# The Video Paper Multimedia Playback System

Jamey Graham, Berna Erol, Jonathan J. Hull, and Dar-Shyang Lee
Ricoh Innovations, Inc.
2882 Sand Hill Rd., Suite 115
Menlo Park, CA 94025
+1-650-496-5700
jamey@rii.ricoh.com

## ABSTRACT

Video Paper is a prototype system for multimedia browsing, analysis, and replay. Key frames extracted from a video recording are printed on paper together with bar codes that allow for random access and replay. A transcript for the audio track can also be shown so that users can read what was said, thus making the document a stand-alone representation for the contents of the multimedia recording. The Video Paper system has been used for several applications, including the analysis of recorded meetings, broadcast news, oral histories and personal recordings. This demonstration will show how the Video Paper system was applied to these domains and the various replay systems that were developed, including a self-contained portable implementation on a PDA and a fixed implementation on desktop PC.

## Categories and Subject Descriptors

H.3.4 [**Information Systems**]: Information Storage and Retrieval – *systems and software.*

## General Terms

Algorithms, Performance, Design, Experimentation.

## Keywords

Paper-based multimedia browsing, retrieval, access, and replay.

## 1. INTRODUCTION

The analysis of multimedia recordings is a challenging task, largely because the obvious solution of watching the recording from the beginning, perhaps taking notes along the way, requires as much time as the length of the recording. Essentially, this explains why so much multimedia content is never used -- it is just too difficult to browse. Over the years, researchers have addressed this problem with online interfaces (e.g., [5]) that display key frames from the video and provide random access as well as the ability to search the transcript. Other work has exploited features of the original recording to improve the utility of an online interface. Video Manga is one example that varied the size of key frames based on an importance measure [6].

Video Paper is an alternative solution for multimedia skimming and retrieval [1]. Text from the closed caption (or a transcript) is displayed together with key frames extracted from the video. A bar code is printed underneath each key frame that, when scanned, plays the video from the corresponding point in time. This allows users to read the paper document and view only those parts of the video that are relevant to their needs. Given a multi-page Video Paper document representing an hour-long meeting or TV program, a reader can quickly skim the content and determine whether there is anything worth listening to in the multimedia recording.


*MM '03*, November 2-8, 2003, Berkeley, CA, USA.
ACM 1-58113-722-2/03/0011.

## 2. SYSTEM DESIGN & APPLICATIONS

The Video Paper system includes a digital multimedia recording process and a postprocessing step that produces the paper document. Users can access the multimedia data on a local area network with a PDA. The PDA reads bar codes and sends commands to a server that control replay of the video on a television attached to a video rendering card on the server. In addition to the bar codes associated with key frames, meta bar codes are included on the Video Paper document that pause the playback, rewind, fast forward, or display the closed caption text on the television.

We also developed a portable version of the Video Paper system in which the video data is written on a small media card that can be inserted in the PDA. A modified version of the control software invokes the video replay on the PDA instead of the television. This allows the Video Paper system to be used in places where there's no network connection, such as on a train, an airplane, a car, etc.

A version of the Video Paper system is shown in Figure 1. The recording process uses a video camera with a 360-degree lens that records a meeting. This is combined with a four-microphone audio localization system that identifies where each speaker was positioned with respect to the camera. This information is used to compute key frames that show who was speaking at each second during a meeting. A presentation recorder captures a separate set of key frames from the slides that were shown on a projector [3].

Postprocessing software chooses key frames from the video and presentation streams to fill the space near the text transcript. Presentation slides that were shown for more than a minimum duration are guaranteed placement on the page. The other key frames are chosen from the images of the speaker that were calculated based on audio localization. An example of a document created by this process is shown in Figure 2.

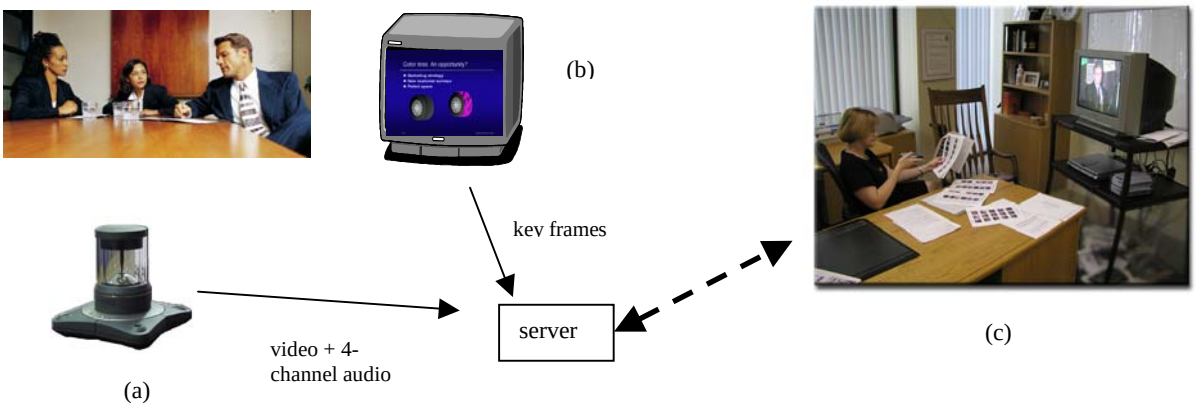


**Figure 1.** The meeting recorder (a) captures video and four-channel audio. The presentation recorder (b) saves key frames from the video shown on a projector. A networked server supports Video Paper access to the recorded data (c).

The Video Paper system has also been tested in our laboratory for other applications, including the analysis of broadcast news, where we observed that the paper representation was similar to a newspaper in the sense that a user could understand what had been presented by skimming the document and only watching selected sections of the video recording. This result was extended in an application to oral histories [4] which showed that Video Paper improved the efficiency of researchers by letting them read-ahead while listening to another section of the recording. It also provided a convenient means for accessing multimedia recordings that previously required the user to load a VHS tape into a VCR and manually fast-forward to a given time-stamp. Now the same work is accomplished by scanning a bar code.

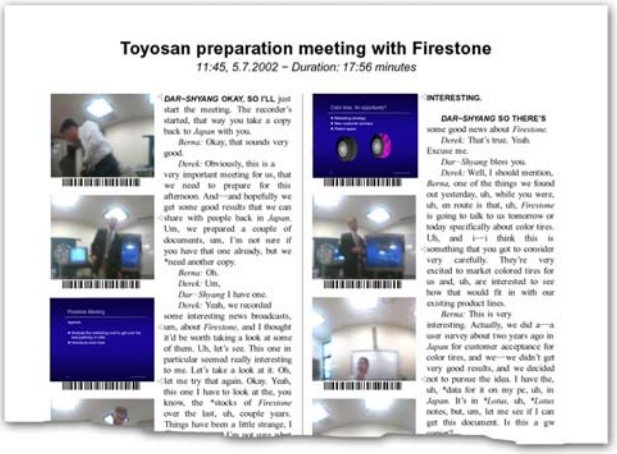


**Figure 2.** A Video Paper representation for a recorded meeting.

## 3. CONCLUSIONS

Video Paper has proven to be an efficient and easy-to-use method for multimedia skimming, access, and retrieval. Various applications, such as replaying recorded meetings, have been developed.

## 4. REFERENCES

[1] J. Graham and J. J. Hull, "Video Paper: A paper-based interface for skimming and watching video," International Conference on Consumer Electronics (ICCE), Los Angeles, June 16-18, 2002, 214-215.

[2] D. L. Hecht, "Printed embedded data graphical user interfaces," IEEE Computer, March 2001, 47-55.

[3] D.S. Lee, B. Erol, J. Graham, J.J. Hull and N. Murata, "Portable meeting recorder," ACM Multimedia 2002, Juan des les Pines, France, Dec. 1-6, 2002, 493-502.

[4] 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.

[5] B. Shahraray and D. C. Gibbon, "Automated authoring of hypermedia documents of video programs", *ACM Multimedia 95*, November 5-9, 1995, San Francisco, CA.

[6] S. Uchihashi, J. Foote, A. Girgensohn, and J. Boreczky, "Video manga: Generating semantically meaningful video summaries, ACM Multimedia 99, Orlando, FL, 383-392.