Automatic Capture and Presentation Creation from Multimedia Lectures

Paul E. Dickson, W. Richards Adrion, and Allen R. Hanson
University of Massachusetts Amherst, pauld@cs.umass.edu, adrion@cs.umass.edu, hanson@umass.edu

Abstract - For more than a decade, the RIPPLES group at the University of Massachusetts Amherst has developed and deployed content delivery systems to support on-campus and distance education. This has been an effective but expensive approach because of its dependence on content captured by human operators and substantial production effort. Presentations Automatically Organized from Lectures (PAOL) is a classroom-capture system that is transparent to and places no constraints (teaching style or pedagogy) on a lecturer. PAOL is unique—no other system can capture unconstrained computer-based materials, capture whiteboard writing and drawing, and produce a speaker video without requiring preinstalled software, electronic whiteboards, or special training. PAOL uses high-resolution cameras, a computer screen capture device, and a wireless microphone to automatically create multimedia Flash presentations, which include a digitally edited instructor video, enhanced images of all material presented by computer or written/drawn on a whiteboard, and an index to support navigation. The PAOL system has been used extensively during system development in several varied settings. We report on usability studies to determine the most effective presentation formats and interfaces and our efforts to capture a full course to further evaluate the impact on teaching and learning.

Index Terms – Distance education, Lecture capture, Content enhancement.

INTRODUCTION

As we move further into the 21st century, the world is faced with significant changes in higher education. In the developed world the number of students attending higher education institutions is growing and growth is greatest in on-line education. Rapid changes in pedagogy coupled with an explosion of knowledge force colleges to consider how to increase productivity and effectiveness. Faced with rising costs, many institutions have limited the size of faculty and depended more heavily on blended (on-line plus in-class) education. In developing countries the number of qualified students seeking higher education is growing exponentially without enough qualified faculty to teach them. Universities in Ghana are faced with 58:1 student teacher ratios [1]. Lecture halls are not large enough to seat all interested students, and many are forced to watch lecturers through the windows.

These changes suggest an increased use of educational technologies, which must be coupled with cost- and learning-effective strategies. The RIPPLES group at the University of Massachusetts Amherst has been developing the MANIC [2] content delivery framework for more than a decade and evaluating its use in a variety of educational and pedagogical settings. Instantiations of the MANIC content delivery system provide a platform [3] by which content can be delivered in a format that includes features proven effective in distance education. Several of these features were developed to support on-campus teaching and learning, including search, navigation, notation, and collaboration tools.

Content delivered using the MANIC technologies typically was provided by human operators and required production crews to create high-quality presentations. While RIPPLES tried a few approaches to automate content creation and capture, the demand for quicker and less expensive production and more flexibility in content delivery led to the development of PAOL, a content capture system. PAOL emphasizes transparency to the lecturer, a key tenant of the RIPPLES group. If one is to increase the adoption of technology in education, one must keep the overhead low. In PAOL, transparency is achieved by using a computer screen capture device, wireless microphones, and high-resolution cameras, all of which function unobtrusively. PAOL captures a small set of images of the computer screen and whiteboard (typically under 100/hour) that represent “significant events” (a slide transition, inking on a tablet, the state of an application, a figure, or writing on a whiteboard) together with a digitally zoomed tracking video of the speaker. These images and video can be compiled, synchronized, and indexed into a Flash MANIC presentation (Figure 1) or the images can be distributed in real-time [4] using a modified Ubiquitous Presenter (UP) [4]-[5] interface and server.

When PAOL content is distributed via MANIC, the system creates a thumbnail index to each significant event. A user can navigate the content by playing the presentation from the beginning or from any index point. The indexes provide easy navigation to the desired points of the classroom record. This format facilitates distance education applications but also is very useful for on-campus students for study and review. The UP-like delivery system can be used in the classroom or later for review and study. When
combined with the MANIC tools for searching, notation, sharing, and collaboration, the PAOL captured content becomes a powerful learning environment.

In the following sections, we describe the PAOL content capture system and the MANIC and UP-like distribution systems. We summarize our extensive experimentation in a variety of classroom settings, usability studies, and experience capturing a full semester course and evaluating the resulting courseware.

**PAOL**

As we have noted, PAOL uses relatively inexpensive, high-resolution cameras, a simple hardware screen capture device, and a wireless microphone to produce content presented in an unconstrained manner on a computer, a whiteboard of many dimensions and aspect ratios, or a combination of both. PAOL is unique among content capture systems in its ability to capture both computer-based and whiteboard content while maintaining transparency. PAOL can handle a variety of lighting conditions—poor lighting, high-contrast projected images and the board, reflections, etc.

**Related Work**

Lecture capture systems have been under development for years. These systems range from the basic seminar-style capture of AutoAuditorium [6] to eClass [7], a comprehensive capture system that indexes computer and whiteboard material by using electronic whiteboard and special computers to capture all material. In between are such systems as the University of California Berkeley's system [8] that captures a video of the lecturer and any slides presented and Authoring on the Fly [9], which records lectures but focuses on methods of indexing captured material.

Many other capture systems exist, and all limit the lecturer in some way, either by restricting presentation modality or requiring special software or equipment. PAOL requires neither and will be compared with the most similar systems.

**System Components**

PAOL is currently installed in a medium-size classroom, the setup and images of which can be seen in Figures 2 and 3. The system is also installed in a slightly smaller classroom with a single camera but with all other components the same. PAOL's computer capture device is an Epiphan Systems VGA2USB. The high-resolution cameras are a pair of Point Grey Research Flea2 color cameras that are run at 1024x410 pixels and 15 frames per second. Figure 3 shows the small size of the Flea2 camera that makes it unobtrusive when mounted in a classroom. Each camera captures a 12-foot width of the front of the room, giving a resolution of 8.5 pixels per inch, and the cameras have a 2-foot overlap of board coverage.

**Capture Algorithms**

PAOL breaks lecture capture into 3 separate subproblems: whiteboard capture, computer capture, and video creation. Whiteboard capture and computer capture are based on the same principles. In each case, index points are created when significant content is presented. Significant content is defined as material that remains unchanged on whiteboard or
computer for a set length of time. We set the time through extensive experimentation. When significant content is identified, the image (whiteboard or computer) and the time it occurred are saved. The image is displayed in the appropriate content window of the created MANIC presentation (Figure 1). Thumbnails are created from these images and the timing information is used to create a thumbnail-based index. As noted, the captured content can also be sent to students using a modified UP system [4]. The PAOL capture algorithms are described in greater detail and with greater comparison with similar systems in [10] and [11].

Whiteboard Capture

The whiteboard capture algorithm analyzes the stream of images captured by the camera(s), identifies significant events, enhances the associated images for greater legibility, and stores the images with timing information. To identify content to store, the algorithm first creates a whiteboard ground truth image by breaking the camera images into blocks and giving each block the average value of the brightest 25% of pixels in that block. The brightest part of each block corresponds to the whiteboard pixels in the block. The whiteboard is then made white and text, drawings and other markings are enhanced by applying equation (1) to each color channel individually. In equation (1), \( P_{\text{out}} \), \( P_{\text{in}} \), and \( P_{\text{ground}} \) are the pixel values for the enhanced image, input camera image, and whiteboard ground image, respectively, and 255 is full saturation of a color channel.

\[
P_{\text{out}} = \min \left( 255, \frac{P_{\text{in}}}{P_{\text{ground}}} \cdot 255 \right)
\]  

Consecutive enhanced images are compared to determine lecturer location. Taking a target enhanced image and substituting the pixels from the previous image for those pixels where the lecturer is located creates a clean whiteboard image. The whiteboard images are then compared to determine when content changes and hence when a new whiteboard image needs to be saved. The saved whiteboard images are then sharpened and the contrast is increased. A series of images from the cameras and the captured whiteboard content associated with them is displayed in Figure 4. These examples show the variety of lighting conditions under which PAOL functions and the quality of results despite these lighting changes. Note that in the first two examples the overhead projector material is also enhanced.

The quality of the text captured is shown in Figure 5, which contains an enlargement of the text from the second example in Figure 4.

![Figure 4 Sample Whiteboard Results](image-url)
FIGURE 5
ENLARGEMENT OF CAPTURED WHITEBOARD TEXT.

One of the strengths of PAOL is the capture of the progression of ideas. An example of a concept being incrementally presented is shown in Figure 6.

FIGURE 6
PROGRESSION OF WHITEBOARD LECTURE CONTENT.

The best systems currently available for capturing whiteboard material are Microsoft's whiteboard capture [12] and TeleTeachingTool (TTT) [13]. Microsoft's whiteboard capture provides cleaner whiteboard images than PAOL but cannot function on large whiteboards or handle faint dry-erase markers as well as PAOL. TTT requires an electronic whiteboard to be installed and requires the lecturer to start a special capture computer. PAOL is more robust than Microsoft's system and more transparent than TTT.

Lecturer Video Creation

A video of the lecturer is created by locating the lecturer in each camera view image and creating a difference image between successive images from the same camera. The difference is used to identify a lecturer location and the location is used to crop a video frame from the original image (Figure 7). The algorithm decides which camera view to use to crop the video frame by determining which camera view contains the lecturer. A smoothing function is used to ensure that the captured video frames do not jump around.

FIGURE 7
DIFFERENCE IMAGE AND CROPPED VIDEO FRAME.

Computer Capture

The computer capture algorithm searches for significant events to capture in the same manner as the whiteboard algorithm by looking for changes. The computer capture algorithm captures images of computer material using a VGA2USB converter. The captured image stream is processed to remove noise and identify stable (significant) screen images. As each significant event is identified, an image is saved as new content. With this method any material presented on screen can be captured and indexed (Figure 8).

FIGURE 8
COMPUTER CAPTURE RESULTS.

Again, the progression of the material presented is important and can contain important information. An example of written content progression captured is presented in Figure 9.

FIGURE 9
PROGRESSION OF COMPUTER CAPTURE CONTENT.
The systems most similar to the PAOL computer capture algorithm are Mediasite [14] and TTT [15]. Mediasite uses an external capture device similar to that used by PAOL to capture screen content and gets similar results. The PAOL computer capture algorithm does a better job of saving significant material whereas Mediasite can capture full videos, which PAOL cannot because it only saves stable material. On occasion PAOL does capture these videos in the lecturer video it creates. TTT gets results nearly identical to PAOL's but requires the lecturer to present the lecture by connecting to a special presentation server. PAOL provides more features overall than Mediasite and is again more transparent than TTT.

**EXPERIMENTAL RESULTS**

During development of the PAOL capture system, data were captured from 96 individual classes involving 9 lecturers. During the 2006-2007 academic year, data were captured using only the computer capture system. In addition, the computer capture system was demonstrated at a number of conferences (including [10]) and other venues. The computer capture system was also tested with the UP delivery system in a software engineering class in Fall 2007 [4]. Once the cameras were installed, a number of individual classes were captured using the computer, whiteboard, and lecture video capture system. The set of all sampled images from these classes serve as a significant base of experimental data that can be used to refine and evaluate the algorithms. The work reported in [10] and [11] describe our analysis of the quality of the captured and stored images in terms of duplicated and/or missed images, noise, etc.

We are currently capturing a full semester class and evaluating the integrated capture-delivery system via usability studies and assessing the impact on teaching and learning.

*Adapting the MANIC Content Delivery System*

The RIPPLES Group has carried out research on multimedia learning technologies and developed and deployed multimedia materials for on-campus and distance education since 1996. As indicated above, the RIPPLES Group uses the MANIC [2] framework to capture both live and authored presentations, coupling them with text, graphics, and search/index mechanisms and delivering the content through streaming servers and downloads and on CD/DVD.

The RIPPLES Group developed courses delivered to students enrolled at the university in on-campus and distance education courses as well as to students at a number of other universities. Overall, CD-MANIC was used successfully as the primary source of course content by over 500 students enrolled in over 30 graduate level courses in engineering and computer science [16].

MANIC initially was developed as a platform for testing streaming multimedia network protocols; its initial incarnation involved synchronizing HTML slides with streaming audio. In 2000, CD-MANIC was developed as a mechanism for delivering bandwidth-intensive high-quality video, synchronized and indexed with lecture slides, to users who faced bandwidth limitations [17]. CD-MANIC was designed to use the network to check for updates to slides, and this activity requires very little bandwidth. As high-bandwidth access became more widely available, CD-MANIC remained a good solution for delivering multimedia content to distance learners because of its portability and anytime, anyplace approach. MANIC has been extended to add support for navigation, notation, and collaboration so that it better supports a constructivist teaching style for distance and on-campus education, including adding extended search capability [18], a synchronous collaboration mechanism [19], and a Learner Logger [17] tool that captures all student interactions. Learner Logger data provide feedback to the instructor and instructional designers.

Recently, RIPPLES began to migrate multimedia course content back to the Web as a means of delivery. Various courses were encoded using Synchronized Multimedia Integration Language (SMIL) and a cross-platform Java version called jMANIC. To exploit the availability of PAOL captured content, a new version of MANIC has been implemented as a Flash application. This provides the cross-platform capability of SMIL and jMANIC but has proved to be a more stable and effective content delivery system for the PAOL high-resolution images. Figure 1 shows the Flash MANIC GUI. We are currently porting the extensive set of searching, navigation, collaboration, and notation tools into Flash MANIC

*Usability Studies and Educational Impact*

As noted, we are capturing a full semester class with the long-term goal of evaluating the integrated PAOL capture system/MANIC delivery system for its impact on teaching and learning. This evaluation is underway during the spring semester 2008. We plan to measure teaching and learning outcomes by looking at student performance, gathering student evaluations, and interviewing and surveying students and instructors.

We will immediately be able to carry out a series of usability studies to determine how well the PAOL system captures the essence and important events in the classroom and how easy it is to navigate to information students are seeking. We also are evaluating the Flash MANIC GUI and tools. The usability studies involve focus groups of students enrolled in the class and students who took the class previously.

**FUTURE WORK**

In the current form, PAOL and MANIC enable the capture and presentation of courses that include whiteboard and computer presentation of materials and do so transparently. For full adoption, the type of material captured by PAOL will need to be broadened. As a next step we plan to develop algorithms that enable PAOL to capture material presented on chalkboards. Chalkboard capture will make it possible
for PAOL to be used in developing countries where chalkboards are still more prevalent than whiteboards.

In many MANIC instances, the indexes are provided as a text table of contents (TOC). The availability of text indexes supports the MANIC search tools. When manually produced, it is relatively easy to add text indexes, and RIPPLES has experimented with OCR extraction of slide titles. For PAOL-produced content, we will have to use a combination of OCR, computer vision and speech-understanding techniques to extract meaningful and accurate text indexes. This work has been underway before and in parallel with the PAOL development. Since the MANIC notation tools provide a TOC index, we can also allow learners to provide their own indexes.

CONCLUSIONS

PAOL captures computer and whiteboard presentations robustly and transparently. It identifies significant events in each medium and enhances content for legibility. The created content can be compiled and delivered by a MANIC content delivery system or distributed to students in class in real time using a modified UP delivery system. It does all of this transparently and with a quality and comprehensiveness greater than any other lecture capture system.

ACKNOWLEDGMENTS

This work was partially funded by the National Science Foundation under grants CNS-0749756 and CNS-9979833 and by the University of Massachusetts under an ITC Strategic Initiative Grant.

REFERENCES


