation room, and play pieces of a movie in which every frame is beautiful, and I think about how and why New Orleans changed. At this point I begin to write a new background file. I recover, I leave convinced.

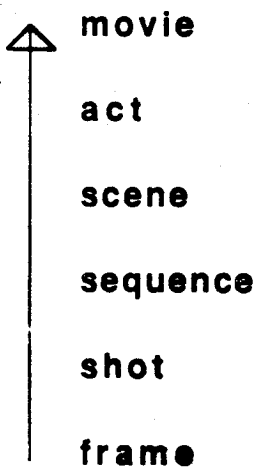
[1]. [Aspen, Automatic Transmission Manual](#) (MIT Architecture Machine Group) [Elastic Movies](#), [Anon](#) (MIT Film/Video Class projects) [Marital Fracture](#) (Roslyn Gerstein, PHD thesis; Russell Sasnett, MSVS thesis)

Narrative descriptions

datum type

People Places Things Actions Events Issues

object type



Editor's knowledge

intercutting between stories

edit on gesture/glance

edit on intonation

pans

zooms

relationships between sound and image

what does audience need to know

what should not be revealed yetwhy?

suspense, humor

Block Diagram

