AUTONOMOUS PRESENTATION CAPTURE IN CORPORATE AND EDUCATIONAL SETTINGS

David M. Hilbert, Thea Turner, Laurent Denoue, Kandha Sankarapandian
FX Palo Alto Laboratory, Inc.
3400 Hillview Ave., Bldg. 4
Palo Alto, CA, USA 94304
{hilbert,turner,denoue,kandha}@fxpal.com

ABSTRACT

While researchers have been exploring automatic presentation capture since the 1990's, real world adoption has been limited. Our research focuses on simplifying presentation capture and retrieval to reduce adoption barriers. ProjectorBox is our attempt to create a smart appliance that seamlessly captures, indexes, and archives presentation media, with streamlined user interfaces for searching, skimming, and sharing content. In this paper we describe the design of ProjectorBox and compare its use across corporate and educational settings. While our evaluation confirms the usability and utility of our approach across settings, it also highlights differences in usage and user needs, suggesting enhancements for both markets. We describe new features we have implemented to address corporate needs for enhanced privacy and security, and new user interfaces for content discovery.

KEYWORDS

Multimedia capture, indexing, retrieval, web 2.0

1. INTRODUCTION

Presentations are ubiquitous in education, business, and government. But presentation archives are rare due to the cost of purchasing, setting-up, and using current recording technology. Even the most usable systems require users to deal with additional software or devices, and to start and stop recordings. Once content has been captured, few systems provide highly streamlined ways for users to search, skim, and share archived content. As a result, useful information passes through projectors all the time and is lost. If we could create useful archives cheaply and easily—without any added burden on anyone—the benefits would be far reaching. ProjectorBox is our attempt to create an autonomous appliance that seamlessly captures, indexes, and archives presentation media, with streamlined user interfaces for searching, skimming, and sharing content.

2. RELATED WORK

There are three main approaches to automatic presentation capture: instrumented environments, screen capture software, and RGB-based appliances. Solutions that leverage instrumented environments, such as [1,3,5,6,12,13,14,15,17], can produce rich presentation records. However, they are notoriously expensive to set-up, operate, and maintain. Thus, such approaches are unlikely to become pervasive in the near future. Solutions that leverage software to record PC screen activity, such as [16,19,21], are simpler to set-up and operate. However, they require presenters to install software and manually start and stop recordings, and fail whenever a non-preconfigured PC, such as a guest presenter's laptop, is used. Thus, not all presentations are captured. RGB-based appliances which intercept the video signal sent from presentation devices, such as a presenter's laptop, to display devices, such as a projector, can capture content from any presentation device and software, with limited impact on presenters [2,18]. However, these solutions also require users to start and stop (or schedule) recordings, and do not produce easily searchable and skimmable archives.

Current approaches assume that presenters, facility operators, or audience members will adjust their practices to garner the benefits of automatic presentation capture. In our experience, even the most modest assumptions—e.g., that presenters will use specific software or start and stop recordings—are unrealistic. Thus, we sought to build presentation capture capabilities that "weave themselves into the fabric of everyday life" as Mark Weiser famously envisioned for ubiquitous computing systems [20]. ProjectorBox realizes this vision by pairing RGB-based capture with intelligent media analysis to automatically create easily searchable and skimmable archives without anyone having to start and stop (or schedule) recordings. We also depart from past research in comparing automatic presentation capture in both educational and corporate settings, uncovering differences in usage, user needs, and opportunities for future improvements.

3. PROJECTORBOX

ProjectorBox is an RGB-based appliance, like Anystream Apreso [2] and Sonic Foundry MediaSite [18], that can capture content from any presentation device running any presentation software. However, it goes beyond existing approaches in that users can set it up in a conference room or classroom and forget about it. It unobtrusively records the video signal sent from PCs to projectors and applies intelligent media analysis to automatically record high-resolution slide images, text and audio without requiring anyone to manage or schedule recordings. A web-based user interface makes it easy for users to search, skim, and share content. And a web service API enables additional services to be built on top of the captured content.

3.1 Requirements

In order to autonomously produce high-quality archives suitable for searching, skimming, and sharing, an RGB-based solution must implement slide classification, presentation segmentation, text extraction, and interfaces for non-linear playback.

The first challenge is to automatically separate presentation content from non-presentation content and free presenters from having to remember to start and stop recordings themselves. Researchers have noted the importance of not distracting instructors with new recording technologies, particularly at the beginning and end of classes when students ask questions [1]. And our own experience [5] has demonstrated that if people must remember to start and stop recordings, then most presentations will simply not be recorded. In terms of RGB capture, this meant we needed to robustly classify screen activity as either "associated with a presentation" or as desktop activity "not associated with a presentation". Thus, we developed and evaluated several slide classification algorithms to address this challenge [11].

Because we envisioned our solution running continuously in rooms used by multiple people for multiple presentations, we also needed to automatically group presentations to allow them to be browsed and retrieved as cohesive units. We describe our approaches to presentation segmentation in [11].

Finally, students want to be able to retrieve presentations based on content, and review specific bits of captured media non-linearly, as opposed to having to play through video sequentially [1]. We also experienced similar requirements in our own corporate conference room [5]. As a result, we apply optical character recognition (OCR) to extract text from slide images and create a full-text index. This allows users to retrieve individual slides within presentations based on content. And our slide skimming and playback interfaces, described below, allow users to easily skim and skip around presentation content non-sequentially.

3.2 Implementation

ProjectorBox is a PC-based system equipped with a high-resolution VGA capture card [7]. This card can capture VGA signals from any computer at any resolution up to 1600x1200. In addition, ProjectorBox can capture audio using any Windows-compatible audio device. We have installed our prototype in multiple small-form-factor PC cases, which are easy to deploy in classrooms and conference rooms and can be integrated with existing presentation podiums.

ProjectorBox consists of two main software components: a capture component and a server (Figure 1). The capture component transmits images and associated audio clips to the server using HTTP. Thus, the

capture and server components can run on the same PC, or a single server can integrate content sent from capture components distributed in multiple classrooms and conference rooms.

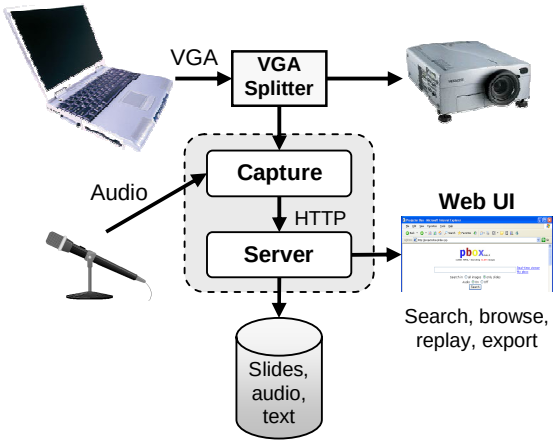

Figure 1. The ProjectorBox architecture.

When the server receives an image, it generates a thumbnail version for the web interface and calls the OCR component to extract its textual content along with the bounding boxes for each word in the image. The image, text and audio data is time stamped and saved in a relational database. The average size of one hour of recording is 30MB (250 KB per minute for the MP3, and 400 KB per slide image, with 40 slides per hour). This is ten times lower than state of the art MPEG4 video encoders for similar high-resolution encodings (e.g. 1024x768 pixels). The server also performs slide classification and presentation segmentation (as described in [11]) and provides the web-based user interface for easy retrieval, skimming, and playback.

3.3 User Interfaces

The web interface supports several methods for quickly retrieving and reviewing content. The main page (Figure 2 left) shows a list of dates and times in a calendar-like list, indicating when content has been captured. If the user knows the date and time of a desired presentation, this browse interface provides a single-click solution to presentation retrieval.

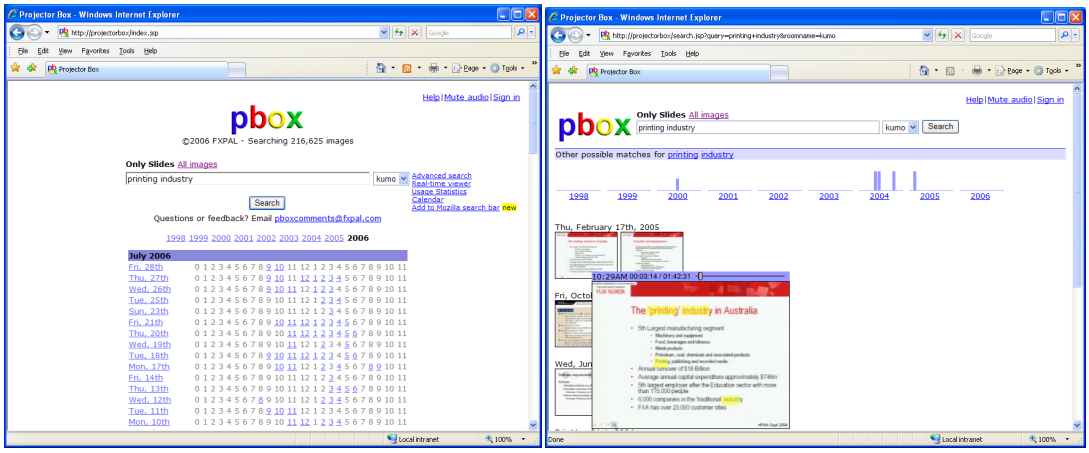

Figure 2. The main page (left) search results page (right)

The main page also provides a text field for full-text search of all captured presentations, allowing users to retrieve slides by content. The search results page (Figure 2 right) shows matching slides organized by

date. Mousing over a slide image shows an enlarged version with matching query terms automatically highlighted in yellow, and plays back the associated audio clip.

Browsing to content by date/time or clicking on a slide in the search results page activates the skimming interface (Figure 3 left). The skimming interface provides an overview of the slides captured on the selected day. Users can mouse over slides to show an enlarged version and play back the associated audio clip. This allows users to easily and non-sequentially skim slides and audio to locate segments of interest. Double-clicking on a slide brings up the slide player (Figure 3 right), for playing back content sequentially or skipping backward and forward in the presentation stream based on slide changes.

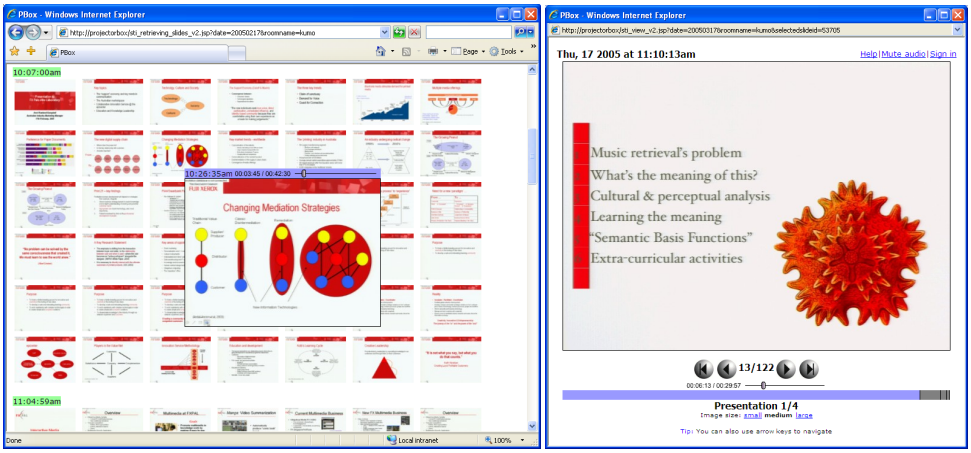

Figure 3. The skimming interface (left) and slide player (right)

While ProjectorBox automatically creates a web-based archive, we learned from previous experiments with a note-taking system [9] that hosting content on a server is not enough. Users want to be able to download presentations to their own laptops so they can easily access and share them anytime, anywhere, independent of network availability. Thus, ProjectorBox allows users to select slides of interest in the skimming interface, and to export slides (and audio) to a variety of printable and sharable formats including Microsoft PowerPoint, Adobe PDF, and Macromedia Flash. Furthermore, because the algorithms for slide classification and presentation segmentation are not perfect, the skimming interface also allows users to quickly re-classify slides and adjust presentations boundaries.

4. EVALUATION

We initially focused on capturing static slide images and room audio. We reasoned that dynamic content (such as slide animations, videos, and software demonstrations) could be captured later if users wanted it. We also decided to forgo capturing room video for the sake of simplicity and to see whether users would demand it. Would users find the system useful? Would they miss content we hadn't captured? To find out, we deployed and evaluated ProjectorBox in our own corporate conference room (22 users over 35 weeks) and in a university classroom at the Naval Postgraduate School (15 users over 8 weeks). Below we discuss the results based on user surveys, web logs, and informal discussions with participants.

Many frequent users: Over 55% of corporate users and 70% of students reported using ProjectorBox "at least once a week", "every few weeks", or "a few times" over the course of the study. The lower percentage of frequent corporate users is partly due to the fact that interns, contractors, and support staff (who typically do not attend presentations) were included in the survey. These users were twice as likely as researchers to be infrequent or non-users. In the remainder of our discussion, we only present results from users who used the system at least once.

Used for missed presentations, review, and awareness: Most corporate users (95%) and students (85%) reported using ProjectorBox to review presentations or classes they were unable to attend. The next most important corporate uses included reviewing presentations users had already seen (85%) and finding out

"what's going on in the lab" (75%). The next most important educational uses included using ProjectorBox to help study for exams (57%) and complete homework assignments (43%).

A safety net, less need to bother others: Most corporate users and students (85% in both groups) reported feeling more comfortable missing a presentation or lecture in the event they were unable to attend. One student pointed out that ProjectorBox offered:

> *Another option prior to seeing the Prof for help or having to rely on your classmates if you missed class.*

This suggests ProjectorBox not only helps people who miss presentations, but also others who would otherwise be spending time to bring them up to speed. We do not know whether absenteeism increased, but users viewed ProjectorBox as a safety net. ProjectorBox also captured information users thought might otherwise be lost, or inconvenient to acquire:

> *What I like best about ProjectorBox is easy access to information that would otherwise be lost.*

> *What I like best about ProjectorBox is anytime, I can get meeting slides. I don't need to ask a presenter to send me the PPT file and save it to my PC.*

Here again ProjectorBox provides a sense of security, and reduces the need to inconvenience others for content.

Reduced note taking, increased attention: Many corporate users (65%) and students (64%) reported taking less notes than usual, and felt they were able to pay more attention in meetings (55%) and in class (71%):

> *Knowing the audio is there to review with the slides makes it less critical to frantically copy what the professor is saying.*

> *I don't take notes. I stopped doing that because I never look at them again anyway. ProjectorBox not only gives you the [slide] image but also what they said. You can get right back to the moment of the presentation and remember what you thought about then.*

We believe the effect on attention was more pronounced for students because students probably spend more time than corporate users taking notes because they are learning new material that they need for homework and exams.

Autonomous recording, easy access, skimming: Several corporate users reported liking ProjectorBox's seamless operation and simple access interfaces, particularly compared to an earlier video-based system [5]:

> *I use it in place of [our automatic video archive system]. I mainly use it to review meetings I missed, but it is way more reliable than [our video system] in terms of being always on and no lag between the meeting and when the content is available.*

> *What I like best about ProjectorBox is access to large amounts of previously unbrowsable data (well, there was video, but what a pain).*

Because ProjectorBox does not require people to manually start and stop recordings and upload content to a server after recordings (like our video system), users could depend on ProjectorBox to capture the content they need and to archive it in real-time. Users also found our simple browse, search, and skim interfaces to be more efficient for their purposes than browsing through folder-style hierarchies to locate and playback video recordings. Finally, our log data confirmed that users in both cases were far more likely to skim or playback portions of lectures and presentations than to play them back in their entirety.

Browse versus search: All users reported browsing by date or time. They preferred to browse instead of search when they were looking for a specific date, time, or event. They preferred to search when looking for a specific topic or fact. About 14% of the students versus 55% of the corporate users used search. Students explained that they knew when a given lecture had occurred and so it was simpler to go directly to it rather than searching for it by topic. Because the course took place over a short period (8 weeks), they could easily find their target content by clicking on the day. Some students speculated that they would be more likely to search if they needed to go back to a lecture farther in the past, and we confirmed that searches typically targeted content further in the past: the median date for content retrieved from browses was about 7 days in the past, while the median date for content retrieved from searches was over 200 days in the past. These results suggest that the value of searching increases as the presentation archive grows. It also suggests we

could focus the calendar-based browse interface on the homepage to just the last several days, and use the reclaimed space for additional content discovery interfaces (as we describe below).

What's missing? We asked our users whether they missed other types of information we were not capturing. Everyone in the corporate setting felt it would be useful to capture software demos and video clips, and many felt room video would be useful at least sometimes (65%) or all of the time (20%). In the educational setting, many felt it would be useful (at least some of the time) to capture software demos (61%) and whiteboard activity (69%). However, few students (23%) felt that room video would be useful. In addition to slide images, text, and audio, 70% of our respondents in both settings felt it would be useful to be able to search the audio. However, users in both settings felt ProjectorBox was useful despite these missing bits of information.

Privacy and security: About a quarter of our corporate users reported having concerns about privacy or security, and 75% said they would like to be able to easily turn ProjectorBox on and off. In contrast, no one in our educational deployment indicated any concerns.

Users liked ProjectorBox: Finally, users in both settings felt ProjectorBox added perceptible value. All students who used ProjectorBox said they would like to use it again in future classes and 95% of corporate users said they would miss it if it were gone. This is encouraging given that ProjectorBox's operational overhead is nearly imperceptible.

5. DICUSSION AND NEW DIRECTIONS

Our study confirmed that both corporate and educational users value simple slide and audio capture, but indicated ways in which we might improve our support for each setting. Students would like to have automatic whiteboard capture, while our corporate users were most interested in enhanced privacy and security, and new interfaces to help find personally relevant presentations they missed and to get a sense of what was going on in the organization.

5.1 Security and privacy

To address the privacy and security concerns expressed by our users, we implemented a touch-screen-based recording console that clearly indicates ProjectorBox's recording status and provides one-click recording control (pause and resume) and one-click routing of captured content into access controlled folders.

The resulting user experience remains very close to our ideal of seamless presentation capture. In cases where access control is not required, the user does not have to do anything: ProjectorBox simply operates as before and seamlessly records presentations which are then made accessible to everyone in the organization. If, however, captured content needs to be protected, the presenter logs into the recording console (optionally using a smart card) and selects a target folder, and ProjectorBox routes the presentation to the selected folder and restricts access to only those who can access that folder. When the presentation is over, the presenter taps the logout button to reset ProjectorBox to its default state. For security reasons, ProjectorBox does this automatically after a specified period of inactivity.

5.2 New ways to discover and access content

Our users were satisfied with our interfaces for browsing, searching, and skimming, particularly when looking for specific presentations or topics. But they were also interested in additional interfaces to help them find personally relevant presentations they didn't know about, and to increase their awareness of what was going on in the organization.

One way to help users discover personally relevant content is to proactively recommend it to them based on their current activities. Thus, we integrated ProjectorBox as a new information source for our proactive contextual retrieval tool, PAL Bar [4]. PAL Bar now proactively recommends presentations related to the user's "context", such as a currently displayed web page or email message [10].

We also designed a new homepage (Figure 4) to help users discover and access content in new ways. The "Tagged Presentations" section (Figure 4, top-left, below the search box) shows thumbnail previews of presentations that have been tagged by the community and provides an easy way for users to discover and

access recent talks of interest. Users can click on presentation tags corresponding to "title", "presenter", "presenter organization", and other arbitrary tags to pivot around the archive showing other presentations with the same tag. An RSS feed allows users to subscribe to newly tagged presentations and a Podcast feed allows users to download talks to their iPods.

The "Untagged Presentations" section (Figure 4, bottom-left) shows thumbnails of recently captured talks that have not yet been tagged, and invites users to tag them. Once a presentation has been tagged, it is moved up into the "Tagged Presentations" section.

Figure 4: New homepage for discovering and accessing content

The "Browse by Date" section (Figure 4, top-right, below the search box) provides a calendar interface that highlights days in which ProjectorBox has captured content. Clicking on a day shows content for that day, and clicking on "Older" and "Newer" navigates backward and forward through the months showing "Tagged Presentations", "Untagged Presentations", and "Topics" for each month.

The "Topics" section (Figure 4, bottom-right) provides a simple visualization, similar to the "tag clouds" popularized by Flickr, that automatically summarizes the most popular topics during the selected month with the most frequently occurring topics appearing in larger fonts. Clicking on any topic retrieves matching slides from the archive.

6. CONCLUSION

Our evaluation results supported our initial vision and design decisions. A simple, easy to use appliance can provide noticeable utility with nearly unnoticeable operational overhead. ProjectorBox was both easy to use and useful across multiple deployment settings despite significant differences in the backgrounds, goals, and

practices of our users. Automatic slide and audio capture helps students review missed classes and study for homework and exams, and corporate users leverage the same capabilities to review missed presentations and stay informed about "what's going on" within the organization.

Our study confirmed that both corporate and educational users value autonomous slide and audio capture, but indicated ways in which we might improve our support for each setting. Students would like automatic whiteboard capture while corporate users were most interested in enhanced privacy and security, and new interfaces to help find personally relevant presentations and to get a sense of what was going on in the organization. Many users requested that dynamic content—e.g., when the presenter shows videos, animations, web pages, and software demonstrations—be captured as video clips in addition to the slide images we currently capture. This will require new algorithms for robustly detecting and capturing dynamic presentation content while ignoring dynamic desktop interactions not associated with any presentation. Finally, our automatic slide classification and segmentation algorithms also remain subjects of ongoing research.

REFERENCES

1. Abowd, G.D., Classroom 2000: An experiment with the instrumentation of a living educational environment. IBM Systems Journal 38, (1999), 508-530.
2. Anystream. Apreso. http://www.apreso.com/.
3. AutoAuditorium. http://www.autoauditorium.com/.
4. Billsus, D., Hilbert, D.M., and Maynes-Aminzade, D. Improving Proactive Information Systems. In Proc. IUI 2005, (2005).
5. Chiu, P., Kapuskar, A., Reitmeier, S., and Wilcox, L. Room with a Rear View: Meeting Capture in a Multimedia Conference Room. IEEE MultiMedia Magazine, 7, 4, (2000), 48-54.
6. Cruz, G. and Hill, R. Capturing and playing multimedia events with STREAMS. In Proc. MULTIMEDIA '94, (1994), 193-200.
7. DataPath. Vision RBG-PRO. http://www.datapath.co.uk/visRGBPRO.htm.
8. Denoue, L., Hilbert, D.M., Adcock, J., Billsus, D., and Cooper, M. ProjectorBox: Seamless presentation capture for classrooms. In Proc. E-Learn 2005, (2005).
9. Denoue L., Singh G., and Das A., Taking Notes on PDAs with Shared Text Input. In Proc. ED-MEDIA 2004, (2004).
10. Hilbert, D.M., Billsus, D., and Denoue, L. Seamless Capture and Discovery for Corporate Memory. In Proc. WWW 2006, (2006).
11. Hilbert, D.M., Cooper, M., Denoue, L., Adcock, J. and Billsus, D. Seamless presentation capture, indexing, and management. In Proc. SPIE Optics East, (2005).
12. Liu, Q., Rui, Y., Gupta, A., and Cadiz, J. J. Automating camera management for lecture room environments. In Proc. CHI '01, (2001), 442-449.
13. Liu, Q., Shi, X., Kimber, D., Zhao, F., Raab, F. An Online Video Composition System. In Proc. ICME 2005, (2005).
14. Mukhopadhyay S. and Smith, B., Passive capture and structuring of lectures. In Proc. MULTIMEDIA '99, (1999), 477-487.
15. Müller, R. and Ottmann, T. The "Authoring on the Fly" system for automated recording and replay of (tele) presentations. Multimedia Systems 8, 3 (2000), 158-176.
16. Quindi. Meeting Companion. http://www.quindi.com/.
17. Rui, Y., Gupta, A., and Grudin, J. Videography for Telepresentations. In Proc. CHI '03, (2003), 457-464.
18. Sonic Foundry. MediaSite. http://www.mediasite.com.
19. TechSmith. Camtasia Studio. http://www.techsmith.com/.
20. Weiser, M. The computer of the 21st century. Scientific American, 265, 3, (1991), 66-75.
21. Ziewer, P., Navigational Indices and Full-Text Search by Automated Analyses of Screen Recorded Data. In Proc. E-Learn 2004, (2004), 3055-3062.