

Whiteboard Content Extraction and Analysis for the Classroom Environment

Paul E. Dickson, W. Richards Adrion, and Allen R. Hanson
Department of Computer Science
Computer Science Building
University of Massachusetts
Amherst, MA 01003, USA
{pauld,adrion,hanson}@cs.umass.edu

Abstract

We describe whiteboard content capture system from Presentations Automatically Organized from Lectures (PAOL) that captures content within the setting of a classroom environment. The system acquires a sequence of images from high-resolution, fixed view cameras and extracts a series of content-rich key frames. The key frames are derived by analyzing the camera images of the front of the room, selecting images that show the progression of the presented material (writing, projected images), digitally removing the speaker's image, and enhancing the images for greater legibility. These images can be distributed in real time or stored and later synchronized with other content (video, computer capture). The system requires no specialized hardware other than fixed-field-of-view cameras, is transparent to the lecturer, and handles both whiteboard and projected presentations. The system extracts on average approximately 70 key frames from roughly 70000 images captured during a 75-minute presentation and runs at close to real time.

1 Introduction

In educational settings many forms of media are used to present material. It is often desirable to abstract or record the material for in-class interaction, post-class study and review, or distance education. A number of systems automatically capture and organize classroom content into indexed presentations [5, 11]. These content capture systems are generally weakest in capturing whiteboard material, often relying on electronic whiteboards or video capture of whiteboards, both of which have limitations. Our whiteboard capture system addresses these limitations and has two goals: first, to capture all material presented on the whiteboard as a series of key frames that represent the progressive development of key ideas; second, to capture all

information transparently so as not to burden the lecturer. Captured content is distributed via several multimedia content delivery systems.

The whiteboard capture system is part of a larger lecture capture system called Presentations Automatically Organized from Lectures (PAOL) [2, 3]. PAOL automatically captures and indexes lectures from any combination of computer and whiteboard content. Whiteboard content capture compliments the PAOL computer content capture [1] and, with captured video and indexes, provides a complete record of classroom activities.

2 Related Work

Whiteboard content can be captured dynamically or as a series of static images. The latter enables us to index and synchronize content but requires determining when something significant has occurred on the whiteboard (text, formula, drawing, etc.). We compare our system with others that specifically capture and store content but not those that merely record a video of the whiteboard.

Electronic whiteboard systems all have limited size and aspect ratios, can be costly, and require special input devices that can be susceptible to interference. They cannot handle projected material or material taped to the writing surface. The SMARTBoard [13] is typical of (up to wall-size) touch screens that can be configured to provide very high-quality content capture. Touch screens are expensive and the price increases with size. Some interactive whiteboards (e.g., Numonics Corporation [8]) use a computer and an LCD projector and track the movements of pen-like devices to simulate pen movements in the projected image. Other interactive whiteboards (e.g., Mimio [7]) use ultrasonic sensors to capture written material. Such systems are less expensive but fail in buildings with ultrasonic light switches and limit the size of the capture area.

We use an image-based capture system because such systems are often less expensive than electronic whiteboards

and equivalent in price to ultrasonic capture bars, requiring only high-resolution cameras and reasonable processing speed, and are easily added to any room to capture any kind of projected or written whiteboard material. WhiteBoardPhoto [10] and ZombieBoard [12] are among the simplest image-based whiteboard capture systems; each acquires whiteboard images and enhances them to improve legibility. Both require the instructor to prompt them to save material. The Wienecke et al. [15] whiteboard capture system is a first step to extract, via OCR, text written on whiteboards. Their method of content extraction, using feature vectors based on pixel intensity, average intensity change, and edges, worked poorly under the variable lighting conditions and dry-erase marker quality commonly found. The system designed by He and Zhang at Microsoft as part of their meeting recorder [4] is the most similar to ours. They build a model of whiteboard color and use it to remove anyone standing in front of the whiteboard and then identify and enhance text. This system does not translate well to whiteboards with relatively unconstrained dimensions. The region based board capture system proposed in [14] was considered but rejected as it relies on a color histogram based board segmentation algorithm which fails when confronted with a whiteboard and white haired lecturers. The region based method of identifying text also relies on more legible text than was typically found in our sample lectures.

3 Whiteboard Capture

A whiteboard capture system should record images of all significant content, i.e., track board events and save a new image when a significant event (a series of changes that become stable for some period) occurs. In PAOL, whiteboard capture is a 2-step process. The main process involves acquiring the images, removing the lecturer from the field of view, and storing those images whenever a stable change is detected. The post-processing step filters the images saved in the first step to remove duplicate images and then enhances the remaining images to improve legibility. This process is performed on the images from each camera stream individually. The main process functions in real time, processing more than 1 frame per second (fps) for a single camera and just under 1 fps for a pair of cameras. In both cases the post-process runs at 5 fps and completes all processing within 2 minutes of the lecture's end. In our instrumented classroom, all whiteboard images are captured by a pair of Point Grey Research Inc. Flea2 cameras run at 1024x410 pixels with a fixed field of view and an effective resolution of 8.5 pixels per inch and automatically synchronize when run on the same firewire bus.

3.1 Main Process

Content capture is not straightforward because of noise and various forms of interference; many problems arise from lighting conditions, reflections, poor markers, bad cable connections, occlusions by the lecturer, etc. The goal of the main process is to decide when to save an image, remove the lecturer, and improve the legibility of the images. First, the input image is averaged to create a ground-truth image for whiteboard color. Next, this average image is used with the original image to create a refined image with enhanced written material. Consecutive refined images are compared to create a difference image. This difference image is then used with a previously created whiteboard image to create a new whiteboard image. Consecutive whiteboard images are compared to determine when the written material has changed and which whiteboard images to save.

3.1.1 Average Image

The average image is created by dividing the input image (Figure 1) into 16x16 pixel blocks. Within each block, each pixel is converted to grayscale and the brightest 25% of the pixels (those that are most likely to correspond to whiteboard pixels) are determined and used to create the average color value for the block. The average image is smoothed from top to bottom because the greatest color change in the whiteboard occurs from top to bottom as a result of overhead lighting as shown in Figures 1 and 2, where the bottom of the board is much brighter and a different color than the top.

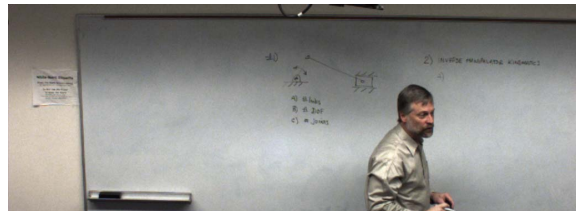


Figure 1. Input image.



Figure 2. Average image.

3.1.2 Refined Image

The refined image is created by remapping the colors of the input image (Figure 1) by applying Equation 1 to each color channel of every pixel to enhance the legibility of the written material.

$$P_{out} = \min\left(255, \frac{P_{in}}{P_{ave}} * 255\right) \quad (1)$$

In Equation 1, P_{out} , P_{in} , and P_{ave} are the pixel values for the output, input, and average images, respectively. All processing is performed in a red, green, blue (RGB) color space with 8 bits per color channel. This process further darkens pixels that are darker than the whiteboard color represented in the average image and sets whiteboard pixels to white. This improves the whiteboard material as shown in Figure 3.

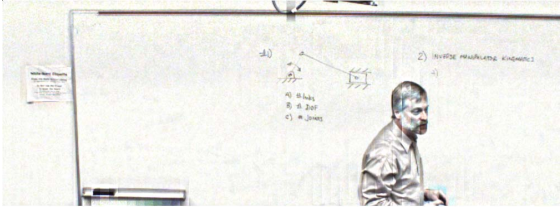


Figure 3. Refined image.

3.1.3 Difference Image

Once the whiteboard image has been refined, the lecturer is removed from the image. First, the lecturer is located. The lecturer is assumed to be the only moving object within camera view and a simple difference image between consecutive refined images shows where movement has occurred. The main process is performed at slightly more than 1 fps for a single camera and slightly less for two cameras.

For the difference image, only corresponding pixels where the sum of the differences between the color values of all 3 color channels exceed an empirically determined threshold of 80 are considered different. In Figure 4 the speckled pixels (invisible in print) at the top of the image and those roughly person shaped within the solid region are the identified difference pixels. If more than 5 of the 8 pixels surrounding a pixel are also difference pixels, the pixel is considered to be a true difference (with respect to the lecturer); all others are considered noise. This process prevents the specular noise at the top of Figure 4 from being confused with lecturer motion.

If the number of true differences is greater than 4500 pixels (roughly 1% of the image), the lecturer is considered found. When the lecturer is found, the region around the

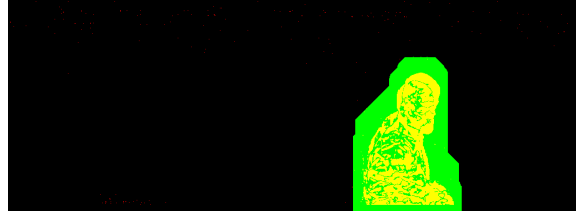


Figure 4. Difference image.

true differences is expanded. Instead of a standard region-growing algorithm, we use a specialized expansion that contains any pixels below true differences and expands the regions upward as it approaches the center of the lecturer. This ensures that all of the lecturer is enclosed within the lecturer region (Figure 4).

3.1.4 Whiteboard Image

The current refined and difference images and the previous whiteboard image are used to create a lecturer-free whiteboard image. The current refined image contains the up-to-date whiteboard pixels not occluded by the lecturer so the new whiteboard image starts as the refined image. All pixels occluded by the lecturer must then be replaced by the corresponding pixel from the previous whiteboard image. In this manner the pixels occluded by the lecturer are filled in with their values from the last time they were visible. Assuming that the lecturer is located correctly, this means that the lecturer will never be visible within the newly created whiteboard image (Figure 5).

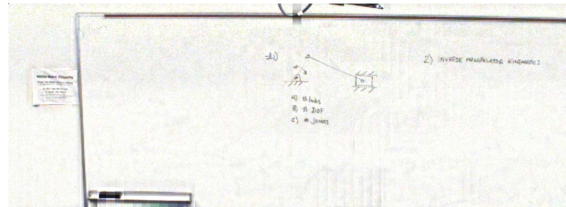


Figure 5. Whiteboard image.

Two issues remain. First, if the lecturer cannot be located then the previous whiteboard image is used as the current whiteboard image. If the lecturer cannot be located we do not know whether this is because the lecturer has moved out of the camera view or is just standing very still and therefore cannot create a new whiteboard image because the lecturer could be occluding a portion of the board. Second, when the initial whiteboard image is stored the refined image is used because no previous whiteboard image exists from which to create the initial whiteboard image. The lecturer may therefore appear in the first whiteboard image and portions of the lecturer may persist in following whiteboard images

until the lecturer no longer be occludes that portion of the whiteboard. Because of this startup problem, the lecturer appears in about 1.4% of all captured whiteboard images.

3.1.5 Whiteboard Image Difference

The system decides which whiteboard images to save by comparing consecutive whiteboard images and looking for differences. If the sum of the differences between all 3 color channels for corresponding pixels is greater than an empirically determined threshold of 200, the pixels are considered different. This threshold ensures that differences counted are only those created by true changes in content. If more than 10 differences are found, the images are considered different and a new whiteboard image is saved. Recall that the post-processing step will cull duplicates from this set.

3.2 Post-process

The post-process serves two functions: it eliminates duplicate captured whiteboard images that may have made it though the first pass and enhances the legibility of the whiteboard material. To eliminate duplicates, consecutive saved whiteboard images are compared and the number of differences of greater than 100 between the 3 color channels per pixel are added. If the sum of the differences exceeds 100, the images are considered to be different and both are saved, otherwise the newer one is eliminated. This process reduces the number of saved whiteboard images by 8%; 8.5% of the 8% of images removed are removed incorrectly, but this is less than 1% of all captured whiteboard material and thus is considered an acceptable loss.

The images that are kept must now be further enhanced for legibility. The first step is to increase the contrast by applying

$$p_{new} = p_{old} - (255 - p_{old}) \quad (2)$$

to each color channel of each pixel where the pixel is likely not whiteboard because pixel average brightness is below 230 or any color channel below 220. In (2) p_{new} and p_{old} are the higher contrast and original whiteboard pixel colors, respectively. Application of (2) heightens the contrast in relation to how different a pixel value is from white, so the darker a pixel before, the greater the contrast after.

Once the contrast has been increased, the images are Laplacian sharpened [6] to improve quality. The data in the contrast-increased whiteboard images are sharpened using Equation 4 for each color channel, where k is set to 0.55 and is the coefficient of sharpening; V is 2 and is the number of pixel shifts used for sharpening; and $P_{sharp}(x, y)$ and $P_{in}(x, y)$ represent the sharpened pixel and the input whiteboard pixel at point x, y , respectively. k was empirically determined to create the most readable results on text.

These results can be seen in Figure 6.

$$C(x, y) = P_{in}(x + V, y) + P_{in}(x - V, y) + P_{in}(x, y + V) + P_{in}(x, y - V) \quad (3)$$

$$P_{sharp}(x, y) = \frac{P_{in}(x, y) - (\frac{k}{4}) * C(x, y)}{1 - k} \quad (4)$$

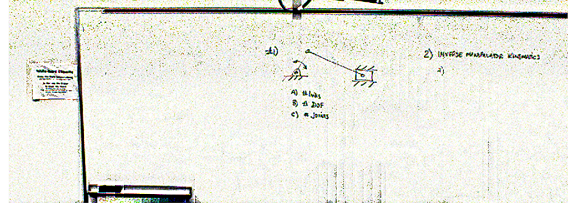


Figure 6. Contrast-increased and sharpened image.

Post-processing increases legibility but also emphasizes noise. Increased noise is a standard side effect of sharpening algorithms. We could achieve pristine whiteboard images with pure white whiteboards and only color where written material appears, but doing this would eliminate lines drawn with faint dry-erase markers, which are generally indistinguishable from noise. Hence, a design decision was made to emphasize anything that might be written material to ensure that no content would be eliminated even if it increased noise.

3.3 Sample Data and Results

The whiteboard capture system was developed and tested on 31 lectures, 30-75 minutes long, given by 12 different lecturers. The lectures included combinations of whiteboard, computer-projected, and overhead-projected material. Various lighting conditions and dry-erase markers (from new to those that should have been replaced) were used. A sample of capture in different environments and the corresponding results is shown in Figure 7. These results show the robustness of the system to adverse conditions and the quality that can be achieved under less than ideal conditions. These results also show that the whiteboard capture system can handle material presented with an overhead projector. Figure 8 shows an enlargement of the content captured. Figure 9 shows the ability of the system to capture the progression of ideas being presented.

In 75-minute lectures with one camera, 70,000 images on average were recorded; of these 72 whiteboard images on average were stored. Of the images saved, roughly 45% were “unnecessary” saves. Unnecessary saves are caused by changes to the images captured that do not correspond to

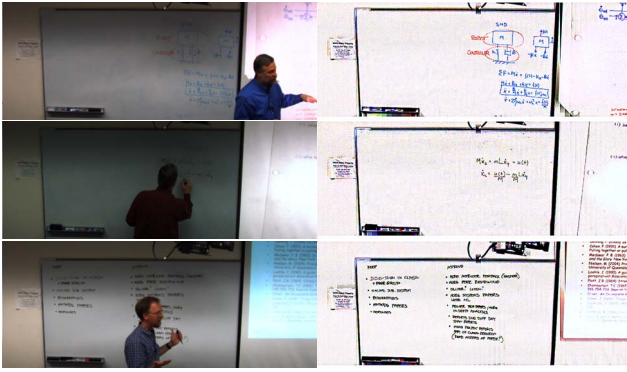


Figure 7. Sample whiteboard capture results.

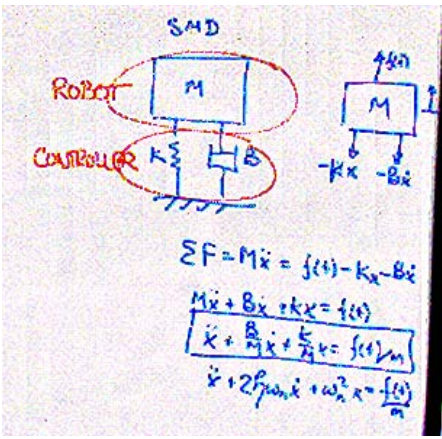


Figure 8. Whiteboard captured content enlarged.

changes in the material presented. An analysis of the causes of unnecessary saves is in Table 1, where *Board* refers to unnecessary saves caused by changes in the area of the input image containing only whiteboard and *Screen* refers to any unnecessary saves caused by changes in the area of the input images containing the projection screen.

The largest category of unnecessary saves is *Other*, the miscellaneous category. Most *Other* examples result from parts of the lecturer appearing or disappearing from the saved whiteboard images. The parts saved are often too

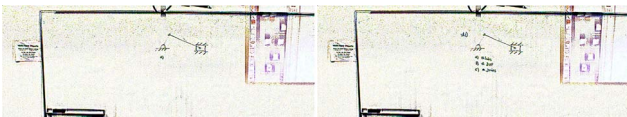


Figure 9. Progression of whiteboard material.

Area	Save Type	% U-Saves
Board	Other	48
Board	Shadow	16
Board	Noise	3
Board	Pen/Eraser	4
Screen	Blooming	9
Screen	Overhead	10
Screen	Computer	10

Table 1. Classification of unnecessary saves (U-Saves).

small to be readily identified as the lecturer. Specks that can be identified appear in roughly 18% of the captured whiteboard images.

Shadow represents saves caused by the lecturer shadow being the only change between images and *Noise* is random camera noise generally caused by low lighting. *Pen/Eraser* unnecessary saves are caused by changes to the location of material (e.g., pens and erasers) on the pen tray attached to the whiteboard. *Blooming* occurs when, for example, an overhead projector overloads the camera with light and is different from *Overhead* saves, which are typically caused by shifting of the transparency on the overhead projector. *Computer* represents all unnecessary saves related to LCD projection.

Unnecessary saves are not ideal but do not greatly affect system performance. They do affect the number of index points, the significance of which is being studied in usability studies. Even with 45% of the stored images deemed redundant, the stored images still represent only 0.1% of the total captured by the cameras—a 1000:1 reduction in storage requirements. For greater detail of how unnecessary saves affect PAOL's performance see [3].

4 Conclusions

PAOL can capture content and produce enhanced images of whiteboards and projected material a series of time-stamped key frames that represent the progression of ideas presented. The time stamps are used to create thumbnail-based index points in the video for use in content delivery systems. The captured images are enhanced to make all written material legible.

This system captures whiteboard material more comprehensively than any other image-based whiteboard system. For a pure white whiteboard and highly visible text, the enhanced images created by He and Zhang [4] are cleaner than those created by our system, but their underlying noise removal process is a simple threshold and marks made by faint dry-erase markers are eliminated. Our system captures

faint writing, but does so at the expense of noisier captured images. He and Zhang's system determines the whiteboard color for each pixel based on an assumption that the brightness of the whiteboard changes linearly across its surface. While adequate for small (meeting-size) whiteboards with a single overhead light source, it would have to be extended to handle the large whiteboards used in most seminar and classrooms. PAOL deals with boards that have bright spots from multiple lights and reflective surfaces in the room that create irregular lighting patterns. When the results of our whiteboard capture are compared with those of other lecture capture systems, our system captures approximately twice as many whiteboard images as those from electronic whiteboard, e.g., TeleTeachingTool [17] because of unnecessary saves. It does far better than camera-based capture systems that sample whiteboard material at a set interval (e.g., Panopto [9]). PAOL also can capture content projected from overhead and LCD projectors.

In PAOL whiteboard content is combined and synchronized with computer capture content and a lecture video and delivered in an indexed Flash-based content delivery system. Processing occurs at close to real time so that the key frames captured by PAOL can be delivered to students in real time using a Ubiquitous Presenter [16]-style presentation system. Because the number of passes of the selection algorithms is constrained by the real time demand, the UP-server presentations have slightly more duplicated images.

5 Future Work

Currently, the whiteboard capture system captures all whiteboard content but with a large percentage of unnecessary saves. We plan to reduce the percentage of unnecessary saves, improve the enhancement steps to reduce noise in the final images, and investigate extending this algorithm to handle chalkboards.

6 Acknowledgments

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

References

- [1] P. Dickson, W. R. Adrion, and A. Hanson. Automatic capture of significant points in a computer based presentation. In *ISM '06: Proceedings of the Eighth IEEE International Symposium on Multimedia*, pages 921–926, Washington, DC, USA, 2006. IEEE Computer Society.
- [2] P. E. Dickson, W. R. Adrion, and A. R. Hanson. Automatic capture and presentation creation from multimedia lectures. In *Frontiers in Education, 2008. FIE 2008. 38th Annual*, Oct 2008.
- [3] P. E. Dickson, W. R. Adrion, and A. R. Hanson. Automatic creation of indexed presentations from classroom lectures. In *Proceedings of the 13th Annual Conference on Innovation and Technology in Computer Science Education 2008*, volume 13, pages 12–16, 2008.
- [4] L.-W. He and Z. Zhang. Real-time whiteboard capture and processing using a video camera for remote collaboration. *IEEE Transactions on Multimedia*, 9(1):198–206, January 2007.
- [5] W. Hurst, G. Maass, R. Muller, and T. Ottmann. The "authoring on the fly" system for automatic presentation recording. In *CHI '01: CHI '01 extended abstracts on Human factors in computing systems*, pages 5–6, New York, NY, USA, 2001. ACM.
- [6] Imatest. Imatest - sharpening. <http://www.imatest.com/docs/sharpening.html>, Apr 2008.
- [7] S. LP. Interactive whiteboard, virtual whiteboard, whiteboards, sanford brands - mimio -. <http://www.mimio.com/>, Nov 2007.
- [8] Numonics. Numonics corporation - interactive whiteboards homepage, the smart boards for all your audio visual needs in the teaching, training, presentation and conferencing fields. <http://www.numonics.com/ipdindex.html>, Nov 2007.
- [9] Panopto. welcome to panopto. <http://www.panopto.com/>, Mar 2008.
- [10] PolyVision. Polyvision products. <http://www.polyvision.com/products/>, Nov 2007.
- [11] L. A. Rowe. Streaming media middleware is more than streaming media. In *M3W: Proceedings of the 2001 international workshop on Multimedia middleware*, pages 5–10. ACM Press, 2001.
- [12] E. Saund. Bringing the marks on a whiteboard to electronic life. In *CoBuild '99: Proceedings of the Second International Workshop on Cooperative Buildings, Integrating Information, Organization, and Architecture*, pages 69–78, London, UK, 1999. Springer-Verlag.
- [13] S. Technologies. Smart - smart board interactive whiteboards. <http://www2.smarttech.com/st/en-US/Products/SMART+Boards/default.htm>, Nov 2007.
- [14] M. N. Wallick, R. M. Heck, and M. L. Gleicher. Chalkboard and marker regions. In *Proceedings of the Mirage—Computer Vision/Computer Graphics Collaboration Techniques and Applications Conference*, pages 223–228, 2005.
- [15] M. Wienecke, G. A. Fink, and G. Sagerer. Towards automatic video-based whiteboard reading. In *ICDAR '03: Proceedings of the Seventh International Conference on Document Analysis and Recognition*, page 87, Washington, DC, USA, 2003. IEEE Computer Society.
- [16] M. Wilkerson, W. G. Griswold, and B. Simon. Ubiquitous presenter: increasing student access and control in a digital lecturing environment. In *SIGCSE '05: Proceedings of the 36th SIGCSE technical symposium on Computer science education*, pages 116–120. ACM Press, 2005.
- [17] P. Ziewer and H. Seidl. Transparent teleteaching. In A. Williamson, C. Gunn, A. Young, and T. Clear, editors, *ASCILITE*, pages 749–758. UNITEC Institute of Technology, Auckland, New Zealand, 2002.