Automatically Linking Multimedia Meeting Documents by Image Matching

Patrick Chiu, Jonathan Foote, Andreas Girgensohn, John Boreczky

FX Palo Alto Laboratory
3400 Hillview Ave, Bldg 4
Palo Alto, CA, 94304, USA
E-mail: *lastname*@pal.xerox.com

ABSTRACT

We describe a way to make a hypermedia meeting record from multimedia meeting documents by automatically generating links through image matching. In particular, we look at video recordings and scanned paper handouts of presentation slides with ink annotations. The algorithm that we employ is the Discrete Cosine Transform (DCT). Interactions with multipath links and paper interfaces are discussed.

KEYWORDS: Automatic linking, video indexing, image matching, scanning, paper interfaces, meeting capture, multimedia

INTRODUCTION

Advances in digital multimedia technology—especially video—have generated interest in capturing meetings, presentations, and lectures (e.g. see [2], [6], [7]). Browsing and reviewing these recordings, however, remain challenging problems.

We describe a way to make a hypermedia meeting record from multimedia meeting documents by automatically generating links through image matching. These are documents that were created for a meeting and not expressly for the purpose of the final hypermedia meeting record. In particular, we look at video recordings and paper handouts of presentation slides with ink annotations. Because these documents have coherent structures by themselves and are intuitive to navigate, they are natural components for the hypermedia record.

In contrast, a customary approach is to author a hypermedia meeting record by creating a new document structure and taking and recombining pieces of meeting documents. See [1], [4], [6]. While this may produce better results, it requires much more effort and a certain amount of design skill.

Another common method is to tightly integrate the contents of the meeting documents as time streams in a multimedia application. See [7], [9]. When the various streams are synchronized, the resulting structure is a simple one that is linear along time, and does not afford complex non-linear or multipath linking. This method also does not work with paper documents.

EXAMPLE: A WORKSHOP MEETING

We give a scenario to illustrate our approach. The meeting documents are: (1) video recording of the meeting shot by a camera operator, (2) scanned paper handouts of presentation slides by the workshop participants. The handouts may contain ink annotations and notes made during the meeting.

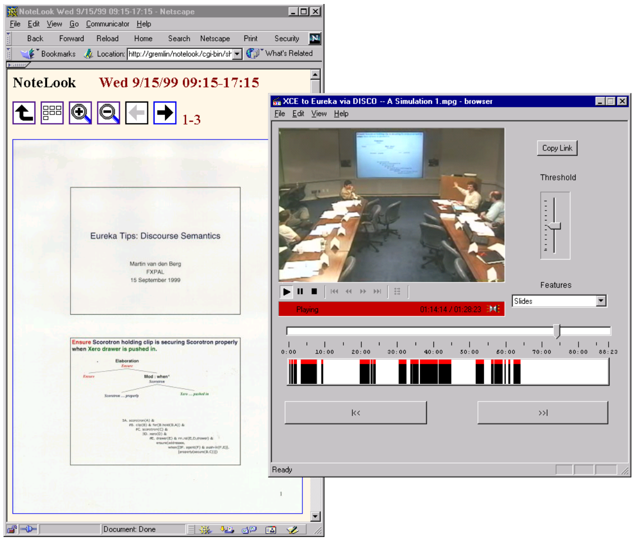
Beyond the obvious presentations by one speaker with a single set of handouts, we describe a more complicated example of a real meeting that we supported. The meeting was an all-day affair that involved several research groups. Six presenters made handouts. The way that the discussion progressed was by turn taking, in which each presenter showed a slide on the issue being talked about. After the meeting, the six sets of handouts were scanned in and manually linked to the video recording (the image matching algorithm was not yet ready).

The hypermedia meeting record consists of a table of contents page with links to these six sets of handouts, plus the video recording (in MPEG and RealVideo). While linked together, each document component also stands on its own. See Figure 1. With this system, it is very easy to find the place in the video where a certain slide by a certain person was discussed. Otherwise, one would be faced with the dreaded task of browsing through 6 hours of video.

IMAGE MATCHING

The video and handout images are matched by looking at image similarity. There are two possible sources for the video images. The first is from the video recording of the meeting, where it is common practice for the camera operator to shoot full screen images of the slides along with the speaker and room activity. A second source is from a video projector used for displaying the presentation slides, which may be recorded for the purpose of image matching.

Final version. To appear in *Proceedings of Hypertext 2000.*



NoteLook Wed 9/15/99 09:15-17:15 - Netscape

Location: http://gremlin/notelook/cgi-bin/sh

NoteLook Wed 9/15/99 09:15-17:15

Eureka Tips: Discourse Semantics

Martin van den Berg
FXPAL
15 September 1999

Ensure Scorotron holding clip is securing Scorotron properly when Xero drawer is pushed in.

XCE to Eureka via DISCO -- A Simulation 1.mpg - browser

Playing 01:14:14 / 01:28:23

Features: Slides

Figure 1: Clicking the lower slide image on the Web page plays the video recording at the corresponding point when that slide was discussed during the meeting. The page is from one of the six sets of scanned handouts.

The image matching algorithm that we have developed is the Discrete Cosine Transform (DCT) with truncated coefficients [3]. Other algorithms may be used (e.g. [7]). The DCT has compact data representations for each image (a video can have many thousands of frames). Truncated DCT coefficients are insensitive to small differences in image translation, rotation, or scale as they can happen while scanning paper handouts. DCT works on black & white or color images.

Ink strokes are ignored because thin lines contribute only to the discarded high-frequency coefficients. After finding the image matches, the ink strokes can be recovered by doing a comparison between the original and annotated images. When more than one set of handouts have been annotated, these ink strokes may be extracted and selectively displayed on top of the common background slide image.

We conducted an experiment in which we matched screen images captured during six staff meetings against the video recordings of those meetings. We scaled down the luminance component of the images to 64x64 pixels and used the 256 lowest-frequency DCT coefficients for the image matching. The match against full-screen slide images in the video was almost perfect, even with the screen images rotated by five degrees (to simulate scan errors) or with ink annotations superimposed on the images.

MULTIPATH LINKS

Sometimes a slide may be visited or displayed multiple times in the course of a meeting. For example, during question and answer period, the speaker may go back to a slide referred to in a question. To handle this, multipath links may be used (see [8]), in which clicking on a handout image provides the user with a menu of different times of video playback points. Alternatively, a set of single path links may be used to represent a multipath link.

PAPER INTERFACES

It is possible to use a paper interface for playing back the video recording from the handout images—after all, they were originally *paper* handouts. This can be achieved with Xerox Glyph technology, which are like 2-d barcodes (see [5]). The playback times deduced from image matching are encoded into Glyphs and printed as a light gray background over which the handout images are also printed. A GlyphPen device enables users to "click" on the images on the paper printouts to activate an attached video player.

ACKNOWLEDGEMENTS

We thank Cathy Marshall and Lynn Wilcox for their valuable comments.

REFERENCES

1. Auffret, G., Carrive, J., Chevet, O., Dechilly, T. Audiovisual-based hypermedia authoring: using structured representations for efficient access to AV documents. *Proc. of Hypertext '99*, pp. 169-178.

2. Chiu, P., Kapuskar, A., Reitmeier, S., Wilcox, L. Meeting capture in a media enriched conference room. *Proc. of CoBuild '99*. Springer-Verlag LNCS 1670, pp. 79-88.

3. Girgensohn, A., Foote, J. Video classification using transform coefficients. *Proc. ICASSP '99, VI*, pp.3045-3048.

4. Hardman, L., van Rossum, G., Bulterman, D. Structured multimedia authoring. *Proc. of ACM Multimedia '93*, pp. 283-289.

5. Hecht, D. Embedded data glyph technology for hardcopy digital documents. *Proc. of SPIE Color and Hard Copy Graphics Arts III*, pp. 341-352. 1994.

6. Ip, H., Chan S. Hypertext-assisted video indexing and content-based retrieval. *Proc. of Hypertext '97*, pp. 232-233.

7. Mukhopadhyay, S., Smith, B. Passive capture and structuring of lectures. *Proc. of ACM Multimedia '99*, pp. 477-487.

8. Pearl, A. Sun's link service: a protocol for open linking. *Proc. of Hypertext '89*, pp. 137-146.

9. Trigg, R. Computer support for transcribing recorded activity, *ACM SIGCHI Bulletin*, 21 (2), 68-71. 1989.