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 writ-
ing. However, it is currently not possible to automatically
obtain general semantic information from a video. Instead
we rely on the structure of writing to approximate regions
of the board that contains 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 lec-
ture 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 (auto-
matic 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 en-
tire board at once is too much data. Likewise represent-
ing writing solely as individual strokes limits the types of
processing that can be done. Moving a single stroke can
destroy the message that the writing conveys. For exam-
ple, consider a stroke that represents a character in a word.
Moving the stroke not only changes the word it was re-
moved 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 indepen-
dently. We call such groups regions. Regions can serve as
the atomic objects in board processing applications, en-
suring 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 lim-
lited. While spatial and temporal information about the
individual strokes is provided by most board capture tech-
nologies, acquisition of higher-level information, such as
word recognition or semantic interpretation, is currently
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 re-
gions can be obtained automatically and processed effi-
ciently. Our method relies only on readily available in-
formation about the strokes and does not require writing
interpretation. In the event that no other capture technol-
ogy 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 related efforts to our regions
concept. In Section 3 we formally define regions, pre-
senting the lifespan of a region and heuristics for forming
regions. We describe several sample applications in Sec-
tion 5. Our algorithm for forming the regions is presented
in Section 4.

2. RELATED WORK
Several researchers have considered the idea of grouping
writing together on the board in service of some larger
specific application. In the ScanScribe system presented
by Saund et. al. [11], the writing in a digital document is
combined together into perceptual groups similar to our
region concept, in that they represent atomic units of writ-
ing. The user then manipulates the individual groups to
change the document. Their system requires manual user
interaction to fine tune the groups.
Another system that uses a similar concept is Flatland [7],
where a user draws on a markerboard; the writing is com-
bined 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. Other researchers have looked at using the whiteboard as an interface device to a computer. For example both ZombieBoard [9, 10] and BrightBoard [12] monitors the writing on the board looking for specific symbols that trigger events.

Onishi et al. [8] introduced a similar 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, merging, as well as having a richer temporal model.

Recently, some researchers have been developing concepts similar to regions that are designed to be treated as first-class objects which can be used in a variety of projects. This is our goal of regions. Lui and Kender [6] introduced a concept of a “learning object” as a first-class partitioning of the board. Our region model differs in that regions, unlike the teaching object, has a richer temporal model that accounts for continued writing, erasing and merging together; this is described in the next section.

3. STROKES AND REGIONS

We use 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 a collection of all the strokes that describe the same idea. Consider, for example, a sentence written on the board. 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. Each region passes through a specific “life-history.” Each region experiences each of the following events, in the specific 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 lists. The alternatives to regions (in Section 2) would handle this case as one large block of writing. Our region model would cause 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 for 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 [1]. 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 occurs in the video. Finally, we combine the strokes together, based on the heuristics described in Section 3, to form the regions. If using a Mimio or some other board capture device, then the first two steps can be skips, since the strokes of writing are captured automatically.

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 chalkboard 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). While 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 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 frame. To perform color classification we extended the algorithm described by Jones and Rehg [5]. 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 consecutive training frames that pixel is used to populate the board histograms, otherwise it is used for the 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.

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 being 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 1 shows a flowchart of the stroke finding algorithm.

4.3. Finding Regions

The region finder uses a set of strokes, provided by the method in the previous section or a capture device. Strokes are combined using heuristics based on the structure of writing.

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.

- 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.

- Whenever a write stroke occurs far from the Active Region in either time or space, the Active Region is considered mature, and a new region is created, becoming the Active Region. As above, 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.

- 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 is changed to when the stroke occurs.
Figure 1: Flowchart of stroke finding algorithm.

Figure 2: Individual strokes found.

Figure 3: Regions from a markerboard lecture.

Figure 4: Regions found on a chalkboard.
• 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 erase stroke.

• Once a region becomes Mature, we check for any live regions that are significantly close to the recently matured region. If so, 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.

If video is used to find the strokes we can determine the exact dimension of the region by subtracting a frame of segmented video from before the birth of the region from a frame after the maturity 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 two regions.

Figures 3 and 4 show results of 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 [2, 3], 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 passive cues from the instructor 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-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 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.

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.
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 [4] 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 beneficial to a wide range of educational applications. These include automatic note-taking, providing an index to events in 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 IBM, NVidia and Intel. Michael Wallick is supported by a Microsoft Research Fellowship.

References