Video Database Design:
Convivial Storytelling Tools

Glorianna Davenport, Associate Professor of Media Technology, Media Lab, MIT
MIT Media Lab, Rm E15-435, 20 Ames St., Cambridge, MA 02139
tel. (617)-253-1607, fax (617)-258-6264, email *gid@media.mit.edu*

Lee H. Morgenroth, Research Assistant, Media Lab, MIT
MIT Media Lab, Rm E15-435, 20 Ames St., Cambridge, MA 02139
tel. (617)-258-8948, fax (617)-258-6264, email morgen@media.mit.edu

Keywords: video database, storyteller systems, knowledge representation.

Abstract

Traditionally, video and film stories have been developed by a single author for a single release movie. Increasingly, video databases will be constructed as content libraries. These libraries will be used to deliver personalized messages to people who know very little about video story construction or editing. The challenge in making these systems usable is to develop storytelling tools for those unsophisticated users.

Story generation presumes some input by the user, first to create appropriate video description and second to suggest a story to tell. This paper offers an overview of some methods of description which have been associated with particular types of video logging and databases in the past. A general problem with these systems has been how to develop video annotations with efficiency and consistency. A new approach, story based annotation, is proposed. In this method a tool set is used for creating a top down story abstraction. Coupled with automatic database selection, this tool set allows the user to encode story based annotations and expert knowledge about editing into the database while producing stories. As the database grows, it becomes structured and annotated by a

process appropriate to the medium, namely storytelling. This structuring process optimizes the database for retrieval of video in a story form.

Introduction: What is a Video Database?

The evolution of large electronic media databases for application in news, ethnographic research, education, and training, as well as for a wide range of personal assistants as in travel, real estate, and other niche video-on-demand markets has been anticipated for almost a quarter of a century. In laboratories throughout the world, research in digital video signals has focused on hardware for storage, transmission, and display; on software for image creation, database management, and navigation; and on design principles for interactive story forms. Despite rapid progress in the development of digital video systems, content construction remains difficult.

One way to think about a digital video system is in the context of a database. Such a system contains a large collection of information elements of a certain granularity. These elements are described according to a range of attributes and can be accessed according to the intent of the user. Who will produce these large digital video databases and how will they affect our communication landscape? Who will have access to such a database and for how much? What tools are needed to use this database effectively? What can we learn about story creation which will affect the content and navigation of future databases? These questions lead us to examine possible video database applications.

The most frequently proposed benchmark for digital video technology has been movies-on-demand. This service will allow the consumer to watch what she wants, when she wants, where she wants, hopefully at a reasonable cost, so long as what she wants is one of 500 or 1000 movie titles. From a consumer stand point, this application does little to

change our perception of television entertainment. However, from a research perspective, the application has pushed the technical framework for variable bandwidth digital video servers.

Today commercial on-line databases for still pictures are beginning to appear. For example, the Kodak Picture Exchange provides publishers of all sorts with access to an electronic photographic archive. The client can search the Kodak repository for a particular photographer or theme, and receive low and medium resolution images over the network. This enables designers to work more efficiently and intimately with their client, while keeping an eye on cost. The negative need only be ordered when decisions about the final publication have been made. Similar still picture services are expected to be developed for news in the near future.

Stills and movies are contained objects; they are complete stories. What happens when we distinguish between story granularity and shot granularity. Consider a database of motion picture clips documenting world geography and culture. Or consider a database of my home video and clips I have captured from television. Or consider a database of video from which we can learn about the technology which goes into building a space ship. Video clips differ from stills or a pre-edited movie in that they constitute fragments of larger stories. As a communications resource, this type of database opens pathways to information and learning.

Once a large collection of data is created, there is still the challenge of making the information usable. Users of information repositories face three daunting problems:

- How do I know what is there?
- How do I find anything there?

• How can I find what I want?

Any database system must deal with these three problems in order to make the content contained within it available for use.

The library analogy

Since the early days of computing, the library analogy has shaped the design of information storage. What is a library but a repository for information objects which want to be accessed by multiple users? Early libraries were laid out alphabetically, according to the author's last name. Soon the problem of juggling shelf space to fit in new works became an enormous handy cap. The Dewey decimal system followed by the Library of Congress system allowed for an infinite and orderly expansion of library inventory while promoting an indexing method which allows for a more complex use of the architectural space. For the first time it became easy to navigate to a topic or collection of similar content. The architectural plan became linked to a conceptual plan which is easily remembered by a librarian or visitor. The architectural layout, the card catalogue, and the librarian together formed a system which alleviated the problems of a database user. The convergence of these interface elements created an environment in which users could interact with information in one of the following modes: serendipitous browsing; directed access; or conversational, personalized story construction.

A library visitor uses serendipitous browsing to access titles which have been carefully crafted by a single author. Visitors use this mode to browse the card catalog or to navigate the physical site. Through serendipitous browsing, the user may find their way wittingly or unwittingly to a specific location. Successive perusal need not follow any obvious pattern. The selection process -- which may be triggered by title, cover, or author --

typically involves more intuition than knowledge. The apparent aimlessness of the user's activity is characteristically unstressful and can be enormously pleasurable.

In focused retrieval the visitor also makes use of the way in which the library is organized. However, focused retrieval is the diametric opposite of serendipitous browsing. Here the visitor wants something. This want can occur at the start of a session or it can be part of an ongoing discovery process. In either case the visitor moves through the information based on a particular worldview. The library card catalogue is an easy to use interface which enables borrowers to perform focused searches and to obtain necessary pointers into the architectural plan. On arriving in the stacks, the user may cycle between focused retrieval and serendipitous browsing.

The third mode of library interaction involves a conversation between visitor and librarian which is journalistic in nature. The visitor introduces the subject of the query. The user may want to know where to find some particular type of material, or more likely may want to more clearly define what they are looking for. The librarian, acting as teacher or guide, offers expert knowledge which expands the historical, cultural, scientific or navigational context of the visitor's knowledge.

The database in film production

Traditionally film and video have been a medium in which stories are created by careful temporal sequencing of sound and image elements. Our eyes and ears perceive these pieces in sequence and our mind reconstructs a story from the parts. When the story is well crafted, we become immersed in the experience and our understanding of what happened is closely aligned with the filmmaker's intent. (Bordwell 1985)

While the end result may be a sequenced whole, the process of production and editing involves a definition of and interaction with the parts. As a film is produced, the physical elements are created. Before editing, these elements can be thought of as a database or library. The index into the database is provided by logs developed during production, which match a film can or video cassette with certain content. The role of the film or video editor is to select elements from this database and order them in a story structure.

Early research in interactive video applications, used the concept the video log to create a descriptive content database of video elements. Using a class/keyword structure, users can typically follow a character or view different clips about a place or a theme. (MacKay and Davenport 1989). Keywords are limited, however, in their ability to encode relationships. A movie can turn on a changing relationship between two characters, or on a set of events which are causally connected. A few experiments have used a more complex knowledge representation. (Davis 1993) However, these experiments have been thin in the content area.

The Database-story continuum

At first, the idea of a database seems at odds with the idea of a structured story, particularly a movie or television program. Clearly these expressive objects are made up of parts, but the rules which govern the creation of a sequential list of sequences or shots are intricate indeed. As we inspect the process of storytelling, we discover that a continuum exists between the finished product and the raw material of which it is created. In film and video production, the first step is to define a story; next, a database of shots and sounds is created; finally selections from the database are sequenced into story. In the realm of video databases, these three phases of production begin to overlap and the distinctions between them begin to blur.

The Problems of a Video Database User

The three problems of the database user discussed in the introduction, 'How do I know what is there?, How do I find anything there?, and How can I find what I want?', are familiar ones to library or video-rental store visitors. They are also familiar to the editor. Without some more or less formal coding scheme, movie editing would be a murky business at best. At the beginning of the editing cycle all the "rushes" are assembled in the editing room. This library contains the collection of granular options from which the editor selects and builds a single story.

How do I know what is there? Generally speaking, in traditional film and video editing, a single editor or filmmaker becomes familiar with the entire corpus of material by viewing all of it, sometimes many times over! Clearly this will not be practical for large databases which can be used in many contexts. Therefore, the user will need to gain familiarity with the database by browsing. However, learning about the content of a database through viewing one shot at a time may be a recipe for failure as there is no structure against which to measure the value of a shot in terms of story. What is needed is a conversational method of browsing which is similar to the librarian who can show the user the content through a collection of representative stories.

How do I find anything there? Strict adherence to a library model of storing footage combined with some level of logging helps the editor find appropriate material. During feature film production, an explicit coding scheme is used; the scheme numerically references scene, shot and take as specified in the preproduction story boards as well as script to film edge numbers or video SMPTE time code. Because footage is often collected out of sequence, footage will often be reordered prior to and certainly during the cut, so that the editor can see all the footage associated with one scene. In the case of film, when a physical cut is made and a shot is extracted from the work print, the edge code

maintains a reference back to the original. (Pincus 1984) The editor or assistant editor, if there is one, is responsible for keeping track of the film down to single frames. In the case of analog electronic recordings, where generational loss is a problem, the editor or editing system must maintain a trace back to the original reels of material. A fully digital system offers significant advantages. In a digital system, footage can be tagged with more complex annotations and possibly key frames. This facilitates both serendipitous and directed browsing by both topic and visual frame.

How can I find what I want? Finding the appropriate shot might involve any one or some combination of all of the interaction modes. If the editor has seen all the rushes and knows something exists, she can identify it from a log of storyboards, and then ask for it by shot and take number, or by frame number. If the editor is looking for a shot because a problem exists with the edit, she must define the nature of the problem and map it to a type of shot. Problems that arise in editing are diverse. A problem can be one of story information or can relate to how a sequence will work. Often the real trick is finding the best solution. The ideal search mode in this case is conversational and iterative. The task is simplified when an assistant editor is on the job: the editor requests, and the assistant editor, who may be even more familiar with the footage than the editor, finds. In the future, representation may be rich enough and the interface conversational enough to support appropriate analogic reasoning.

A Test Scenario

Until recently, large digital video databases referred theoretically at least to the movies-on-demand model or in practice to single projects. In most implementations, the total database was pre-edited to fit some specified storage limitation such as a certain number of video discs. To meet other research objectives, these projects tended to focus on a

particular research domain such as city planning in New Orleans, or children learning at the Hennigan School in Boston.

One of our early perceptions in the development of video database tools was that they needed to serve many projects and many users. In the fall of 1992, we looked at a set of video logging tools which we had built (Aguierre Smith and Davenport 1992) and modified them such that the data sets could be stored in Framer (Haase 1993) , a powerful persistent knowledge representation language written by Professor Ken Haase. This modification was the first step in a longer process of building a tool set which linked storytelling with video logging.

In order to develop our ideas about the relationship between story and a database of shots and sounds, we began simultaneously to build a database which might be put to use by a travel agency. The concept behind the video collection process was to create generic footage about place. The footage would be made up of one third location shots, one third interviews, and one third events of place. A story telling system for this purpose must be simple enough for a travel agent to use without incurring an enormous amount of training.

A Storytelling Tool Set

A unique set of tools, Stratagraph, Homer, the Sequencer, and Photobook, that have been developed at the MIT Media Laboratory begin to suggest story based methods of interacting with a database of video. This system supports the three types of navigation discussed in the library analogy.

The first modules developed for this system assumed that the task was one of data description or annotation. Users could log video using a constrained logging tool. The

annotations were displayed in the Stratagraph, which supports serendipitous or directed browsing. The approach to annotation used in these applications is stream based. Many descriptions can be layered on top of the video stream. To familiarize themselves with the video content, users can click on an annotation such as "place/shirehall" and the system will move to the first instance of that annotation. The Stratagraph simultaneously displays all the additional annotations for those frames. From these annotations, we discover that this shot is also about Woodbridge, is a medium shot, and is shaky etc. If we want, we can also move to the next instance of the description, or we can examine the surrounding footage in the video stream.
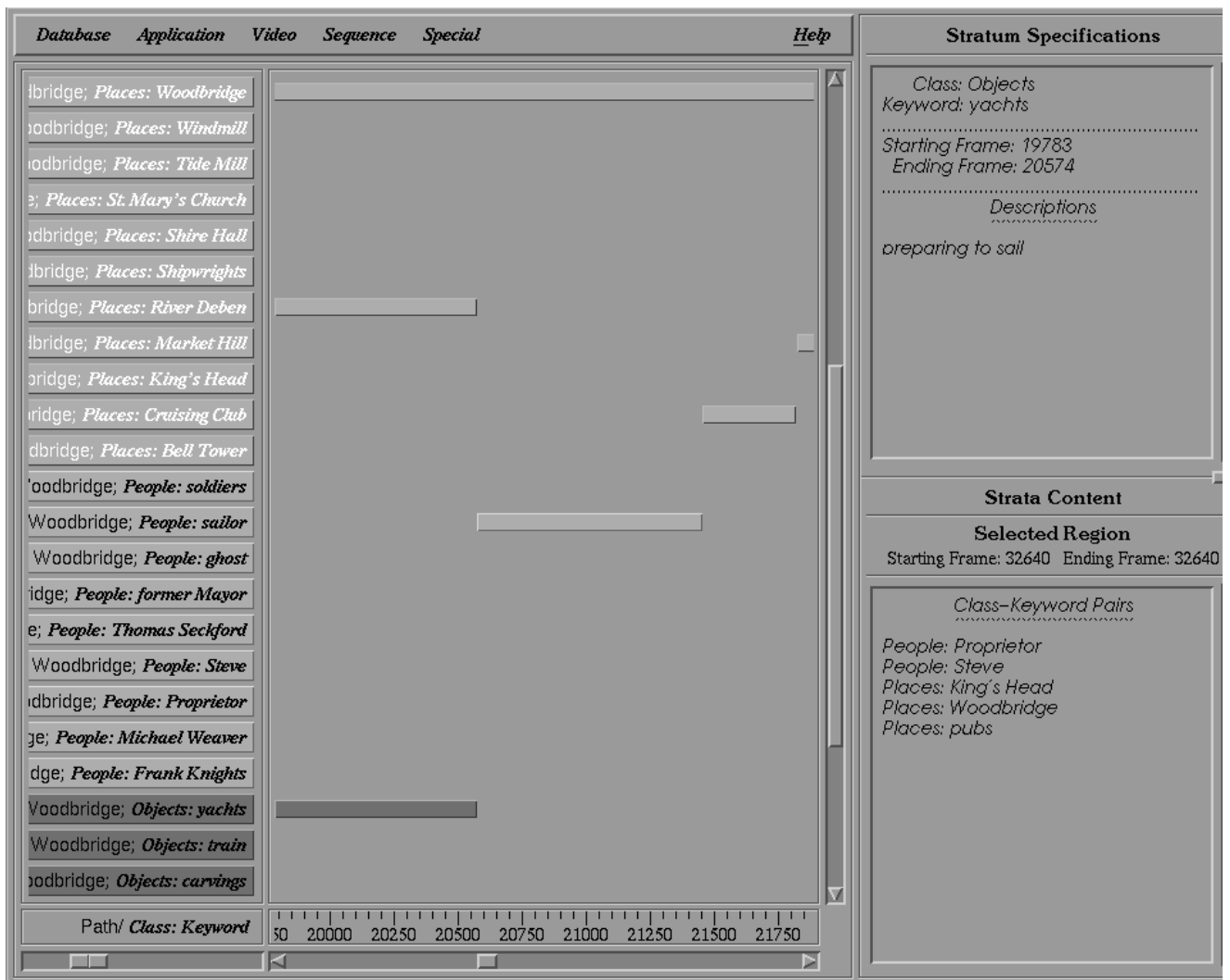
Figure 1. The Stratagraph.

In parallel with Stratagraph, users can learn about context and content by requesting that a story model browse the database for them. An early version of Homer, the story modeling tool, was built in 1992 (Morgenroth 1992). The story model forms a powerful and conversational mode of interaction between the user and the system. As the librarian offers the library visitor context, past story models offer the user of a video database a point at which they can begin to learn about the content. In the past year, Homer has been rewritten, and now includes a graphical workspace. This interface allows the user to play with the ideas of granularity, sequence, and story. Users can call on the library of story models or build their own in the following fashion.

Woodbridge

Figure 2. A single Block Homer story model.

Stories are built in Homer using abstract story chunks, called Blocks. Each Block gets its size from the duration specified by the user. Block sizes can range from one second, to several hours. Each Block also has a number of descriptions that determine story content. The maker can design a story by creating a progression of Blocks. Blocks can also be layered to create sequence structures. The Block based story model is simple enough that a non-expert can understand and even compose story models.
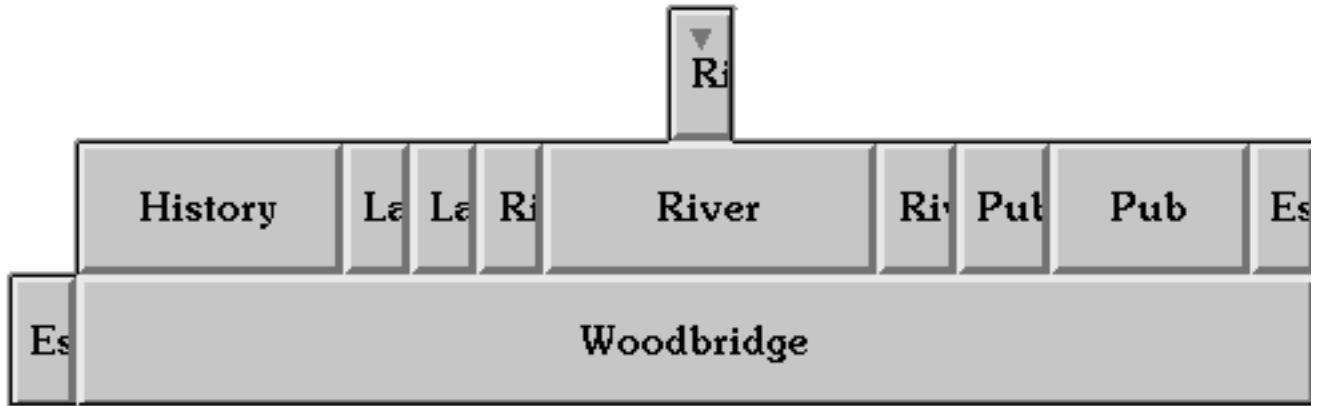
Figure 3. A complex layered Homer story model.

Because the Homer system does not use extensive rules to determine where to parse shots and because story models do not yet have a sense of style, a Sequencer, is included in the tool set. The Sequencer give the user hands on access to the Homer structured story. Using this tool, shots can be reordered, or replaced, and cuts can be trimmed. The Sequencer is also linked to Photobook (Pentland 1994), an image based search tool. Photobook allows the user to search for shots that 'look like' another shot in the database. This type of search is especially useful during the low level processes of editing, where the user will 'know what they want when they see it.'
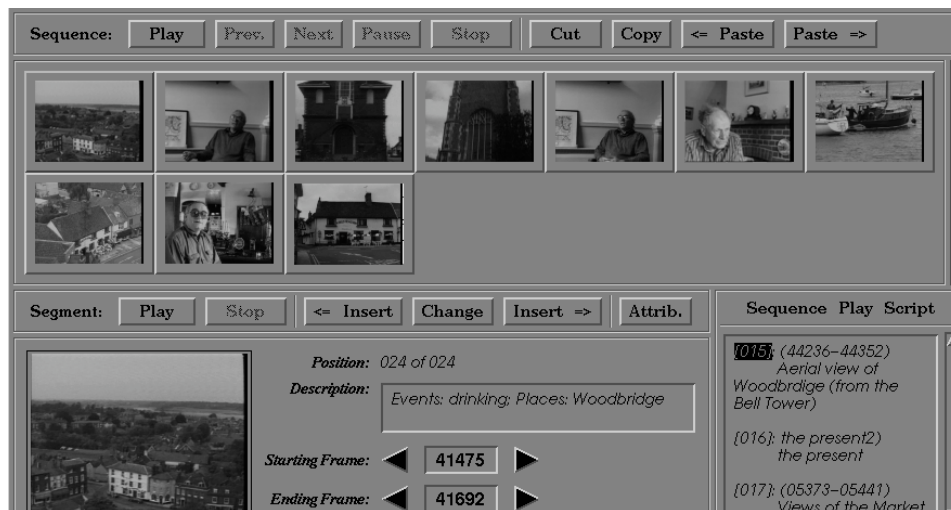

Figure 4. The Sequencer

As the tool set developed it became clear that the process of database creation and annotation is intimately linked to the potential uses of the database in question. In our travel database, we are interested in short stories which allow the end-user or client to get a feel for Woodbridge or Boston. These short stories should help the travel agent close deals on travel arrangements. Because the process of story creation is closely linked to use, as we fill a story model with appropriate shots, we can use this model to create new references in the database. In fact our current research suggests that in the future, most annotations will most appropriately be made either automatically or by story being reflected back into the database description. It is the idea of a database with an imbedded story structure that allows for convivial styles of browsing story.

Conclusion

While the technology to enable the creation of video databases is beginning to bear fruit, the question of how to effectively use these databases is still a thorny problem. Video stories are meaningful structures made from relatively less meaningful parts. Video editing is time consuming and requires strict adherence to library type tracking systems. It is a simple matter to describe video and make it available for recombination in a digital database environment. The challenge is to provide tools which allow even unsophisticated users to extract meaningful stories from such a database of video.

The tool set described in this paper is dramatically different from contemporary video edit systems in that it uses the concept of story models to structure an annotated video database around story. Story models in Homer can be used to generate stories from a database of video. These same models can be used to apply descriptions to video. These descriptions are more useful than conventionally logged descriptions because they are grounded in story.

If future story based systems are to be practical tools for travel agents or video researchers, the tools will need a stronger understanding of the world of stories. Causal relationships and story conflicts need to be incorporated into the representation. Framer, the knowledge representation language that underlies our current generation of tools, may allow us to create a system with the necessary domain knowledge to better understand and produce story.

References

Aguierre Smith, Thomas G. and Glorianna Davenport. "The Stratification System: A Design Environment for Random Access Video." In *Workshop on Networking and Operating System Support for Digital Audio and Video in San Diego, CA*, ACMYear.

Bordwell, David. *Narration in the Fiction Film*. Madison, Wisconsin: The University of Wisconsin Press, 1985.

Davis, Marc. *Media Streams: An Iconic Visual Language for Video Annotation*. MIT Media Lab, 1993. To be presented at the 1993 IEEE/CS Symposium on Visual Languages, Bergen, Norway.

Haase, Kenneth. "Framer: A Persistent Portable Representation Library." In *AAAI-93 in* Year.

MacKay, Wendy and Glorianna Davenport. "Virtual Video Editing in Interactive Multimedia Applications." *Communication of the ACM* 32 (7 1989): 802-810.

Morgenroth, Lee. *Homer: A Story Model Generator*. B.S. Thesis, MIT, 1992.

Pentland, A., Picard, R. W., Sclaroff, S. "Photobook: Tools for Content-Based Manipulation of Image Databases." In *SPIE Conf. Storage and Retrieval of Image and Video Databases II in San Jose, CA*, Year.

Pincus, Edward, and Steven Ascher. " "Picture Editing" chapter 11." In *The Filmmaker's Handbook*, New York: New American Library, 1984.