

MARKER AND CHALKBOARD REGIONS

Michael N. Wallick, Rachel M. Heck and Michael L. Gleicher

{michaelw,heckr,gleicher@cs.wisc.edu}
University of Wisconsin-Madison
Department of Computer Science
1210 West Dayton Street
Madison, WI 53706 USA

ABSTRACT

When recording a classroom lecture, it is useful to capture the writing on the board for processing by applications. To provide more useful information for applications, we group together strokes based on the meaning of the writing. However, it is currently not possible to automatically obtain this meaning, or general semantic information automatically. Instead we rely on the structure of writing to approximate regions of the board, where each region represents a single thought or idea. These regions provide an abstraction of the board that can be used in several applications, such as a note taker, a lecture indexing program, an automatic video editor, and a program for creating multimedia presentations.

1. INTRODUCTION

Many lectures center around a lecturer writing on a board. Applications that process this writing can potentially add value to lecture recordings. For example, the writing might be re-arranged to provide a transcript of the lecture (automatic note-taking), elements of the writing can be used as indices into the lecture, or information about the writing can be used to aid the creation of novel presentations of the lecture.

In order to perform processing on the board, the writing must be represented at an appropriate granularity. The processing that can be done on the entire board at once is extremely limited. Likewise representing writing solely as individual strokes reduces the types of processing that can be done. Moving a single stroke can destroy the message that the writing conveys. For example, consider a stroke that represents a character in a word. Moving the stroke not only changes the word it was removed from but also the words it was moved close to. To perform nontrivial processing operations, strokes can be combined into groups that can be manipulated independently. We call such groups *regions*. Regions can serve as the atomic objects in board processing applications, ensuring that related writing is kept and processed together.

Ideally, a region would represent a complete thought or idea, for example a sentence or a diagram. Unfortunately, the information available for forming regions is quite limited. While spatial and temporal information about the individual strokes is provided by most board capture technologies, acquisition of higher-level information, such as word recognition or semantic interpretation, is often infeasible. Therefore, an automatic process for grouping writing into regions must rely on spatio-temporal data and heuristics about how people write on boards.

In this paper we present a method for grouping writing into regions. The region model is simple enough that regions can be obtained automatically and processed efficiently. Our method relies only on readily available information about the strokes and does not require writing interpretation. Unlike similar methods, our region model contains a temporal element, that provides information about the state of the region through time. In the event that no other capture technology is available, we present a method for obtaining strokes from a video of people writing on the board. We also present several sample applications that take advantage of the region concept.

The remainder of this paper is organized as follows. In the next section, 2, we discuss other efforts related to our regions concept. In Section 3 we formally define regions, presenting the lifespan of a region and heuristics for forming regions. Our algorithm for forming the regions is presented in Section 4. We describe several sample applications in Section 5.

2. RELATED WORK

Several researchers have considered the idea of grouping writing together on the board either in service of some larger specific application or as a first class object in itself. The major difference between our regions and the work of others is that regions provide a temporal model which describes the state of a region through time. The ScanScribe system by Saund et. al. [11] groups writing together, in a similar manner to regions. However, their system requires

manual user interaction to fine tune the groupings. Another system that uses a similar concept is Flatland [8], where a user draws on a markerboard; the writing is combined in a similar manner to what we describe. Flatland enables users to preserve and interact with their writing. The writing in Flatland is represented by a bounding box around the writing. Our region model allows more advanced interaction with the regions and provides methods for post processing the board and writing, as well as providing a more advanced temporal model that the region moves through.

Onishi et al. [9] introduced a similar region concept. In their work, a blackboard is segmented into areas of “written rectangles” and “written regions.” A written region is the combination of several written rectangles. The rectangles are formed by using edge detection to find stationary edges, which is assumed to be writing. Our implementation uses a different algorithm for extracting the writing from the video. Additionally, our model of regions allows for more complex interaction, including erasing and merging and provides a more rich temporal model.

Recently, other researchers have developed concepts similar to regions that are designed to be treated as first-class objects that can be used in a variety of projects. This is our goal with regions. Lui and Kender [7] introduced the concept of a “learning object” as a first-class partitioning of the board. Our region model differs in that regions, unlike the teaching object, have a richer temporal model that accounts for continued writing, erasing and merging together; this is described in the next section.

Other researchers have looked at using the whiteboard as an interface device to a computer. For example, both ZombieBoard [1, 10] and BrightBoard [12] monitors the writing on the board looking for specific symbols that trigger events. These systems use a region-like concept of grouping writing and then identify the symbols within that writing.

3. STROKES AND REGIONS

We employ two concepts: strokes and regions. A stroke is the writing or erasing that occurs between pen-down and pen-up operations. When using video this can be estimated as the writing that occurs over a small amount of time. *A region is the collection of all of the strokes that describe the same idea.* Consider, for example, a sentence written on the board. The letters, or parts of the letters, are the strokes, and the entire sentence is the region. In this section, we discuss these concepts in greater detail.

For our analysis, we abstract each stroke as its axis-aligned bounding box in time and space. As such we represent a stroke as a six-tuple. The first four are the size and location of the stroke. The other two entries are the time that

the stroke occurs and if it was writing or erasing. We represent strokes in terms of their bounding box rather than the actual pixels that the stroke covers because a bounding box is compact and easier to work with, and the precision that is lost is not necessary for our purposes.

Related strokes are combined to form regions. The regions contain temporal information, represented as a “life-history.” Each region experiences each of the following events in order:

Birth A region is born when the first stroke of writing is drawn on the board.

Maturity A region is mature when the last stroke of writing is drawn on the board. During the time between birth and maturity the region is said to be growing or immature.

Death A region dies under one of three conditions: the region is erased, the lecture ends, or the regions merges with another region that is close by.

Heuristics are used to determine if strokes are related using only spatio-temporal information. These heuristics are based on the natural structure of writing that communicates a message. Our heuristics are as follows:

1. Only one idea will be presented at a time; it is difficult and confusing to explain multiple ideas at once.
2. Writing pertaining to a specific idea will appear close together in time and space; it would not make sense to have large gaps between related strokes.
3. There will be either a temporal pause or physical space between different ideas; an instructor uses this spacing as a cue to the audience that a new idea is coming.

In our model, we only allow one region to be immature at a time. This is based on the heuristic that only one idea will be explained at a time. The simplified model is still capable of handling complex cases correctly by allowing regions to be merged later. For example, an instructor may write two separate parallel list. The alternatives to regions would handle this case as one large block of writing. Our region model would have each entry in the list as a distinct regions that would ultimately merge into two regions, one for each list.

4. ALGORITHM FOR BUILDING REGIONS

In order to form the regions from a video we need to know each stroke’s axis-aligned bounding box, the time that the stroke appeared, and if it was writing or erasing. Obtaining this information can be done with the use of specialized hardware, such as a markerboard scanner [13] or an

electronic markerboard [2]. However, this equipment can be expensive and cannot be easily moved. As an alternative, strokes can be extracted from a video of the lecture. Video cameras are less expensive and more portable than specialized hardware. We present a method for extracting strokes from a video and a method for combining the strokes to form regions. It should be noted that the strokes used to form the regions can come from any mechanism, not just our video technique.

Our algorithm works in three steps (described in this section). The first is to segment the board, or remove the instructor or anyone else who may be obstructing the board at any point in the video. The second step is finding the individual strokes of writing or erasing that occur in the video. Finally, we combine the strokes together, based on the heuristics described in Section 3 to form the regions.

4.1. Segmenting the Board

In our implementation we use a video camera placed in the back of a classroom to record the board during a lecture. The strokes cannot be directly detected from the video because the instructor (or other students) are often blocking the board. To correct this, we segment the video to determine which pixels are the board and use inpainting to remove occlusions.

In order to perform segmentation, we use color classification to identify instructor and board pixels. Color was chosen for several reasons: the board is generally a unique color that simplifies finding non-board pixels; segmentation based on color will work with little motion or on single images, such as those acquired with a high-resolution digital camera; and color classification will work with multiple occlusions (i.e. other participants or more than one presenter). Color classification can fail if the instructor is wearing clothing that is a similar color to that of the board. This problem does not occur often enough to warrant concern. We segment the video by manually marking the location of the board and identifying pixels as being either board colored or not. A new video is created (which we refer to as the segmented video) by replacing any non-board pixels with a board pixel from the same location in a later or previous frame.

To perform color classification we extended the algorithm described by Jones and Rehg [6]. Their method uses three dimensional histograms manually populated by pixels known to belong to the object each represents. For example, a board histogram is populated with known board pixels. For classification, unknown pixels are looked up in the objects histogram. Because the camera is not moving and we know the location of the board, we are able to automatically populate the color histogram. To do this, we select several training frames from the video. If a pixel in the board location remains relatively constant between

training frames that pixel is used to populate the board histograms, otherwise it is used as instructor histogram.

4.2. Finding Strokes From Video

Once the board has been segmented, we are able to find strokes of writing. For our purposes, a stroke is the writing that occurs over a small amount of time, such as a few seconds. Strokes are found by subtracting two frames from the segmented video that are close in time. Difference between the two frames indicates that there has been some change on the board: writing or erasing. If there is a difference, we repeat the process, decreasing the space between the two frames, until there is no longer a difference. This gives a better estimate of the time that the stroke occurs. Figure 1 shows strokes found over time.

To distinguish writing from erasing, we note that writing creates high frequencies and erasing removes them. Using this fact, a stroke can be identified as either writing or erasing by applying a high pass filter to the area of difference in the later frame. The presence of high frequencies indicates that there has been writing; likewise, the lack of high frequencies means that there has been erasing.

Figure 2 shows a flowchart of the stroke finding algorithm.

4.3. Finding Regions

Each stroke is processed in chronological order to form the set of regions. The rules described earlier in this paper are applied to advance the system. These rules are based on the type of stroke and its proximity to the region that is currently being formed, or the Active Region. Each region contains the information about its birth, maturity and death. Maturity and death are updated as the strokes are processed.

1. If there is no Active Region and a write stroke appears a new region is created and is considered the Active Region. The dimensions of the stroke are used for the dimension of the region. The birth and maturity times are set to the time the stroke occurs, and death is set to the end of the video.
2. Whenever a write stroke occurs far from the Active Region in either time or space, the Active Region is made mature, and a new region is created, becoming the Active Region. The original dimensions of the stroke are used for the new region. The birth and maturity times are set to the time that the stroke occurs, and the death is set to the end of the video.
3. A stroke that occurs close to the Active Region in time and space extends the Active Region. The dimensions of the Active Region are increased to include the current stroke, and the maturity time is changed to when the stroke occurs.



Figure 1: Individual strokes found.

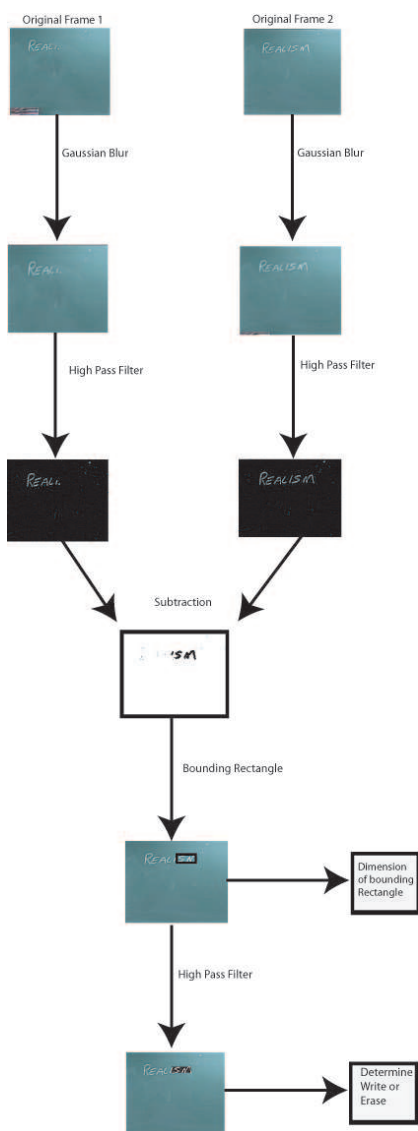


Figure 2: Flowchart of stroke finding algorithm.

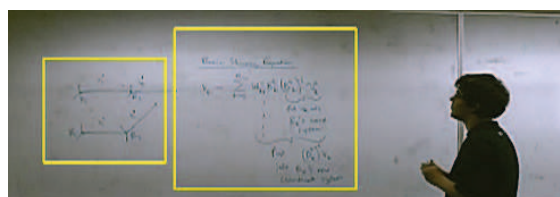


Figure 3: Regions from a markerboard lecture.

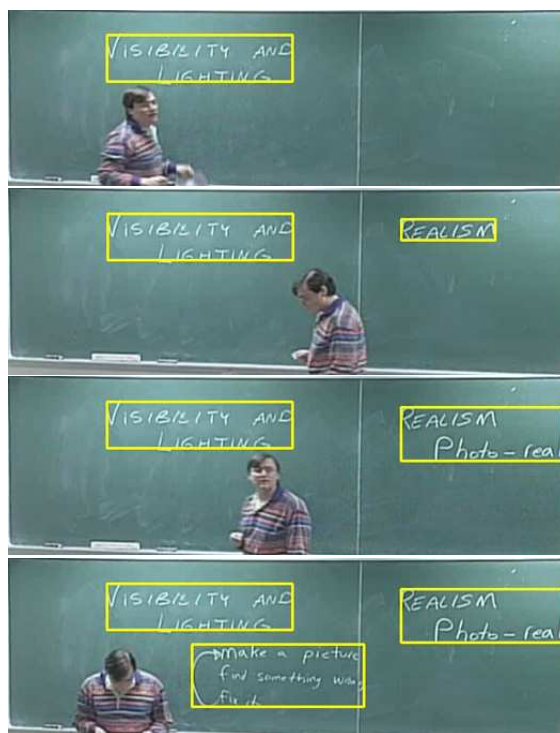


Figure 4: Regions found on a chalkboard.

4. When an erase stroke is detected, all existing regions are inspected. If the erase stroke greatly overlaps a region, then the region is erased, and the death time of the region is set to the time of the stroke.
5. Once a region becomes Mature (another region is born or sufficient has passed since the last stroke was added), we check for any live regions that are significantly close to the recently matured region. If one is found, then the two regions are merged together. The older region has its death time set to the birth of the younger region, and the boundary of younger region is extended to include both regions.

Subtracting a frame of segmented video from before the birth of the region from a frame after the maturity of the region can be used to find the exact dimensions of the region. The difference between the two frames represents all of the writing in the region; this technique defends against errors in the stroke finding process.

For our experiments, we use 10 seconds as the cut-off for close in time, approximately 12 inches for a stroke to be close to the Active Region, and 3 inches as the maximum distance between merged regions. The distance is measured as the shortest Euclidean distance between regions. Figures 3 and 4 show results from our system. The regions in Figures 3 were found on a markerboard, and the regions in Figure 4 were found on a chalkboard.

5. USES OF REGIONS

With a few simple extensions to the region model, we are able to develop several applications that make use of a chalk or markerboard as input. A few of these example applications include a virtual video editor, an automatic note taker, a lecture indexing system, and an application for developing multimedia presentations from a video. Although we do not discuss it in this paper, it is easy to imagine how to replace the board unit representations of the projects described in Section 2 with our regions.

We first developed the idea of regions to support our project *Virtual Videography* [3, 4], an automatic video editing system. For this project an unattended camera in the back of a classroom records a lecture. Regions are used as a basic building block for creating video. They provide a convenient means of framing the correct parts of the board, and ensure that writing that should be shown in the final video will not get cut off. A virtual editor uses the the regions, combined with other passively acquired cues from the instructor, such as pointing to a region, to decide what part of the board to focus on. Figure 5 shows example output from our video editing system.

Since each region represents a thought or idea written on the board, they can be employed as an automatic note-



Figure 5: (Top Row) Original Video. (Bottom Row) Edited video output from the Virtual Videography System.

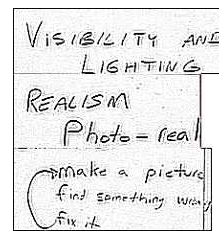


Figure 6: Output of a region-based automatic note taker. The regions were ordered based on birth time.

taking device. Using stroke information alone, a note-taker is limited to displaying the entire contents of the board at any given time. Our implementation of a note-taker uses an image of each mature region. A small amount of image processing is applied to the output in order to make the chalk appear black against a white background. The result is a linear record of the writing that has appeared on the board. Figure 6 shows an example of the output of the region-based note-taker.

Regions also provide a natural index into a lecture. In response to this observation, we developed a video playback program based on regions. In addition to the standard VCR controls of a playback program, the viewer is also presented with a a thumbnail image for each region. Clicking on one of the regions causes the playback



Figure 7: Screen shot from a region-based playback program. The top image shows the video being played, and the bottom part shows all the regions in the video.



Figure 8: A region of writing (left) was replaced with type written text (right).

to jump to the birth of that region. The region list acts as a visual table of contents for the lecture. Figure 7 shows an example of our playback program. The top part shows the current video playback, and the bottom shows the different regions that are available to the user. The whiteboard capture system [5] being developed at Microsoft Research uses a method for automatically capturing a markerboard. During playback, the viewer may click on a stroke of writing to be taken to that point in the video. This allows a finer granularity of indexing, but makes it harder to navigate the video at a level of thoughts and ideas.

The ability to associate information with a region makes creating multimedia presentations easier. Someone who is familiar with the lecture may wish to provide a pointer to a program or document for each region. A lecture talking about our region model may include a link to this paper. Keywords may be associated with each region, giving a viewer a means of searching through a video.

In addition to augmenting the regions with information, someone who is familiar with the lecture can completely replace the region with a new image, video or animation. Replacing the writing on the board can enhance the video by adding more information or simply make the writing easier to read. Figure 8 shows redrawn regions.

6. CONCLUSION

In this paper we presented a concept called “regions.” These regions are formed by combining strokes of related writing from a classroom lecture. Ideally the regions are representative of a single thought or idea written on the board. Because current technology does not provide a means of automatically extracting this information, we make use of temporal and spatial heuristics based on the natural structure of writing to form the regions.

Regions provide an approximation of semantic understanding that is not possible to obtain with strokes of writing alone. This approximation can be used in a wide range of educational applications. These include automatic note-taking, providing an index to a lecture, automated video editing, and multimedia presentation creation.

Acknowledgements This work was supported in part by NSF grants CCR-9984506 and IIS-0097456, Microsoft Research, and equipment donations from Intel. Michael Wallick is supported by a Microsoft Research Fellowship

References

- [1] M. Black, F. Berard, A. Jepson, W. Newman, E. Saund, G. Socher, and M. Taylor. The digital office: Overview, 1998.
- [2] Brother. Brother cypoint whiteboard. Computer Hardware Product, 2002.
- [3] Michael Gleicher and James Masanz. Towards virtual videography (poster session). In *Proceedings of the eighth ACM international conference on Multimedia*, pages 375–378. ACM Press, 2000.
- [4] Michael L. Gleicher, Rachel M. Heck, and Michael N. Wallick. A framework for virtual videography. In *Proceedings of the 2nd international symposium on Smart graphics*, pages 9–16. ACM Press, 2002.
- [5] Li-wei He, Zicheng Liu, and Zhengyou Zhang. Why take notes? use the whiteboard capture system. Technical Report MSR-TR-2002-89, Microsoft Research, September 2002.
- [6] Michael J. Jones and James M. Rehg. Statistical color models with application to skin detection. Technical Report CRL 98/11, Cambridge Research Laboratory, 1998.
- [7] Tiecheng Liu and John R. Kender. Spatial-temporal semantic grouping of instructional video content. In *Image and Video Retrieval, Second International Conference, CIVR, 2003*.
- [8] Elizabeth D. Mynatt, Takeo Igarashi, W. Keith Edwards, and Anthony LaMarca. Flatland: new dimensions in office whiteboards. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 346–353. ACM Press, 1999.
- [9] M. Onishi, M. Izumi, and K. Fukunaga. Blackboard segmentation using video image of lecture and its applications. In *International Conference on Pattern Recognition*, pages Vol IV: 615–618, 2000.
- [10] Eric Saund. Bringing the marks on a whiteboard to electronic life. In *Proceedings of CoBuild; Second International Workshop on Cooperative Buildings*, 1999.
- [11] Eric Saund, James Mahoney, David Fleet, Dan Larner, and Edward Lank. Perceptual organization as a foundation for intelligent sketch editing. *2002 AAAI Spring Symposium on Sketch Understanding*, 2002.
- [12] Quentin Stafford-Fraser and Peter Robinson. Brightboard: a video-augmented environment. In *CHI '96: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 134–141. ACM Press, 1996.
- [13] Virtual Ink Corporation. Mimio, 2004. Computer Hardware Product. <http://www.mimio.com/index.shtml>.