VIDEO MOSAIC: LAYING OUT TIME IN A PHYSICAL SPACE

Wendy E. Mackay

Rank Xerox Research Center, Cambridge Lab
61 Regent Street
Cambridge, CB2 1AB ENGLAND
+44 223 341 502
mackay@europarc.xerox.com

Daniele S. Pagani

Rank Xerox Research Center, Grenoble Lab
6 chemin de Maupertuis
38240 Meylan FRANCE
+33 7661 5084
pagani@xerox.fr

ABSTRACT

Paper video storyboards are still in use by even very experienced video producers with access to the most advanced video editing software. An analysis of the characteristics of paper and on-line editing provide an overlapping but distinct set of benefits (and problems). Paper provides the user with the ability to lay out various temporal sequences over a large spatial area and the ability to quickly sketch, annotate and rearrange the relevant video clips. On-line editing provides users with the ability to generate and store a variety of video arrangements. Video Mosaic provides users with the ability to combine the best of both worlds: elements of a paper video storyboard are used as input to an on-line video editing system to take advantage of the best aspects of each. We developed a Unix and a Macintosh version of Video Mosaic. This paper describes the design of Video Mosaic, compares alternative approaches to creating this type of application, and suggests directions for future work.

KEYWORDS: Video editing, Augmented reality, Storyboards, Paper user interfaces.

1 INTRODUCTION

We are interested in the problem of effectively creating and managing temporally-based information, such as video. Although a number of effective user interfaces have been developed to help users create and manage static data, such as text and graphics, tools for editing dynamic data are still in their infancy. Computers can provide us with flexibility in editing and changing dynamic data, but current tools tend to concentrate on the production of a final version, rather than on supporting the user in the development and exploration of ideas. Information with a temporal component presents a unique set of problems: How does the user "see", in one glance, data that is by its very nature dynamic? How can a user "sketch" ideas and evaluate different alternatives, without being forced into a commitment about the final presentation? How can the user make the transition from informal representations of the information to the final result?

Creative people have been developing tools to support the creation of temporally-based presentations for decades, even centuries. Musicians use a highly developed annotation system that provides a spatial representation of multiple strands of music as they change over time. An experienced musician can quickly jot down musical ideas that he or she "hears" internally. The composer can then work with live musicians or, more recently, computer-based synthesizers in order to hear and compare different alternatives. Similarly, choreographers use an annotation scheme to represent steps by various dancers in a ballet. Most choreographers combine this technique with working with live dancers, in order to experiment, see and compare different alternatives. In this century, cinematographers have developed techniques for visualizing the film and "sketching" their ideas, through use of a *storyboard*. Storyboards provide a spatial representation of temporal information, namely the sequence of motion events as they unfold over time. In each case, there is a critical interaction between the information recorded informally in a spatial layout on paper and the resulting dynamic work that occurs over time.

Today's on-line video editing systems provide flexible access to video and audio clips for the creation of interactive video programs. Yet despite this flexibility, many video producers continue to use paper storyboards for initially sketching out their ideas and communicating them to their colleagues. What is it about paper storyboards that makes them so useful? Is it simply a question of improving the interface to CRT-based video editing software or is there something fundamental about the flexibility of paper that makes video producers reluctant to give it up? The next sections describe the advantages and disadvantages of storyboards and on-line video editing systems and then discuss an approach that will enable users to move smoothly back and forth between the spatial layout of the storyboard and the temporal presentation of the resulting video.

The next section provides background information about storyboards, on-line video editing systems, and augmented reality. We then describe the underlying system, based on the EVA video-analysis architecture (Mackay, 1988, 1989) and a variation of the DigitalDesk (Wellner, 1992, 1993), and discuss basic functions that must be supported. Next, we describe two alternative designs, implemented on Unix and a Macintosh, and compare the differences. We also describe a different application which illustrates how the basic idea can be extended in a variety of directions, beyond the basic storyboard application. Finally, we discuss the advantages and disadvantages of this approach and conclude with a discussion of future research.

Proceedings of Multimedia '94, San Francisco, CA: ACM.

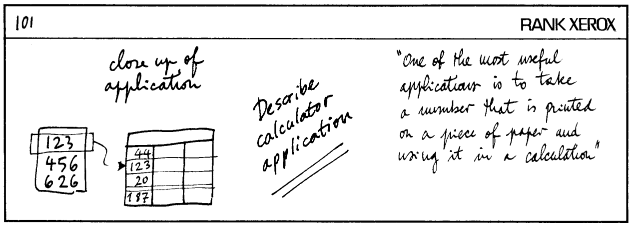

Figure 1: Part of a hand-drawn storyboard.

2 BACKGROUND

2.1 Storyboards

A storyboard is a paper tool that helps video producers and film-makers create video tapes, documentaries, movies, etc. Storyboards outline the action and capture the key elements of the story. Like a comic book, the storyboard shows a sequence of rough sketches of each action or event, with any accompanying dialog (or subtitles) and related annotations such as notes about what is happening in the scene, the type of shot (e.g. pan or zoom), and the type of edit.

Producers use storyboards to refine their ideas, generate 'what if' scenarios for different approaches to a story, and communicate with the other people who are involved in creating the production (e.g. camera, sound and actors or 'talent'). Some storyboards are very informal "sketches" of ideas which include partial information and are generally created before any video has been shot (see Figure 1). Other storyboards follow a pre-defined format and are used to direct the production and editing of the final material, such as that shown in Figure 2.

Paper storyboards provide a limited spatial representation of something that is fundamentally linear. Storyboards make it easy to jot down notes and get a quick overview of a lengthy visual presentation. If the elements of the storyboard are placed on separate cards, the designer can easily experiment with different linear sequences and insert or delete video clips with ease. However, paper storyboards pose some problems:

- *No link between paper and video*. Once the paper storyboard is created, the producer has no easy way to take advantage of the storyboard to create the video production. Systems that provide 'on-line' storyboards generally require the producer to work with a more finished version than that provided by ordinary pencil and paper.
- *Hard to visualize the final result.* Although an experienced video producer is able to visualize an action sequence from a set of still drawings, it may be difficult to communicate the idea to others, not only the people involved in the production but also potential clients or audiences.
- *Difficult to rearrange storyboard elements*. Once the elements of a storyboard have been placed on a series of pages, the resulting document becomes difficult to rearrange. It must be copied, cut into sections, rearranged, and then recopied in the new sequence.
- *Difficult to search for a particular element*. Especially for long presentations, the producer must search through the entire document in order to find a particular image or subtitle. It is also difficult to find all possible examples of particular types of shots (e.g., a zoomed in establishing shot) or edits (e.g., an l-cut).

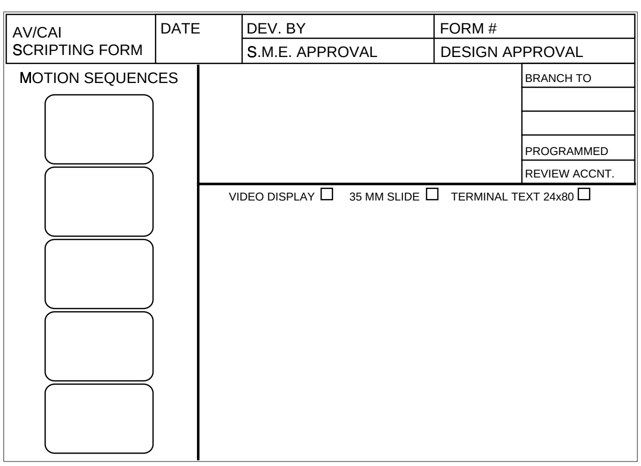

Figure 2: Storyboard form designed to assist developers at Digital to create interactive multimedia software.

2.2 On-line editing systems

On-line video editing systems use a variety of approaches. A common approach is based on hypertext, in which nodes or "chunks" of video are linked to others, such as Video Noter (Trigg, 1989), Firefly and MAEstro. The February, 1994 CACM special issue describes this approach. A contrasting approach concentrates on the relationships among streams of video, such as EVA (Mackay, 1989) Athena Muse (Hodges and Sasnet, 1993) and Streamer (Elliot, 1993). The October 1989 SIGCHI Bulletin discusses how these approaches differ. A third approach models the characteristics of analog video editing systems, (e.g. Editflex, Montage, and Editdroid) which enable editors to preview multiple edits before actually recording them. Apple's QuickTime provides a common infrastructure for managing video and audio clips, which is used by more sophisticated video editing applications such as Adobe Premiere and Avid VideoShop. Matthews et al. (1993) describe a direct manipulation interface for video editing and Weber and Poon (1994) describe a pen-based interface for video logging. These systems have advantages and disadvantages, depending upon the specifics of the application. All are designed to be controlled from a keyboard or electronic pen and display video on a CRT screen or video monitor. Video sources include cameras, videotape, videodisc or files with digitized video.

At first glance, it appears that adding a computer screen as an interface to video editing equipment would solve most of the above problems. Indeed, issues such as rearranging video clips and finding text associated with video become significantly easier with an on-line system. However, on-line editing systems pose their own problems. Despite the explicit transfer of most of the functions of paper storyboards and analog editing functions to computers, on-line video editing systems still seem to lose something in the translation. The following are some problems with on-line video editing systems:

- *Lack of screen real estate*. This is one of the most problematic issues. Even large CRT monitors have very little room for multiple video images. Video images must be small in order to fit and may be cumbersome to move.
- *Difficult to compare different versions*. One of the most frustrating aspects of video editing is trying to examine and compare collections of video clips to each other. A computer can manage access to a variety of video clips, but limited screen real estate makes it very difficult to see these video clips at the same time.
- *Poor video quality*. Although we can expect that faster hardware and better algorithms will result in improved video

quality, current systems provide small quantities of low quality video. The analog world is moving to higher-resolution HDTV and demand is increasing for managing larger numbers of video clips simultaneously, which has created a moving target for improving the quality of digitized video.

- *Lack of portability*. Users must make tradeoffs between size and portability. Despite the decreasing weight of laptop and palmtop computers, paper continues to be much lighter and easier to handle.
- *Difficult to annotate*. Unlike paper, the user must use a text editor or graphics program even to make a quick note or sketch.

Some of these limitations, such as video quality, may be short-lived. However, many of the limitations are inherent in the CRT-based model of computing. Large-screen systems, such as the Xerox Liveboard, may begin to address the screen real estate problem but are not portable. Similarly, palm-top computers may make it easy to transport the information, but don't address the real-estate problem. Our solution is to take advantage of the best characteristics of each system, paper and electronic, through an approach called "enhanced" or "augmented" reality.

2.3 What is Augmented Reality?

Most computer users live in two overlapping, but largely separate worlds consisting of paper and electronic systems. In a special issue of CACM, Wellner, Mackay and Gold (1993) introduce an approach that seeks to enhance the real world with computation:

> *Computer-Augmented Environments merge electronic systems into the physical world instead of attempting to replace them. Our everyday environment is an integral part of these systems; it continues to work as expected, but with new integrated computer functionality.*

The idea is to enable users to continue to use existing skills to manipulate these objects, while enhancing them with the power of the computer. The July 1993 special issue of CACM provides an overview of this approach and presents a variety of techniques and applications. Augmented reality borrows techniques created in other fields, particularly virtual reality (Rheingold, 1991) and image recognition. Techniques for identifying marks on paper are also very useful, such as the glyphs developed for Xerox's Paperworks so that paper can act as an interface to a computer via a fax machine.

Mackay et al. (1993) look specifically at paper as an object that can be augmented along a number of different dimensions. Paper is essentially static: once written or printed, the marks on the paper do not change (except, of course, if changed by hand through erasure or white-out). Yet we are often interested in changing the information on paper, in different ways. Numbers can be changed along a "computational" dimension. So, a mathematical formula becomes more useful if it can be used for active calculations. Similarly, a static table of numbers becomes more useful if it is amenable to "what-if" speculations on a spreadsheet (Wellner, 1991). Text can be transformed along several dimensions, e.g. translated into other languages (Newman and Wellner, 1992), checked for spelling, or analyzed for meaning, grammar or writing style. Hand-drawn sketches can be formalized. Two-dimensional drawings can be rendered into three dimensions (Carter, 1993).

Storyboards attempt to represent time-based action sequences of film or video in a static form, on paper. Video Mosaic uses an augmented reality approach to extend the static information that appears on paper and provide it with an active temporal dimension. Although Video Mosaic is designed to support video editing, the techniques described here will work with any time-based information.

3 VIDEO MOSAIC

Video Mosaic evolved from two research projects at RXRC Cambridge (EuroPARC). The first is EVA, the Experimental Video Annotator (Mackay, 1989a, 1989b), which enables users to link streams of multimedia data (video, audio, text, button presses, etc.) and edit and change the relationships among them. EVA was initially developed to enable researchers to make live annotations of usability study sessions, which could later be used to facilitate the analysis of the video data. The second is Wellner's DigitalDesk (1991, 1993) and related projects such as Marcel (Newman and Wellner, 1992), Mackay et al. (1993) and Carter (1993) which use paper as an interface to various on-line devices, such as calculators, electronic dictionaries and drawing tools.

Video Mosaic is an attempt to combine the power of paper video storyboards with the full capabilities of video editing software. Paper is portable, easy to annotate and change, and easy to lay out over a large area to compare alternatives and view relationships. On-line editing systems facilitate searching, make it easy to generate alternative sequences, store and reference related on-line (and off-line) data, and can sometimes handle multiple video sources. Video Mosaic is designed to take advantage of the best features of both, to enable users to move smoothly between temporal and spatial views of video and other time-based data. Users can annotate and manipulate their individual storyboards as well as share them electronically with others, checking how different segments fit together or commenting on each other's work. Users can make associations among different media (video, audio, subtitles, notes, and control information) at a particular point in time. They can also make associations among different clips over time, either in a linear fashion or in more complex arrangements, such as within a hypermedia document.

Perhaps the best way to describe Video Mosaic is to begin with a description of the system, from a user's point of view. The next section describes a video editing problem faced by a video producer, and the ways in which she uses a storyboard to work out her ideas. The following two sections describe the set of technical requirements which must be addressed by Video Mosaic in order to solve her problem and the user interface challenges that this approach raises.

3.1 User Scenario

A video producer has been asked to create several new video tapes. The first will be submitted to a conference and must act as a stand-alone description of her system. This tape will be approximately 12 minutes long and will use both audio and video. The second will be used as part of a summary of the activities of her lab, in conjunction with other brief descriptions of other projects. This tape will be approximately 2 minutes long and will also use both audio and video. The third will be used as a backup for her own talks, essentially video illustrations of her work. This tape will be about 5 minutes long and will only use video.

The producer already has a number of video tapes, with clips that have been stored on-line in digitized form. She also has ideas for a number of video clips that have not yet been shot. She begins by creating a hand-written storyboard for each of the tapes. She uses Video Mosaic to project a template onto her desk for each storyboard element. She identifies a video clip either with a rough, hand-drawn sketch of the action (which is recorded by Video Mosaic) or by selecting a on-line video clip. She then writes the text for the soundtrack (using a keyboard) or selects text from an on-line script if the video has already been shot. Finally, she makes some hand-written notes about what's going on, whether the video already exists, the best type of shots, etc. which is also recorded by Video Mosaic. Video Mosaic assigns each element a unique identifier and associates it with a storyboard.

As she works, the producer uses Video Mosaic to identify and track individual storyboard elements. She points to a set of elements in a particular order and asks the system to play them back for her. Video Mosaic retrieves the corresponding sequence of video clips (if they exist) or her rough sketches (if the clips haven't been shot yet), with the associated text and annotations and displays them on the table next to her paper storyboard elements. She experiments with different storylines for presenting the material. She includes and deletes different elements, creates new elements, subdivides existing elements, and views various arrangements. When she's satisfied, she asks Video Mosaic to save the particular storyboards and print them for her in a linear form, on several sheets of paper. She then shows these storyboards to her colleagues and asks for their comments. They scribble some notes and make suggestions for new clips to shoot. She goes back to Video Mosaic, which records the new hand-written annotations and lets her modify the clips and related text. She searches through the text subtitles for all instances of a particular event and compares them to select the best one. She then records and digitizes the new video clips and replaces the rough sketches with the new material. When she is ready, she tries different arrangements for each of the three video presentations she needs to create. In some cases, she reuses the same element and in others, she chooses related but different elements to present the ideas. When she is satisfied with each, she asks Video Mosaic to edit video versions from each storyboard and print them for her.

In order to support this scenario, we need to create two basic objects: a storyboard and a storyboard element, which have both electronic and paper representations. The storyboard, which specifies a video production created by editing and composing several video clips, must include the following basic information:

- Unique identifier
- Author and contact information
- List of related storyboard elements

A storyboard is composed of 'elements', one for each video clip to be edited, that include the following information:

- Unique identifier
- Video 'best frame' (from an existing video clip) or a hand-drawn sketch of the action.
- Script of the clip, subtitles or ASCII text
- Hand-written note to specify the scene, shot type, etc.

These components comprise the minimum necessary for Video Mosaic. However, it should be easy to add other components to support related applications. For example, a musical score could be added with a playable audio track. Or eye-movements or keystroke logs that were recorded with the video during a usability study could be added as an associated data track. Key words that identify particular events could be used for later searching and identification of particular clips.

The next section describes the basic functions that must be supported by Video Mosaic. These functions are described in general terms and different specific solutions can be found for each, depending upon cost and the requirements of the specific application.

3.2 Technical design requirements and options

3.2.1 Identification of paper storyboard elements

Each storyboard element has both a paper and an electronic form. Video Mosaic needs a method of uniquely identifying each storyboard element and the storyboard (or storyboards) with which it is associated. Note that if a storyboard element is 'deleted' from a particular storyboard, it continues to exist in its paper form and may be useful later. Thus we need to be able to track the storyboard elements separately from the storyboards, and yet provide easy methods of linking them to each other, once they exist in their paper forms. It must be possible for the system to detect each individual paper element or storyboard.

Once a paper element or storyboard has been provided with a unique identifier, it must be detectable by Video Mosaic. This implies either very sophisticated image recognition software and hardware, or simpler techniques for detecting the identification mark from the paper, such as barcodes, glyphs, or optical character recognition of alphanumeric codes. The barcode reader is a simple and inexpensive option, but requires the system to print a barcode on each paper element. Glyphs, developed at Xerox PARC contain more information in less space than a barcode, are less susceptible to variations in position, and can be read either by a hand-held scanner or a video camera. User-readable codes can be detected via optical character recognition software and a scanner or video camera, however this tends to be less reliable and is very susceptible to changes in orientation.

3.2.2 Recognition of user commands

Users must be able to easily issue commands to Video Mosaic to perform functions such as playing a particular storyboard, capturing a new hand-written annotation, or printing out a sequence of storyboard elements. Various techniques are possible here. One method is to use a small vocabulary of spoken commands and a voice recognition system. Another alternative is to use 'paper buttons' (Wellner, 1991) which are placed under a camera and recognized with an optical character recognition system. (A microphone can be used to detect when the user taps on the table, to indicate when the user is issuing the command.) The advantage of this approach is that it uses the same technology for identifying storyboard elements and issuing commands. A graphics tablet or other device to specify the x,y coordinates of where the user is pointing can be used in conjunction with projected menus of commands. Another alternative is to provide commands via barcodes, which can be read with a barcode reader.

3.2.3 Capture of user annotations

The user must be able to make comments on the storyboard and easily record them for later use. A low resolution alternative is to grab a video image from a video camera. A higher resolution alternative is to use a regular or a hand-held scanner.

3.2.4 Display of information to the user

The user must be able to watch the video clips as they play, with all of the associated information. One method is to mount a video monitor into the desk itself, which can provide a high-quality video display. A projector mounted overhead can display the related text and graphic information next to the monitor. Several projection systems can be used, from large-screen systems to lightweight LCD projection screens designed for use with overhead projectors. The projector must be able to display information from the computer and, if a separate video monitor is not used, must also be able to display video clips.

3.2.5 Storage of video information

The user must be able to store a variety of video clips. A write-once analog videodisc player provides a flexible, high-resolution (but expensive) method of storing video information. High quality video cards, such as Parallax, can display the resulting video via the workstation screen. A cheaper, but slower and lower-resolution alternative is to use a video-tape player for video storage and lower-quality video cards, such as VideoPix. A number of digitized video alternatives exist, such as QuickTime for the Macintosh and various video compression cards for Unix workstations. These involve compressed video and are currently slower and lower resolution than the analog alternatives, but will increase in quality over time.

3.3 User interface design issues

Substitution of different technologies to accomplish similar functions must be evaluated not only from a cost and ease-of-implementation standpoint, but also from the impact it will have on the user. Some characteristics of the system create both new possibilities for interaction and also new limitations. These issues are more difficult to identify and require further research before these techniques are practical for everyday use.

3.3.1 Video overlays

Users of Video Mosaic come to the system with two sets of different expertise. When using computer window systems, they expect that placing a video window on top of another window will obscure the lower window and make it possible to see the window on top. Similarly, if working with paper, the papers that appear on top obscure the papers below. If we project video or other information from below the desk, as in the embedded video monitor, this principle applies as the user expects. However, the system behaves oddly (from the user's perspective) when video is projected on top of paper. Basically, the user sees both the underlying information on the paper and the projected image.

For some applications, this is precisely what is required. For example, a user might have a still image and want to see what it looks like if animated. Seeing one superimposed upon the other provides a useful contrast. However, sometimes the information is not related and it is simply difficult to see.

3.3.2 Erasing information

On-line systems make it easy to erase or remove information without a trace. Paper requires a physical action (such as using a rubber eraser on a pencil drawing or using "white-out" to superimpose a white mark on top of printed text). What does it mean to erase when these two environments are mixed? What happens when information is erased from paper, but not the electronic version, or vice versa?

3.3.3 Issuing commands

Most commands are issued as single words or selected as items from a menu. This provides a very simple, but also limited user interface. We can provide users with more flexible input, either by adding traditional input devices, such as keyboards, or through hand-writing or gesture recognition techniques as they improve.

4 SYSTEM OVERVIEW

We created two versions of Video Mosaic, one in a Sun/Unix environment and the other in an Apple Macintosh environment. These versions allowed us to explore various approaches to providing the functions in the architecture and took advantage of features unique to each hardware and software platform. This section provides more specific details about how each version was implemented and then compares the different approaches.

4.1 Version 1: Unix Video Mosaic

Figure 3 shows a picture of the Unix version of Video Mosaic. Note the use of the video camera, positioned over the desk and the video monitor, embedded in the desk surface.

4.1.1 Hardware and software setup

Video Mosaic V1.0 (Unix) comprises the following equipment:

1. *Video camera mounted above the desk.* Provides several types of input to the system, including: grabbing a handwritten annotation for inclusion in the storyboard, identifying a 'button' command, and identifying a particular storyboard or element. (Note: another version uses the video camera to identify where the user is pointing).
2. *LCD video projector mounted above the desk.* Projects a computer screen including text, annotations and other information from the storyboard as well as menu commands, such as "print" and "play".
3. *Video TV monitor (11.5x9 cm) installed in the desk.* Provides a high-quality image of the video from a storyboard. The video projector projects text and annotations on either side of the monitor.
4. *Desktop ink-jet printer.* Provides the user with a quick means of generating new paper storyboard elements.
5. *Write-once video-disc recorder and player.* Provides an efficient means of storing and retrieving video under computer control.
6. *Microphone mounted below the desk.* The user taps on the

Figure 3: The Unix version of Video Mosaic

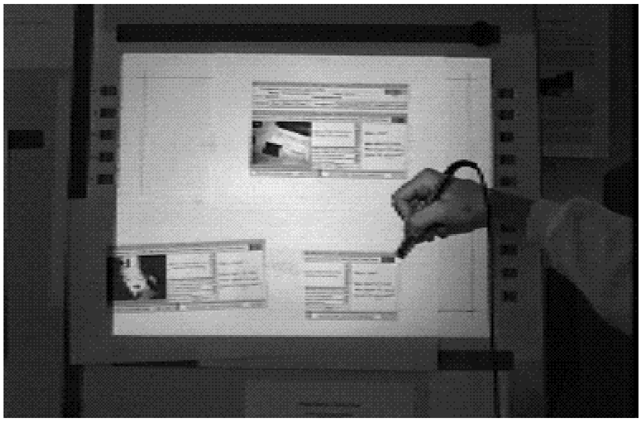

Figure 5: The Macintosh version of Video Mosaic

desk to identify when the camera should grab a new image.

7. *Sun Unix Workstation*. Runs the video editing software and controls the hardware devices. The software is written in C and X-11 under Unix. Additional software includes the Xerox ScanWorX optical character recognition system.

4.1.2 Storyboards and storyboard elements

The user creates and edits storyboards by interacting with a storyboard header, used to identify the storyboard, and several storyboard elements (one for each video clip). Figure 4 shows a typical storyboard element, associated with a video clip. It includes a video best frame when printed on paper or moving video when used on-line (left); a handwritten annotation, that can be digitized with the camera for the on-line version (middle); a text editor for entering the text of the script of the clip or comments (right); buttons for issuing commands (top).

4.2 Version 2: Macintosh Video Mosaic

Figure 5 shows a picture of the Macintosh version of Video Mosaic. The hardware setup is somewhat different, in particular, the method of identifying where the user is pointing and the issuing of user commands.

4.2.1 Hardware setup

Video Mosaic V2.0 (Macintosh) comprises the following equipment:

1. *Video camera and associated pointer*. Detects where the user is pointing with a red light source (LED, laser pointer or barcode reader wand) and transforms the location into x,y screen coordinates.
2. *LCD projection panel (640x480 pixels, 2 million colors), associated with an overhead projector*. Projects video, text, annotations and other information from the storyboard as well as menu commands, such as "print" and "play".
3. *Desktop ink-jet printer*. To print out the storyboard elements corresponding to each video clip.
4. *Hand-held scanner*. Provides the user with a quick means of capturing sketches on storyboard elements.
5. *Video digitizing and compression board*. Provides an efficient means of digitizing, compressing and storing video on the computer.
6. *Barcode reader*. Allows the user to identify storyboard elements and to issue commands.
7. *Apple Macintosh*. Runs the video editing software and controls the hardware devices. The software is written in HyperCard 2.2 with external commands (XCMDs) written in Pascal and takes advantage of QuickTime to handle video.

4.2.2 Storyboards and storyboard elements

The user begins a session by creating a new storyboard. Figure 6 shows a typical storyboard header, which includes: information about the storyboard (name, author, date, Subject Matter Expert approval, design approval); a barcode to identify the storyboard; commands to create, delete and copy storyboard elements; commands to play, reorder and print a storyboard. The reordering command allows the user to change the sequence of clips to be played by selecting the corresponding paper elements with the bar code reader in the desired order.

Figure 7 shows an example of a MacIntosh Video Mosaic element. The video component (left) allows the user to select and view the video clip. The "File" button allows the user to select a video segment already on the disk, whereas the "Record" button digitizes a new video clip from camera or tape. The "In" and "Out" buttons permit the user to select the first and last frame of the subsection of the clip to be used in the production. "Best" allows the user to select the preferred frame for printing. The subtitle component (middle) allows to enter text synchronized with the video. When the video runs, the corresponding subtitle is displayed. When clicking on the "Show list" button, the list of subtitles is displayed for editing: clicking on a subtitle line moves the video to the corresponding time and shows the times when the subtitle is displayed ("Start") and deleted ("Stop"). The user can edit the text of the subtitle at any time and can change the timing of the subtitle either by clicking the "Start" and "Stop" buttons or by

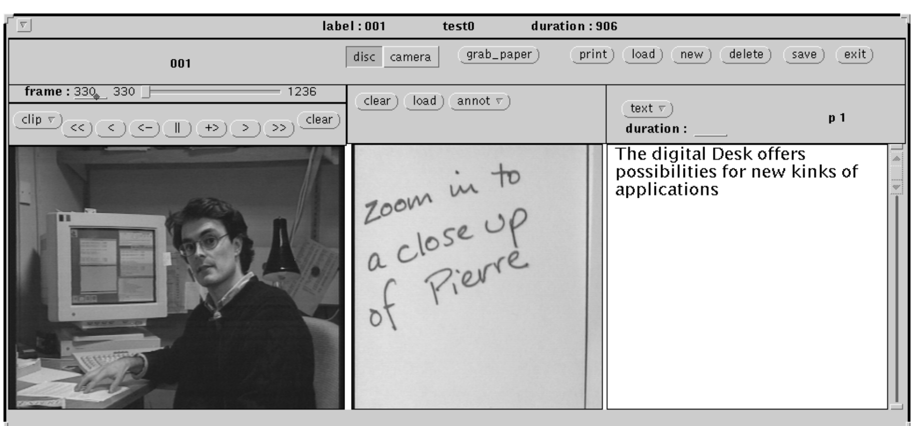

Figure 4: A Video Mosaic storyboard element (Unix version)

moving the rectangles in the timeline at the bottom. The left component of the element displays a handwritten annotation (same as Unix version). The barcode is used to identify the element once it is printed on paper.

The Unix version was less usable than the later MacIntosh version. Some problems were technical, such as requiring the user to place storyboard elements under a stationary camera instead of pointing at them (with a bar code reader) as layed out on the desk. A more indepth discussion of usability alternativies is beyond the scope of this paper.

5 ALTERNATIVE USE OF VIDEO MOSAIC

Most of the discussion so far has been centered around the support of a video producer to edit and create a video sequence to be delivered on traditional media such as videotape. However, the Video Mosaic approach suggests a wider range of applications than just video editing. A simple example is the problem of using printed paper documents that contain video, audio, animations or dynamic graphics. Currently, a user must choose between viewing the document on-line, which makes reading the text very tiring but permits viewing of dynamic material, or reading a printed version of the document, which is much more portable and much easier to read, but can only provide a still image from a video clip and no sound or animation. Video Mosaic makes it easy to use a combination of paper and electronic presentations of a multimedia document, taking advantage of the benefits of each.

An example of the effective use of Video Mosaic for this purpose was the creation of multimedia field notes for our EuroCODE research project. This project involves the creation of a distributed media space for engineers responsible for the construction of the world's longest suspension bridge, in Denmark. Our research activities include visits to the various sites at the bridge, where we interview people and observe their daily work practices. In addition to taking written notes, we collect various artifacts and videotape people and the work they do. Upon our return to the lab, the various members of the group summarize the results of these interviews, digitize the video, scan the paper artifacts and produce a 50+ page document (our field notes), which we share and use as input to our design. Fifty pages is too long to read on-line. On the other hand, the small, still frame images from the field clips are generally inadequate as illustrations of activities at the bridge site.

We found that both versions of Video Mosaic could be used to help solve the problem. We added an identifier next to each video image that appears in the text (a number for the Unix version, a barcode for the Mac version) and provided a 'control page' at the back of the document to issue commands and identify the members of the project. Furthermore, we took advantage of our existing media space (Gaver et al., 1992) in which each office has a video camera, microphone, and monitor for displaying video, which are connected to the Raven media space controlled from a Sun Unix workstation.

The reader of the field notes can read the printed version of the field notes anywhere. Viewing the full video clips and making connections to the live media space requires being in the office. The user simply indicates the video clip (either with the camera or the barcode reader) and Video Mosaic selects the clip and plays it. The user can also turn to the control page to establish live audiovisual connections with another member of the team through the media space and discuss the document.

6 CONCLUSION AND FUTURE WORK

This paper presents the Video Mosaic system, which enables users to use paper storyboards as a means of controlling and editing on-line video. The system is based on the EVA architecture, which uses a stream as the underlying primitive in order to relate video to other time-based information, either on or off-line. It would be easy to design a system where paper storyboards are linked to an on-line hypermedia system, with a corresponding set of advantages and disadvantages.

One can easily imagine a range of uses for a paper interface to a multimedia system. The perspective of "dimensions" along which paper can be extended, in the context of augmented reality, provides a useful method for thinking about new applications. At RXRC Cambridge, we have experimented with computational, editorial, and graphical extensions, as well as temporal. We can also think of additional kinds of temporal extensions. Thus a musical score can be played and edited using these same techniques. Similarly, sets of data (e.g. information about weather patterns) can be displayed dynamically as a form of scientific visualization. Cartoonists could treat a sequence of hand-drawn cartoons on paper as 'key frames' and watch the computer-generated animation proceed smoothly from one

Figure 6: A storyboard header (Macintosh version)

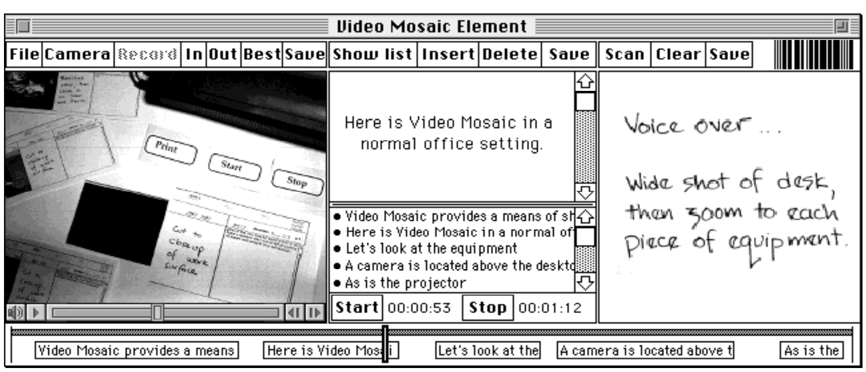

Figure 7: A storyboard element (Macintosh version)

keyframe to the next. As described earlier, there are numerous possibilities for using information on paper to generate live connections, e.g. telephone calls or video conferences.

Educational uses might include "Dynamic Flashcards", in which students could work with ordinary paper flashcards, but obtain access to dynamic information, answers or a live video connection to a teacher, based on their activities. (Of course, some educational applications already exist, such as children's books that use a barcode reader to let the child hear parts of the story.) Another use might be the layout of an office or kitchen. Sometimes, working with small paper representations of different objects (sink, dishwasher, refrigerator) is the easiest way to view and arrange the system. But such a system would also benefit from being able to project alternatives, make copies of desirable layouts and display different views (e.g. a 3d view walking into the kitchen). This is the sort of system that could be combined effectively with a virtual reality system, to provide both a tactile, easily manipulable paper version and a computer-generated virtual version. Yet another interesting possibility is the area of exploratory data analysis of multimedia data. Experimenters must discover patterns in associated streams of data, which may include video, text, codes, numerical data, and various other forms of temporal records. Much of this information is printed on paper, which makes it easier to handle and scan. However, discovering patterns and associations among different types of data is often better accomplished with the computer. One could imagine a Video Mosaic style of interface that enabled researchers to work with both forms of the data together.

Video Mosaic is a first step in exploring the relationship between informal, paper-based representations and highly-interactive, electronic systems for managing temporal data. We have presented some solutions for some of the technical and user problems, but clearly additional work is required to make such systems truly effective. The key is to think about the characteristics of each medium: For what purposes is paper best suited? For what purposes is a computer and/or computer network best suited? Are there situations in which each has characteristics that are beneficial, and using only one or the other provides a less useful system? If the answer to the last question is 'yes', it makes sense to explore the concept of a paper-based interface to the computational system.

Our future work involves continuing development of the EVA architecture, which will enable us to expand the video, audio and other editing capabilities of the system, within a distributed environment (and supporting both live and stored video). We will also continue to investigate a variety of issues with respect to paper-based interfaces, not only new technologies for providing input and display capabilities but also variations in the forms of the user interface. Our hope is that this approach will provide a more accessible way to creatively 'sketch' and explore ideas for highly interactive, temporally-based presentations.

ACKNOWLEDGMENTS

Special thanks to Gilles Velay and Chaoying Ma, who programmed sections of the Unix version of Video Mosaic. Pierre Wellner contributed invaluable software, advice and ideas. Glorianna Davenport inspired some of these ideas through our discussions about the problems of on-line video editing systems. Michel Beaudouin-Lafon contributed discussions and a review of the manuscript.

REFERENCES

(1993) Carter, K. Computer-Aided Design: Back to the Drawing Board. In *Proceedings of Creativity and Cognition*, Loughborough, April.

(1993) Elliott, E.L. Watch-Grab-Arrange-See: Thinking with motion images via streams and collages. M.S. dissertation, MIT, Cambridge, MA.

(1992) Gaver, W., Moran, T., MacLean, A., Lovstrand, L., Dourish, P., Carter, K. and Buxton, W. Realizing a video environment: EuroPARC's RAVE system. In *Proceedings of CHI'92 Conference on Human Factors in Computing Systems.* (May 3-7, Monterey, CA), pp. 27-35. ACM, NY.

(1993) Hodges, M. and Sasnet, R. Multimedia Computing: Case Studies from MIT Project Athena. Addison-Wesley, M.A.

(1988) Mackay, W. EVA: Experimental Video Annotator. Video presentation at *CHI '88 Conference on Human Factors in Computing Systems*. (May, Washington, D.C.), ACM, NY.

(1989) Mackay, W. EVA: An Experimental Video Annotator for Symbolic Analysis of Video Data. In *SIGCHI Bulletin*, Vol. 21(2). Special Issue: Video as a Research and Design Tool. October.

(1989) Mackay, W., and Davenport, G. Virtual Video Editing in Interactive Multi-Media Applications. In *Communications of the ACM*, July, Vol. 32, No. 7.

(1993) Mackay, W., Faber, L., Launianen, P. and Pagani, D. Design of the High Road Demonstrator,*D4.4*, 30 September 1993, EuroCODE ESPRIT project 6155.

(1993) Mackay, W., Velay, G., Carter, K., Ma, C., and Pagani, D. Augmenting Reality: Adding Computational Dimensions to Paper. In *Communications of the ACM*, July 1993, Vol. 36, No. 7, pp. 96-97.

(1993) Matthews, J., Gloor, P., Makedon, F. VideoScheme: A Programmabnle Video Editing System for Automation and Media Recognition. In *ACM Multimedia 1993.* (1-6 August), Anaheim, California. pp. 419-426.

(1992) Newman, W. amd Wellner, P. A desk supporting computer-based interaction with paper documents. In *Proceedings of CHI'92 Conference on Human Factors in Computing Systems.* (May 3-7, Monterey, CA), pp. 27-35. ACM, NY.

(1991) Rheingold, H. *Virtual Reality*. Summit Books. New York, N.Y.

(1989) Trigg, R. Computer Support for Transcribing Recorded Activity. In *SIGCHI Bulletin*, Vol. 21(2). Special Issue: Video as a Research and Design Tool. October.

(1994) Weber, K. and Poon, A. Marquee: A tool for real-time video logging. In *Proceedings of CHI'95 Conference on Human Factors in Computing Systems.* (April 24-28, Boston, MA), pp. 58-64. ACM, NY.

(1992) Wellner, P. The DigitalDesk calculator: Tactile manipulation on a desktop display. In *Proceedings of UIST'92, the ACM Symposium on User Interface Software and Technology*. (Nov 11-13, Hilton Head, SC), ACM, NY.

(1993) Wellner, P., Interacting with Paper on the DigitalDesk. In *Communications of the ACM*, Vol. 36, No. 7, pp.86-96.

(1993) Wellner, P., Mackay, W., and Gold, R. (Editors), Computer Augmented Environments: Back to the Real World. In *Communications of the ACM*, July, Vol. 36, No. 7, p 24-26.