SMART HANDOUTS: PERSONALIZED E-PRESENTATION DOCUMENTS

Berna Erol, Dar-Shyang Lee, and Jonathan J. Hull

Ricoh California Research Center
2882 Sand Hill Rd. Suite 115, Menlo Park, California, USA
{berna_erol, dsl, hull}@rii.ricoh.com

ABSTRACT

A novel system is described that significantly enhances the usefulness of handwritten notes taken during a presentation by creating a multimedia document that includes scanned images of handouts, personal notes, and links to a multimedia recording of the presentation. Notes are linked to the e-presentation media with automatic content analysis without any special notes capture device. Layout segmentation and template matching automatically detects the presence of presentation handouts during scanning. Presentation-level and slide-level linking of handouts to e-media use text and image features from slides. Experimental results show 95% accuracy in linking of the scanned handouts to the e-presentation media.

1. INTRODUCTION

Many e-presentation capture systems have been developed [1, 2]. These systems capture audio and video, presentation slides, notes, and whiteboard data, resulting in media-rich documents. Today, capturing multimedia data is easier than it has ever been and the real challenge is in making it accessible to users. Most of today's e-presentation systems support web-based access and provide cross-indexing between slides and an audiovisual recording of a presenter. While a web-based e-presentation playback interface can be useful for people who are motivated to search, browse, and access recorded presentations and lectures, e.g., a student right before an exam, it is not as useful in a corporate environment.

When people look for specific information or review a presentation, their own personal notes are often excellent starting points for retrieval. In our more than four years of research and regular use of an e-presentation capture system in a corporate environment, we observed that even though an electronic note taking system was available, users generally preferred to take notes on presentation handouts.

In this paper, we present a method that allows a user to take notes on presentation handouts with a regular pen, and after the presentation is finished, scan the handouts and invoke a process that automatically creates a pdf file that contains links to a multimedia recording of the presentation. Referred to as a *Smart Handout,* this is a personalized e-presentation document which shows the presentation slides, notes, metadata, and has media links to the electronic presentation recording as shown in Figure 1. When a user needs to review what occurred when a particular slide was presented, he can simply click on a slide or a video key frame in the Smart Handout. This invokes the web-based presentation playback interface and starts the playback from the time associated with the slide or key frame. Additional metadata present on the handouts, such as the Q&A activity and key frames, helps the user recall the presentation and efficiently navigate to the point that interests her. The fact that the user does not need to access an additional interface for searching for a particular presentation recording significantly reduces the effort needed to access e-media.

This paper describes a novel algorithm that automatically generates Smart Handouts. A new layout segmentation and template matching algorithm automatically detects whether a scanned document is a regular document or a presentation handout. A presentation matching algorithm, which is based on OCR and n-gram matching, retrieves the presentation recording where the slides in the handouts were presented. Linking of the scanned slides with slides captured in a presentation recording uses edge histograms. Users' handwritings on scanned slides is detected and segmented for better matching accuracy.

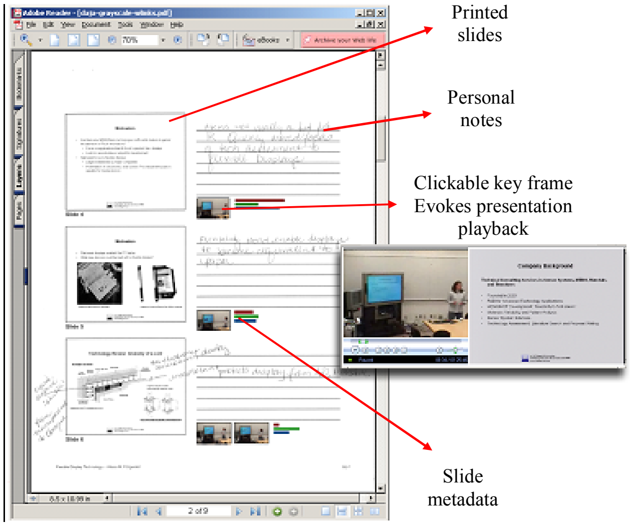


Figure 1. Example of a Smart Handout.

The rest of the paper is as follows. In the next section we give an overview of the related work in the literature. Section 3 presents our method for automatic generation of *Smart Handouts*. Experimental results and conclusions are given in Section 4 and Section 5, respectively.

2. RELATED WORK

There are e-presentation/e-learning capture systems described in the literature that allow cross-referencing of user annotations and notes [2]-[8]. In some of these systems, notes are entered on a computer or a PDA that timestamps each typed note [3][5]. For handwritten notes capture, Tablet PCs, PDAs, and special pads are used in systems such as [6][7]. These note-taking systems synchronize notes with the e-learning media, but require the use of specialized devices. These devices may not be accessible by all users at all times thus creating a usage barrier. In fact, a survey we performed of 21 participants revealed that

0-7803-9332-5/05/$20.00 ©2005 IEEE

the majority (>80%) of people prefer to take notes on paper presentation handouts rather than on a laptop or a PDA.

Our method automatically creates personalized e-presentation documents without using any specialized notes-capture devices. A similar presentation access method is suggested in [4]. However, they do not present experimental results on automatically matching handouts to presentation recordings. This leaves utility of the DCT-based matching method they suggest an open question, particularly for matching of handouts that are printed in black and white. In contrast, we present experimental results that show how our method is suitable for matching slides printed in black and white, grayscale, and color, demonstrating that it's suitable for a practical implementation. Furthermore, we propose solutions for handout detection and slide segmentation, which is not addressed by the prior art.

In our earlier work [8], we presented a system that printed bar codes on handouts before a presentation took place. A slide image-matching algorithm mapped the bar codes onto the times when the slides were displayed. Our previous method required a special printer that assigned unique barcodes to each slide and saved the original source file. The slide images captured during a recording were matched to the original presentation slides. In this paper, we eliminate the need for bar codes and using a special printer. As a trade off, we no longer have access to the original document for slide matching (i.e., the PowerPoint file). This problem is overcome by segmenting and matching the scanned slides directly to the recorded slides. This significantly improves the usefulness since our new method can be employed with any e-presentation system that saves slide images.

3. SMART HANDOUT CREATION

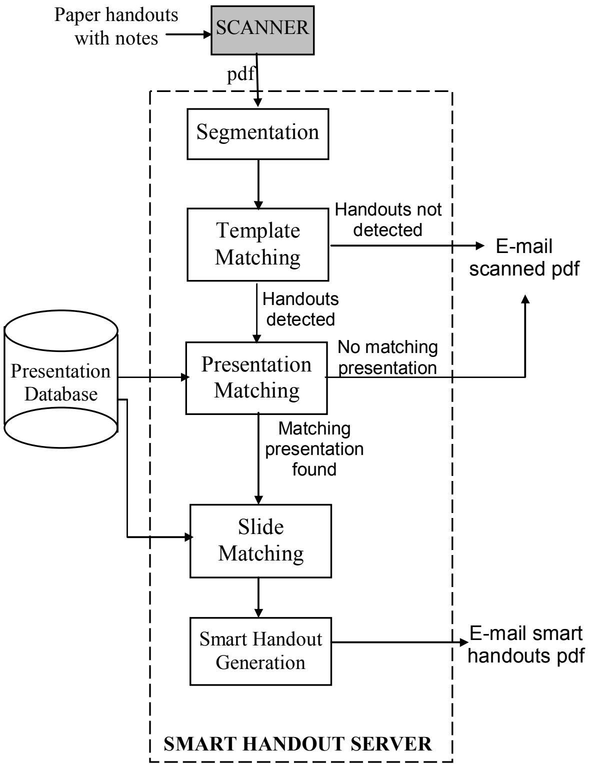


Figure 2. Smart Handout server processing

Our conference room is equipped with a PTZ camera, an omni-directional audiovisual capture device with 4 channel audio capture, a whiteboard capture system, and a presentation recorder. The *Presentation Recorder* (PR) captures the video output of a presenter's laptop as it's routed to a projector. A presenter's screen images are captured once a second and every captured image is time-stamped and saved if it is significantly different than the previously captured image. These images are synchronized to the captured video via time stamps. Each PR image is OCR'ed and indexed with the extracted text. Moreover, for each slide, metadata such as slide duration and audio activity, which is based on changes in sound source direction, are computed.

After a presentation, if a user would like to have an electronic version of her notes, she inputs her e-mail address and scans the handouts on a regular scanner. A pdf file is generated from the scanned pages and passed to the *Smart Handout Server* as shown in Figure 2.

The server converts individual pages in the pdf document into JPEG images. Then, segmentation is applied to each page to detect possible slide regions. Commercially available presentation document authoring software, such as PowerPoint^{TM} and FrameMaker^{TM}, support a limited number of layouts for printing handouts. Motivated by this, a template matching technique was developed that detects whether a scanned document is a presentation handout or a regular document. If the scanned document is not a presentation handout, the pdf containing only the scanned document is e-mailed to the user. Otherwise, the presentation matching step retrieves the relevant presentation recording, and matches slide images in the scanned handout to slides captured by the presentation recorder. Then, the scanned document is populated with e-media links and the resulting document is e-mailed to the user. The details of these processing steps are given in the following sections.

3.1 Segmentation

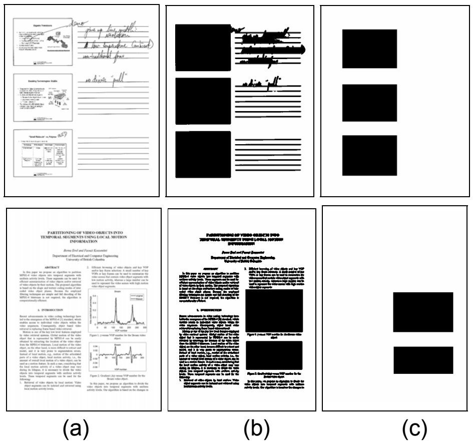


Figure 3. Slide candidate segmentation: (a) input images, (b) outer connected components, (c) candidate slide regions.

First a smoothing filter is applied to reduce half-toning effects that may occur after printing and scanning, followed by binarization with global thresholding. Examples of two scanned documents, one presentation handout document and one regular document, is given in Figure 3. Connected component analysis is applied to the document images to find the outer-most components, as shown in Figure 3.b. Slide handouts may contain regions that do not belong to the slides, such as a user's handwritten notes, a presentation title, and page numbers. Erosion with a resolution-dependent structuring element disconnects slide regions from any overlapping handwriting segments. Then features of each connected region, i.e., height, width, width-to-height ratio, and compactness, are analyzed to eliminate non-slide regions. Figure 3.c shows examples of segmented slide region candidates in two input documents.

Currently, our system requires the document scan be performed using the automatic feed of the scanner, instead of using manual page-by-page scan. This ensures that all scanned pages of a presentation handout are placed in a single pdf file. As a side benefit, the scanned pages have minimal skew. Nevertheless, if page by page scan is desired, skew correction should be performed prior to segmentation.

3.2 Template Matching

In this step, segmented slide candidate regions are compared against 6 commonly used presentation handout layouts: 1, 2×1, 3×1, 2×2, 2×3, and 3×3 slides per page. Each scanned page is represented by a feature vector, $S(P_n) = \{\vec{\mathrm{f}}_\mathrm{n}^1, \vec{\mathrm{f}}_\mathrm{n}^2, ..., \vec{\mathrm{f}}_\mathrm{n}^{\mathrm{m_n}}\}$, where P_n is the page n, m_n is the number of slide candidate regions on P_n, and $\vec{f}_n^i = \{\mathrm{w}_\mathrm{n}^\mathrm{i}, \mathrm{h}_\mathrm{n}^\mathrm{i}, \mathrm{cx}_\mathrm{n}^\mathrm{i}, \mathrm{cy}_\mathrm{n}^\mathrm{i}\}$, where $\mathrm{w}_\mathrm{n}^\mathrm{i}$ and $\mathrm{h}_\mathrm{n}^\mathrm{i}$ are the normalized width and height of a candidate slide region on page *n* and, $\mathrm{cx}_\mathrm{n}^\mathrm{i}$ and $\mathrm{cy}_\mathrm{n}^\mathrm{i}$ are the x and y coordinates of the slide region relative to cx_n^1 and cx_n^1, which are the coordinates of first region in the raster scan order. The same feature vector is also computed for each handout template T_x, $S(T_x) = \{\vec{\mathrm{r}}_\mathrm{x}^1, \vec{\mathrm{r}}_\mathrm{x}^2, ..., \vec{\mathrm{r}}_\mathrm{x}^{\mathrm{m_x}}\}$. Then, a directed distance between $S(P_n)$ and $S(T_x)$ is computed as follows:

$$d(P_n, T_x) = \sum_{i=1}^{m_n} \min_{\vec{r}_x^j \in S(T_x)} \left\{ \left\| \vec{f}_n^i - \vec{r}_x^j \right\| \right\}.$$

Some scanned pages, particularly the last pages, may contain fewer slides than that of the template T_x. Using the above distance measure, instead of a correlation based measure for example, ensures robust matching of scanned pages to the correct template regardless of the number of slide regions on a page. The matching handout template is found as the template that has the smallest directed distance to the scanned page. For a given input document, if 2/3rd of the pages match to the same template, the scanned document is identified to be a valid presentation handout. In that case, each slide in the handout is segmented out and numbered based on the page number and the raster scan order. This results in a collection of scanned slide images, *S={S1, S2, ... Sm}*. These images are used for presentation and slide matching as described in the following sections. If the input document does not match to any of the handout templates, then further processing is not applied and the server e-mails the scanned document to the user, without modifying its contents.

3.3 Presentation Linking

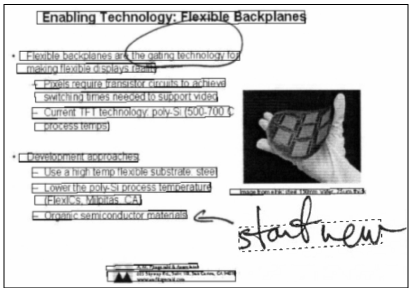
In order to retrieve the presentation recording session where the scanned handouts were presented, we employ n-gram matching. Scanned slides are OCR'd and n-grams are formed with words that contain at least 4 characters. The n-grams are compared to n-grams of the text extracted from each recording session in the database. More details of this algorithm can be found in [8]. If there is more than one matching presentation recording, which may occur if the same slides were presented in more than one presentation session, then the recording with the most recent date and time is selected as the matching presentation.

3.4 Mapping of Slides with Handwriting

The segmented slide images from scanned handouts, $\{S_1, S_2, ... S_N\}$, are mapped to the slide images captured by the *Presentation Recorder* in session *i*, $PR_i=\{I_{t1}, I_{t2},..., I_{tM}\}$. Since each PR image, I_{tm}, is time-stamped with *tm* at capture time, we can determine when each slide on the scanned handout was presented. Using this information, each slide is linked to the video stream.

We employ an edge histogram-based slide matching algorithm for finding matching I_{tm}'s to S_n. In [8], we employed a similar slide matching algorithm to map PowerPoint slides to PR images, which yielded a high accuracy. Here, we map scanned slide images to PR images. In this case the difficulty is increased because of image degradation caused by printing and scanning and the existence of handwritten annotations.

To improve the slide matching accuracy, we identify regions that potentially contain handwriting and exclude them from the matching process. Given a slide image, text-like regions are identified by finding strong edges with the Canny edge detector, smearing the edges with a 64x2 smearing filter, thresholding, and performing connected component analysis. The connected components that do not possess a specified height and width ratio are filtered out as non-text regions. Usually, text regions with handwriting are less horizontal than machine-printed text regions. Furthermore, because letters are connected in handwriting more often than in machine print, the average height-to-width ratio of connected components in a handwritten text region is much smaller than that of machine print. Motivated by these, we compute fitted line *Li* in the direction of the text region spread. We also compute $c_{av} = \frac{1}{N}\sum_{i=1}^{N} c_i$ and $c_i = \frac{h_i}{nc_i\ w_i}$, where N is the number of text boxes, h_i is the height, w_i is the width, and nc_i is the number of connected components (corresponding to letters) in text box *i*, respectively. Finally, the text boxes that do not have horizontal spread, $|\theta(L_i)| \geq 10^\circ$, or that have low height to weight ratio, $c_i < \frac{c_{av}}{2}$, are marked as handwriting text regions. An example of handwriting detection in a slide image is given in Figure 4.



Enabling Technology: Flexible Backplanes

- Flexible backplanes are the gating technology for making flexible displays reality
  - Pixels require transistor circuits to achieve switching times needed to support video
  - Current TFT technology: poly-Si (500-700 C process temps)
- Development approaches
  - Use a high temp flexible substrate: steel
  - Lower the poly-Si process temperature (FlexICs, Milpitas, CA)
  - Organic semiconductor materials

start new

Figure 4. An example of handwriting detection on a slide.

The detected handwriting regions are ignored during slide matching, yielding a significant improvement in the matching accuracy. Note that our method cannot be used to detect other user markings such as arrows, lines, etc. Nevertheless, since our method is based on edge histograms, these markings do not affect matching accuracy as much as the edge-dense handwriting segments.

3.5 Smart Handout Composition

Once each slide on a scanned handout is mapped to one or more images captured by the *Presentation Recorder*, the final pdf document is composed by including video key frames and media links. Recall that the PR images are time-stamped and they are synchronized with the video. First, video frames corresponding to these times are extracted from the video stream. Then, template matching results are used for determining the optimum positions for inserting video key frames in the scanned handouts. For each given handout template, the preferred locations for video key frames are designated in advance. Before inserting a key frame in the pdf file, luminance variance analysis is applied to the region to detect the presence of user's markings. If the luminance variance is lower than 16, then the key frame is inserted as an opaque image, if it is higher than 16, then, the key frame is inserted with 50% transparency so that the user's markings are visible. The Adobe Acrobat SDK is used for inserting key frames and links in the pdf file [9].

If the presenter shows a particular slide more than once then that slide is mapped to multiple PR images. In that case multiple video key frames are inserted for a given slide. Such an example is on the 3rd slide of the handouts shown in Figure 1. Each key frame is also linked to a CGI script that starts the video and slide replay in a web browser.

4. EXPERIMENTAL RESULTS

Our presentation database contains 343 recorded presentation sessions. The total number of screen images captured by the *Presentation Recorder* is 44958. All these images are time-stamped and indexed with the OCR output.

The test set is composed of 53 handout documents containing 1341 slides. These were printed before various presentations, distributed to several users for note taking, and collected afterwards for scanning. The handouts were printed using various layouts in PowerPoint (e.g., 2, 3, 4, 6, and 9 slides per page) in color, grayscale, and black&white. Some handouts contain slides with dark foreground on a light background and some others contain light foreground on a dark background. The paper handouts were scanned at 200 dpi binary on a device with an automatic sheet feeder.

The first set of experiments shows the accuracy of handout detection with template matching. The results are presented in Table 1. Handout detection was applied to 53 scanned handouts and 956 scanned document images in the UW database, which is a commonly used test set for document image analysis research. 100% of the handout documents and 99.7% of the regular documents are correctly classified. The documents that are incorrectly classified as handouts contain large tables that are similar in appearance to a PowerPoint handout layout.

Input doc. type	Number of documents	% of correct detection
Presentation Handouts	53	100%
Regular Documents	956	99.7%

Table 1. Handout detection results

After handout detection, each handout are matched against the 343 recorded presentations in the database using n-grams with n=2. The accuracy of presentation level matching is 100%. Once a presentation is identified that matches to a given handout, the segmented slide images are matched against all the Presentation Recorder images captured in that session. The retrieval results are presented in Table 2. 53 slide handouts are automatically segmented to obtain 1341 scanned slides. The total number of *captured presentation recorder images*, i.e. 6115, is significantly larger than the number of slides as these include many screen shots, such as videos and demos, which are not actual slides. The total number of *matching presentation recorder images* that are marked as ground truth items is 1474, which is still larger than the number of actual slides, because some slides are shown more than once during the presentations. The total number of *correctly matched presentation recorder images* is 1406, yielding a 95% overall matching accuracy. As can be seen from the table, the handouts with 4- and 6-slide layouts are retrieved with a higher accuracy, 96% and 97% respectively, then the other handouts. Based on our observations, this is because of two reasons: the segmented slide regions are large enough to make an accurate match and when 4- or 6-slide-per-page layouts are used, users do not put markings in the slide region of the handout as often as they do when they use handouts with 2- or 3-slide-per-page layouts, which provides a better matching accuracy.

Template (number of slides per page)	Number of pres. handouts	Number of slides on handouts (S_n)	Total number of PR images (I_{tn})	Number of matching PR images (I_{tn})	Number of correctly matched PR images (I_{tn})	Prec.	Recall
2	12	173	1122	197	187	1	0.95
3	12	299	1751	344	328	1	0.95
4	11	308	1254	317	304	1	0.96
6	9	270	1037	292	284	1	0.97
9	9	291	951	324	303	1	0.94
Total/Ave	53	1341	6115	1474	1406	1	0.95

Table 2. Results of matching scanned handout slides to the screen images captured by the Presentation Recorder.

5. CONCLUSIONS

In this paper, we presented a new system for creating personalized e-presentation documents by using presentation handouts as templates and populating them with a user's handwritten notes and e-presentation media. Experimental results demonstrated the accuracy of techniques for creating the e-presentation document. Delivery of such personalized multimedia documents is a valuable alternative to web-based access and retrieval of e-presentation media.

6. REFERENCES

[1] R A. Steinmetz, "Media and Distance: A Learning Experience", IEEE Multimedia, pp. 8-10, 2001.

[2] Jason A. Brotherton, Gregory D. Abowd , "Lessons learned from eClass: Assessing automated capture and access in the classroom", ACM Trans. on CHI, pp. 121 - 155, 2004.

[3] P. Chiu, J. Boreczky, A. Girgensohn, D. Kimber, "LiteMinutes: An Internet-Based System for Multimedia Meeting Minutes," Proc. WWW10, pp.140-149, 2001.

[4] P. Chiu, J. Foote, A. Girgensohn, and J. Boreczky, "Automatically linking multimedia meeting documents by image matching," ACM Hypertext, pp. 244-245, 2000.

[5] D. Bargeron, J. Grudin, A. Gupta, E. Sanocki, et al, "Asynchronous Collaboration Around Multimedia Applied to On-Demand Education," Proc. HICSS, 2001.

[6] Richard Anderson, Ruth Anderson, C. Hoyer, B. Simon, et al, "Lecture Presentation from the Tablet PC," Workshop on Advanced Collaborative Environments, 2003.

[7] K.N. Truong and G.D. Abowd, "StuPad: Integrating student notes with class lectures," Proc. of CHI, pp.208-209, 1999.

[8] B. Erol, J. J. Hull, J. Graham, and D-S. Lee "Prescient Paper: Multimedia Document Creation with Document Image Matching", IEEE ICPR Conference, 2004.

[9] Adobe Acrobat SDK, http://partners.adobe.com/public/developer/acrobat/devcenter.html

[10] University of Washington English Document Image Database, info available at http://documents.cfar.umd.edu/resources/database/ UWASH_Database