A Paper-Based Interface for Video Browsing and Retrieval

Jamey Graham and Jonathan J. Hull
Ricoh Innovations, Inc.
Menlo Park, CA
{jamey, hull}@rii.ricoh.com

Abstract

A paper-based interface for browsing video is proposed. A paper document shows key frames selected from a video, a transcript for the parallel audio track, and bar codes that, when scanned, invoke a multimedia player. The paper document provides a stand-alone representation for a video recording that lets a user both understand the content of the file and replay only selected parts of the multimedia that are necessary to gain a better understanding. This approach applies the two-dimensional display characteristics of a newspaper to multimedia retrieval. By so doing, the user's browsing and search efficiency is greatly improved. This ***poster*** *describes an implementation of the Video Paper system using a Pocket PC with a bar code reader as the remote control device and an archive of TV programs on the Pocket PC or an external server.*

1. Introduction

Video is difficult to browse and search because it is essentially a one-dimensional medium. When we watch a video, at any given time we can only see a small portion of the available information. It's a challenge to visualize the surrounding context. While there are many elegant online solutions for this problem (e.g., [2, 3, 8]), they can be difficult for new users to understand.

Newspapers, on the other hand, are familiar to everyone and are designed for easy browsing. Paper is a low cost, high resolution display medium with many advantages, including portability, low power requirements, and usability in almost any environment. Layout rules, developed over hundreds of years, provide a two-dimensional representation that allows a user to perceive massive amounts of information in a single glance. The structure of the text, including carefully chosen titles and well engineered paragraphs, combined with photos that attract a reader's attention, help people decide which stories they might be interested in and how much time they should devote to reading them.

We set out to apply the principles of newspaper design to the problem of video browsing and understanding. We wanted to provide a paper representation for a video recording that was "stand-alone" in the sense that a reader could understand as much as possible about a video by merely glancing over the document. However, that document should also include an easy-to-use means for replaying portions of the video so that the user could see and listen to the multimedia recording whenever a more in-depth explanation was required. Ideally, the need for selective replaying would be reduced as much as possible since every time the user has to play the video to search for information, they would be back in the mode of one-dimensional search and passive uptake that we are trying to minimize.

Our solution for interacting with multimedia content such as a TV broadcast or a recorded video is called *Video Paper*. This system includes paper versions of the multimedia content that include a text representation for the audio (we use the closed caption when available), formatted with guidelines designed for the newspaper industry. We use fonts chosen for their readability at a small point size, multiple columns and short lines of text. The line spacing, bolding and capitalization techniques also contribute to improving the ability of the user to skim the document.

We also display key frames extracted from the video at various locations. Figure 1 depicts a sample of the proposed Video Paper interface. Bar codes refer to corresponding points in the recorded video. Swiping a bar code causes the video to begin playing at that time. This allows users to read the paper document and view only those parts of the video that are relevant to their needs. (Glyphs [5] could also be embedded in the key frames for similar functionality.) Given a multi-paged document of this type representing, say, an hour long TV program or recorded meeting, a reader can quickly skim the contents of the program to see if anything relevant might be present in the text.

Poster Submission: ICME 2003, Baltimore, MD, July 6-9, 2003.

Figure 1. Example Video Paper document.

An example of using Video Paper for analyzing a news broadcast, like the one shown in Figure 1, would be looking for a story about Idaho. A quick glance over that page lets a user to see the map in the lower left corner and focuses his attention on that section, thus eliminating the need to look elsewhere in the document.

2. System Architecture

The Video Paper system architecture is shown in Figure 2. Television programs are recorded from satellite transmissions, broadcast sources, or cable. The recording process preserves the closed caption content. The server converts the video data to MPEG2 format and generates the video paper representation.

Video paper contains key frames and a formatted version of the closed caption. A bar code is generated for every key frame that identifies the video recording and the position of the key frame within the video. This is so that the action of scanning the bar code can cause the video replay to begin at the corresponding position.

A remote control device is comprised of a PDA (e.g., Compaq iPAQ) with a bar code reader and a wireless interface (e.g., 802.11b). Software on the PDA decodes scanned bar codes and sends commands to the server that controls replay of the video on a television attached to the video rendering card on the server. In addition to the bar codes attached to key frames, meta control bar codes are included that pause the replay, rewind, fast forward, or display the closed caption on the television.

We also developed a portable version of the video paper system in which the MPEG2 video data is written on a PCMCIA disk drive that is added to an iPAQ together with a bar code reader. A modified version of the remote control software invokes the video replay on the iPAQ instead of on a separate television. This allows the Video Paper system to be used in places where there's no network connection, such as on a train.

3. Key Frame Selection Algorithm

The algorithm for key frame selection is an important part of maximizing the usefulness of video paper documents. Following a clustering step, video frames are input to a routine that recognizes the presence of a fixed

3. Experimental Results and Applications

A complete working prototype of the video paper system has been developed that includes a PC server that stores recorded videos and generates video paper. Running on a 2 GHz Windows 2000 system, less than one minute of run-time is required to produce a video paper PDF file for a one-hour recording. The MPEG2 video file can also be trans-coded to a low bit-rate version for the portable system. These files typically require about 100 MB per hour of recorded data. Thus, up to 50 hours of video can be saved on a 5GB PCMCIA disk.

The remote control shown in Figure 2 is connected to our WiFi LAN and allows users to replay any of the video paper documents on demand. We also developed a version of video paper for our Meeting Recorder [8]. This system captures 360-degree panoramic video and presentation slides while they're being displayed in our conference room. The Video Paper representation includes key frames from both sources and a transcript for the audio track.

Video Paper is also being used for oral histories [7]. These are interviews of the observers or participants in historical events. Typically, the interviews are recorded and transcribed. Because of the difficulty in accessing the audio, subsequent analysis is mostly limited to reading the transcript. Oral historians have started video taping their interviews, but these are even more difficult to access afterward because of the difficulty in synchronizing the video replay with the transcript.

A recent experiment subjected Video Paper to 13 subjects who reviewed two 75-minute interviews. They were given a list of questions and used Video Paper to find the correct answers. Post-hoc analysis of their experiences showed they were completely at ease with the concept of Video Paper and reported that it provided substantial benefits with minimal overhead. They used the audio to hear particularly compelling passages and used the video to help get an idea about the personality of the interviewee.

Interestingly enough, several people established a multi-tasking search strategy in which they read one part of the transcript while listening to the audio from another. This unintended result demonstrates the usefulness of the two-dimensional display provided by Video Paper.

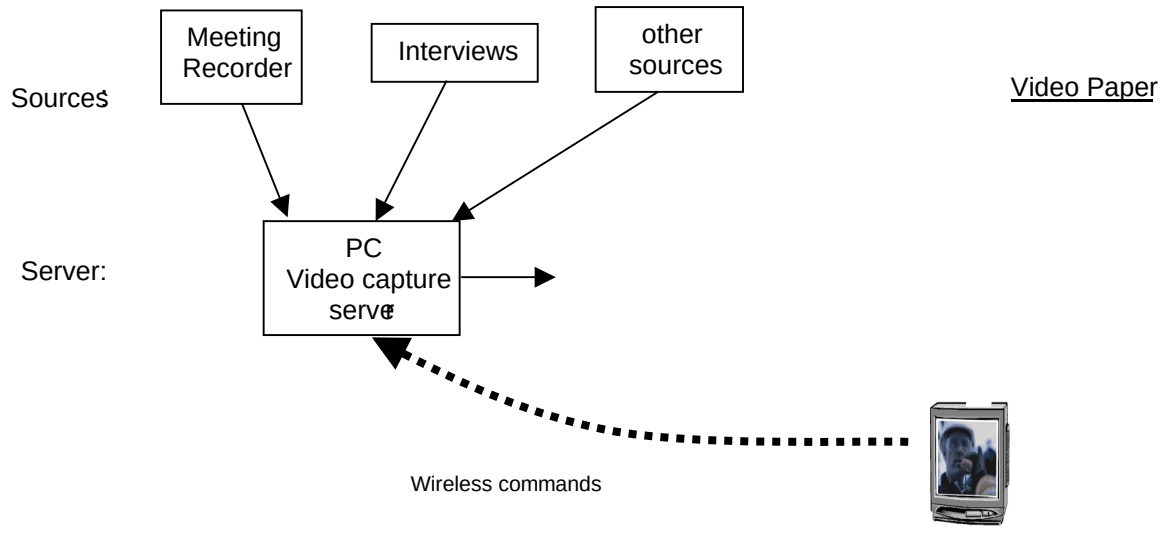

Figure 2. Video Paper System Architecture

6. Related Work

Paper-based representations that include links to electronic information provide fertile ground for exploring new multimedia retrieval technologies. In spite of its similarity to several recent efforts, Video Paper is different from all of them. For example, in one case users manually add links to text strings in paper documents that can retrieve the representation for that text in another language [1]. In another instance, users take notes on paper attached to a device that simultaneously records audio [9]. Subsequently, one can page through the notepad and replay the audio that was captured when a notation was written.

Another example allows users to manually select interesting key frames while a video is playing or a presentation is given [5]. A paper representation shows those images on a time line with a machine-readable code that can activate the video replay.

All of these techniques contrast sharply with our emphasis on *automatically* producing a paper representation for a multimedia recording that a person can understand without necessarily using an electronic device. Our documents also have bar codes that can trigger a multimedia player to start at any randomly selected point.

The layout of a Video Paper document is carefully designed so that a reader can easily browse and access a long video recording. An online solution with similar characteristics provided an html interface to TV programs [8]. However, there was no provision for paper-based access to the multimedia data in this system.

7. Conclusions

A novel solution for video and retrieval was proposed and evaluated that uses the characteristics of newspapers and paper-based displays to provide an efficient and easy-to-use method for finding information in videos. Recent results with news broadcasts and recorded meetings confirm the intuition of the system's developers that Video Paper is an ideal interface for browsing and selectively accessing recorded video.

An application to oral histories showed how Video Paper can provide easy access to multimedia data and thereby enriching the reader's experience in ways that were not possible with previous technology.

References

1. Arai, T., Aust, D., and Hudson, S. E., "PaperLink: A technique for hyperlinking from real paper to electronic content," Conference on Human Factors in Computing Systems (CHI 1997), 327-334.

2. Boreczky, J., A. Girgensohn, G. Golovchinsky, and S. Uchihashi, "An Interactive Comic Book Presentation for Exploring Video," Conference on Human factors in Computing Systems (CHI 2000), pp. 185-192, 2000.

3. P. Chiu, A. Kapuskar, S. Reitmeier, and L. Wilcox, "Notelook: Taking Notes in Meetings with Digital Video and Ink," ACM Multimedia '99, Oct. 1999, Orlando, FL.

4. J. Graham, "The Reader´s Helper: A Personalized Document Reading Environment," Proceedings of the CHI 99 conference on Human factors in computing systems: the CHI is the limit, May 1999, 481-488.

5. D. L. Hecht, "Printed Embedded Data Graphical User Interfaces," IEEE Computer, March 2001, 47-55.

6. S. Klemmer, J. Graham, G. Wolff, and J. Landay, "Books with voices: Paper transcripts as a physical interface to oral histories," ACM Conference on Human Factors in Computing Systems (CHI-2003), Fort Lauderdale, Florida, April 5-10, 2003.

7. D. S. Lee, B. Erol, J. Graham, J. J. Hull, and N. Murata, "Portable Meeting Recorder," ACM Multimedia 2002, Juan Les Pins, France, Dec. 1-6, 2002, 493-502.

8. B. Shahraray and D. C. Gibbon, "Automated Authoring of Hypermedia Documents of Video Programs", *ACM Multimedia 95*, November 5-9, 1995, San Francisco, CA.

9. L. Stifelman, B. Arons, and C. Schmandt, "The Audio Notebook: Paper and Pen Interaction with Structured Speech," ACM CHI Conference, Seattle, WA, March 31 – April 5, 2001, 182-189.