WHY TAKE NOTES? USE THE WHITEBOARD CAPTURE SYSTEM

Li-wei He, Zicheng Liu, Zhengyou Zhang

Microsoft Research, One Microsoft Way, Redmond, WA, USA
Email: {lhe, zliu, zhang}@microsoft.com

ABSTRACT

This paper describes our recently developed system which captures both whiteboard content and audio signals of a meeting using a digital still camera and a microphone. Our system can be retrofit to any existing whiteboard. It computes the time stamps of pen strokes on the whiteboard by analyzing the sequence of captured snapshots. It also automatically produces a set of key frames representing all the written content on the whiteboard before each erasure. Therefore the whiteboard content serves as a visual index to efficiently browse the audio meeting. It is a complete system which not only captures the whiteboard content, but also helps the users to view and manage the captured meeting content efficiently and securely.
Keywords: Whiteboard capture, meeting archiving, video analysis, audio/video indexing, image classification

1. INTRODUCTION

The Whiteboard Capture System is a part of our Distributed Meetings project, aiming at dramatically improving the meeting experience of knowledge workers using ubiquitous computing technologies.

The work presented in this paper focuses on the particular meeting scenarios that use whiteboard heavily such as: brainstorming sessions, lectures, project planning meetings, patent disclosures, etc. In those sessions, a whiteboard is indispensable. It provides a large shared space for the participants to focus their attention and express their ideas spontaneously. It is not only effective but also economical and easy to use -- all you need is a flat board and several dry-ink pens.

While whiteboard sessions are frequent for knowledge workers, they are not perfect. The content on the board is hard to archive or share with others who are not present in the session. People are often busy copying the whiteboard content to their notepads when they should spend time sharing and absorbing ideas. Our system is an attempt to alleviate meeting participants the mundane tasks of note taking by capturing whiteboard content automatically along with the audio. We use a high-resolution digital still camera to capture one image of the whiteboard about every 5 seconds. Our system distills a small set of key frame images from the captured image sequence. Time stamps of the pen strokes contained in the key frames are also computed. The users can view the key frame images, print them as notes, or cut and paste them into documents. If the users want to find more about the discussion on a particular topic, our browsing software allows them to click some pen stroke associated with that topic and bring up the audio at the moment when the stroke was written.

Many technologies exist to capture the whiteboard content automatically. They generally fall into two categories. The devices in the first category capture images of the whiteboard directly [5,7]). Ours falls into this category. Devices in the second category track the location of the pen at high frequency and infer the content of the whiteboard from the history of the pen coordinates. This can be done either by using ultrasonic techniques [3] or by making the whiteboard essentially a giant touch-sensitive monitor [6]. Due to space limitation, the reader is referred to our technical report [2] for a detailed comparison between ours and the aforementioned ones.

The paper is organized as follows: Section 2 explains the design choices that we made and presents the architecture of our Whiteboard Capture System. Section 3 explains the technical challenges we encountered while building the system. Section 4 reports the details of the deployment process of our system.

Note that the drawings in this paper were done on a whiteboard and were captured by our system.

2. SYSTEM DESIGN

The design goals of the Whiteboard Capture System are:

1. Work with any existing whiteboard.
2. Capture the whiteboard content automatically and reliably.
3. Use the whiteboard content as a visual index to efficiently browse the recorded meeting.

2.1 Capture Device

One important design goal of our system is to work with any existing whiteboard, *without requiring* using special pens and erasures. As long as the users turn on the system before erasing, the content will be preserved. Our Whiteboard Capture System uses a 4-mega pixel digital still camera to capture the whiteboard content. The camera provides images that are 2272 pixels by 1704 pixels -- equivalent to 31.6 dpi for a 6' by 4' board. Using a mass market consumer product as opposed to proprietary hardware can potentially lower the cost.

2.2 Capture Interface

Another important feature differentiating our system from others is that people are not required to move out of the camera's field of view during capture as long as they do not block the same portion of the whiteboard during the whole meeting. This feature is very important because the users do not change any behavior when using our system. It does, though, add significant complexity to our analysis algorithm because we have to reconstruct the whiteboard content from different snapshots. A description of our algorithm is detailed in Section 3.

2.3 Browsing Interface

Since most people probably do not want to listen to the recorded meeting from start to end, we provide two browsing features to make non-linear accessing of the recorded information very efficient (see Fig. 1 & caption).

1. **Key Frames:** Key frame images contain all the important content on the whiteboard and serve as a summary to the recording. They can be cut and pasted to other documents or printed as notes.
2. **Visual Indexing:** We provide two levels of non-linear access to the recorded audio. The first is to use the key frame thumbnails. The user can click a thumbnail to jump to the starting point of the corresponding key frame. The second is to use the pen strokes in each key frame. Together with the standard time line, these two levels of visual indexing allow the user to browse a meeting in a very efficient way.

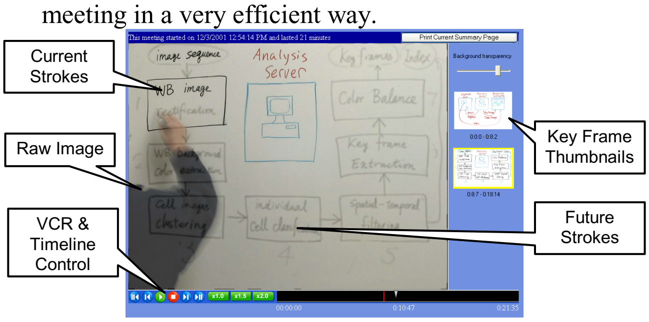


Figure 1: Browsing interface. *Each key frame image represents the whiteboard content of a key moment in the recording. The main window shows a composition of the raw image from the camera and the current key frame image. The pen-strokes that the participants are going to write in the future (Future Strokes) are shown in ghost-like style.*

3. TECHNICAL CHALLENGES

The input to the Whiteboard Capture System is a set of still digital images (see *Figure 2*). We need to analyze the image sequence to find out when and where the users wrote on the board and distill a set of key frame images that summarize the whiteboard content throughout a session. Compared to other systems, our system has a set of unique technical challenges: 1) The whiteboard background color cannot be pre-calibrated (e.g. take a picture of a blank whiteboard) because each room has several light settings that may vary from session to session; 2) Frequently, people move between the digital camera and the whiteboard, and these foreground objects obscure some portion of the whiteboard and cast shadow on it. Within a sequence, there may be no frame that is totally un-obscured. We need to deal with these problems in order to compute time stamps and extract key frames.

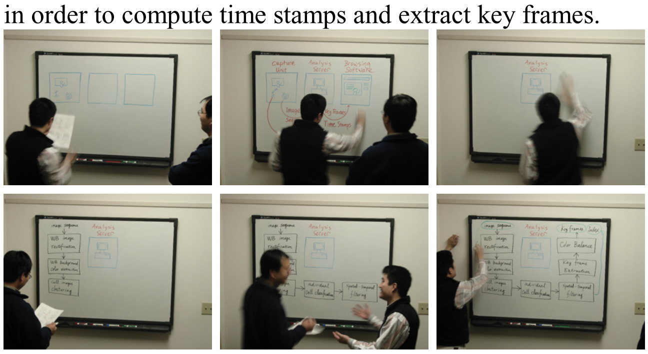


Figure 2: Selected frames from an input image sequence. *The session lasts about 16 minutes and contains 195 frames.*

Rather than analyzing images on per-pixel basis, we divide the whiteboard region into rectangular *cells* to lower computational cost. The cell size is roughly the same as what we expect the size of a single character on the board (about 1.5 by 1.5 inches). Since the cell grid divides each frame in the input sequence into *cell images*, we can think of input as a 3D matrix of cell images.

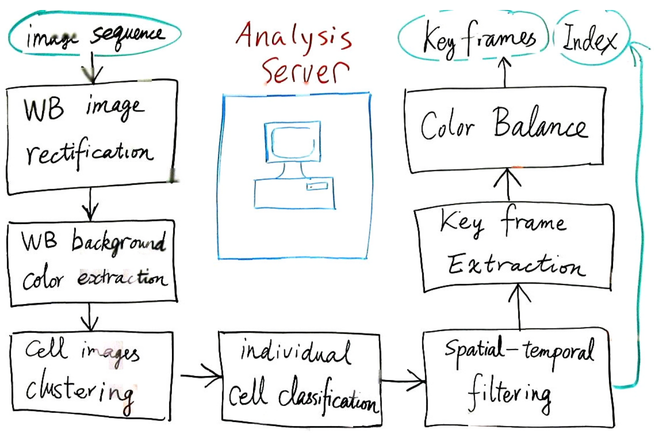


Figure 3: The image sequence analysis process.

Figure 3 shows an outline of the algorithm. We will use the sequence shown in *Figure 2* as a running example as to illustrate our algorithm. Due to space limitation, we can only provide a general brief description, and the reader is referred to our technical report [2] for details.

3.1. Rectify the Whiteboard Images

Before feeding the image sequence to the stroke extraction process, we crop the non-whiteboard region and rectify the images. Because the lens of the G2 camera has fairly low radial distortion, we only need to specify the four corners of the whiteboard. This is currently done manually by clicking a captured image during the one-time calibration step, although this could be done automatically. With the four corners, a simple perspective warp is performed for each image in the sequence using bi-cubic interpolation.

3.2. Computing the Whiteboard Color

For cell classification, we need to know for each cell what the whiteboard color is (that is, the color of the whiteboard itself without anything written on it). The whiteboard color is also used for white-balancing in producing the key frames, so it needs to be estimated accurately to ensure the high quality of the key frame images.

Figure 4: Whiteboard color estimation. *On the left is the result of our technique and on the right is the actual blank whiteboard image.*

This is not a simple task for several reasons: (i) the whiteboard color is not uniform, and is usually much brighter near the lights; (ii) there are maybe already strokes written on the whiteboard before meeting, so a cell may contain strokes throughout the sequence; (iii) foreground objects may be brighter than the whiteboard (e.g., when a person wears a white T-shirt and/or holds a piece of white paper). We have developed a robust and accurate technique combining histogram, *least-median-squares* [4], and local plane fitting. The left image of *Figure 4* shows the result for the input sequence in *Figure 2* where a person was holding a white paper in some of the frames. This should be compared with the image of the actual blank whiteboard shown on the right.

3.3. Clustering Cell Images over Time

During the meeting, the content of each cell usually changes over time. For each cell, we would like to cluster all the cell images in the sequence into groups, where each group contains the cell images which are considered to be the same. We use two tests to determine if two cell images are the same or not: Normalized Cross-Correlation and mean color difference based on Mahalanobis distance [1].

3.4. Classifying Cells

This step is to determine whether a cell image is a whiteboard, a stroke, or a foreground object. We use the following heuristics: 1) a whiteboard cell is uniform in color and is grey or white (i.e., the RGB values are approximately the same); 2) a stroke cell is mostly white or grey with one or two primary colors mixed in; 3) a foreground object does not have the characteristics above. The classification is therefore to determine whether the color distribution of the current cell image and the whiteboard color distribution are the same, or not the same but having strong overlap, or totally different. Again, we use Mahalanobis distance [1].

3.5. Filtering Cell Classification

Notice that the above classification algorithm only uses the color information in a single cell. More accurate results can be achieved by utilizing spatial and temporal relationship among the cell groups.

Spatial filtering. The basic observation is that foreground cells should not appear isolated spatially since a person usually blocks a continuous region of the whiteboard.

Temporal filtering. The basic observation is that it is virtually impossible to write the same stroke in exactly the same position after it is erased. In other words, if for any given cell, the cell images of two different frames contain the same stroke, then all the cell images in between the two frames must have the same stroke unless there is a foreground object blocking the cell.

3.6. Key Frame Image Extraction

Key frame images are the summary of a whiteboard session. The user would expect the key frame images to have the following properties: 1) They should capture all the important content on the board; 2) The number of the key frames should be kept to a minimum; 3) They should only contain the pen strokes and the whiteboard, but not the person in front; 4) They should have uniform white background and saturated pen colors for easy cut-and-paste and printing.

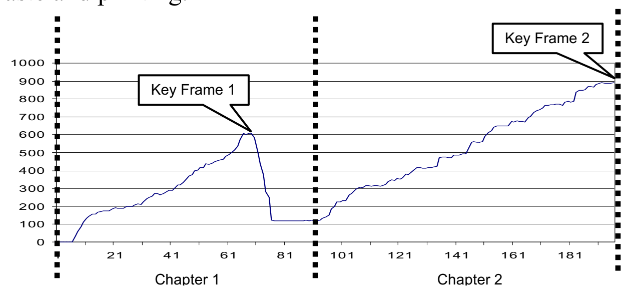


Figure 5: Key frame extraction: *A plot of the number of strokes vs. time for the sequence in Figure 2.*

The key frame extraction algorithm uses the cell images classification results from the previous step. We choose the frame just before an erasure starts as the key frame, which ensures that the content is preserved in those frames. *Figure 5* shows the stroke cell count plotted against frame number in the example session (*Figure 2*). The valleys in the curve divide the session into chapters. Once the frames are selected, we need to reconstruct the images corresponding to what the whiteboard looks like at these time instants. But we cannot simply use the raw images because they may contain foreground objects. We reconstruct the images by gathering the cell images in the sequence. Refer again to [2] for details.

3.7. Key Frame Color Balance

The reconstruction process removes people from the whiteboard images, but they still look like the raw images: grayish and noisy. The color balance process consists of two steps:

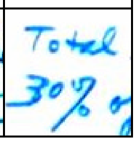
1. Make the background uniformly white and increase color saturation of the pen strokes. For each cell, the whiteboard color computed in Section 0 is used to scale the color of each pixel in the cell.
2. Reduce image noise. We remap the value of each color channel of each pixel in the key frames according to an S-shaped curve.

One of the two key frames from our running example is shown in *Figure 3*. The beginning and ending times of the chapters and the file names of their key frame images are saved in the index along with the time stamps of the strokes produced in Section 0.

4. SYSTEM PERFORMANCE AND USAGE

We have equipped three conference rooms with our system. Information about those rooms is listed in *Table 1*. The sizes of whiteboards in those rooms vary and so do the qualities of the key frame images produced. As we can see from the sample images, the writings on a 12'x5' board are fuzzier than the ones on the other two boards because the resolution is maxed out for a 4 mega-pixel input image. Nevertheless, they are still quite legible.

Table 1: Information about the three installation sites.

	Room 1	Room 2	Room 3
Board Dimension (ft)	4x3	8x5	12x5
Key Frame Image Dimension (pixel)	1200x900	2400x1500	2400x1000
Resolution (dpi)	25	25	16.7
Sample Images (80x80 pixels, approx. 96 pt font on the board)			

The analysis server runs on a PIII 800MHz dual CPU PC. It takes about 20 minutes for one hour recording. The installed systems are used frequently not only by our own team members but also by colleagues from neighboring groups. Over the course of 6 months, we recorded 108 sessions totaling 48 hours -- averaging 27 minutes per session and 4.5 sessions per week. The average number of key frames per session is 2.7.

All users of our system believe that our system is very useful for meetings that use whiteboard extensively. We are also pleasantly surprised by some users who found new ways to use the system which we did not intend initially. Take an example of status meetings which usually did not require writing on whiteboard. People still turned on the whiteboard capture system. When it was someone's turn to speak, the manager wrote his/her name on the board so that the speech segments could be easily found later in the recorded audio by clicking on the names in the key frame image. Another example is during a brainstorm session, when someone thought of a good idea, he wrote a star on the side of the board and said it aloud. The audio can then be retrieved later by clicking on the star.

5. CONCLUDING REMARKS

Meetings constitute a large part of knowledge workers' working time. Making more efficient use of this time translates to a big increase in their productivity. The work presented in this paper focuses on the meeting scenarios that use whiteboard heavily: brainstorming sessions, lectures, project planning meetings, patent disclosures, etc.

Our system alleviates the participants of those meetings the mundane note-taking task, so they can focus on contributing and absorbing ideas during the meetings. By providing key frame images that summarize the whiteboard content and structured visual indexing to the audio, our system helps the participants to review the meeting efficiently at a later time. Furthermore, the users who did not attend the meeting can often understand the gist of the meeting in a fraction of meeting time. From our initial deployment, we found the meeting participants started to find creative ways to use the recording and indexing capabilities of the system are discovered.

6. REFERENCES

[1] Duda, R., Hart, P. and Stork, D. *Pattern Classification*, Second Edition, John Wiley & Sons, New York, 2001.

[2] He, L., Liu, Z. and Zhang, Z. *Why Take Notes? Use the Whiteboard Capture System.* Technical Report MSR-TR-2002-89, Microsoft Research, Sept. 2002.

[3] Mimio. http://www.mimio.com/. Virtual Ink Corp.

[4] Rousseeuw, P. and Leroy, A. *Robust Regression and Outlier Detection*, John Wiley & Sons, New York, 1987.

[5] Saund, E. Image Mosaicing and a Diagrammatic User Interface for an Office Whiteboard Scanner. Technical Report, Xerox Palo Alto Research Center, 1999.

[6] SMARTBoard. http://www.smarttech.com/.

[7] WhiteboardPhoto. http://www.websterboards.com/.