Human-Centered Webcasting of Interactive-Whiteboard Lectures

Gerald Friedland and Raúl Rojas
Institute of Computer Science, Freie Universität Berlin
{fland|rojas}@inf.fu-berlin.de

Abstract

In our system for recording and transmitting lectures over the Internet, the board content is transmitted as vector graphics, producing thus a high quality image, while the video of the lecturer is sent as a separate stream. It is easy for the viewer to read the board but the lecturer appears in a separate window. As a result, two areas of the screen are competing for the viewer’s attention, causing the widely known split attention effect. To eliminate this problem, the lecturer is extracted from the video stream and his or her image is pasted onto the board image at video stream rates. The lecturer can be dimmed from opaque to semitransparent, or even transparent. The article presents a detailed analysis of the underlying psychological problems and explains the multimedia techniques that are applied to achieve the solution.

1 Webcasting Chalkboard Lectures

If one wants to webcast a regular chalkboard presentation held in a classroom or lecture hall there are mainly two ways to do this.

One possibility is to take a traditional video of the chalkboard together with the lecturer acting in front of it and then use standard webcasting products, such as Microsoft Windows Media or RealMedia to transmit the video into the Internet. The primary advantage of broadcasting a lecture this way is that the approach is rather straightforward: The setup for capturing a lecture is well-known and off-the-shelf Internet broadcasting software is ready to be used for digitizing, encoding, transmitting, and playing back the classroom event. Furthermore, the lecturer’s workflow is not disturbed and there is no need to become accustomed to new devices. Even though some projects have tried to automate the process [9, 18], a major drawback of recording a lecture the “conservative way” is that it requires additional man power for camera and audio devices operation. Yet the video compression techniques used by traditional video codecs are not suitable for chalkboard lectures: Video codecs mostly assume that higher frequency features of images are less relevant. This produces either an unreadable blurring of the board handwriting or a bad compression ratio. Vector format storage is not only smaller, it is also favorable because semantics is preserved. After a lecture has been converted to video, it is for example not possible to delete individual strokes or to insert a scroll event, without recalculating and rendering huge parts of the video again. Some projects have therefore tried to recognize board content automatically, see for example [24]. In most cases, however, this is hard to achieve, because chalkboard drawings are sometimes also difficult to read because of low contrast. Figure 1 shows an example of a traditional chalkboard lecture webcast with a commercial Internet broadcasting software.

Knowing the disadvantages of the conservative approach, several researchers have investigated the use of pen-based computing devices, such as interactive whiteboards or Tablet PCs to perform lecture webcasting (see for example [1]). Using a pen based device provides an interesting alternative because it captures the handwriting and allows to save it in a vector-based format. Vector-based information requires less bandwidth, can be transmitted without loss of semantics, and is easily rendered as a crisp image on a remote computer. Still, a disadvantage is the low resolution of these devices and the requirement for professors to change some teaching habits and technical accessories. One of the systems that supports the creation of remote lectures held using a pen input device, is the E-Chalk system [8], created in 2001. E-Chalk records the creation of the board content together with the audio track of the lecturer and transmits both synchronized over the Internet. The lecture can be received remotely either using a Java Applet client or using MPEG-4.

2 Handwriting only is not Satisfying

During an evaluation, many students reported they found it disturbing that the handwritten objects on the board appear from the void during distance replay. The lecture appears impersonal because there is no person acting in front of the board. The replay lacks important information be-
cause the so-called “chalk and talk” lecture actually consists of more than the content of the board and the voice of the instructor. Often, facial expressions of the lecturer bespeak facts beyond verbal communication and the instructor uses gestures to point to certain facts drawn on the board. Sometimes, it is also interesting to get an impression of the classroom or lecture hall. Psychology suggests (see for example [14]) that gestures and facial expressions are part of a person’s semantic of encoding ideas. The understanding of words partly depends on gestures as they are also used to interpret and disambiguate the spoken word [16]. These shortcomings are highlighted when board activity is temporarily abandoned for verbal-only explanations or even non-verbal communication. In order to transport this further information to a remote computer, we added an additional video server to the E-Chalk system. As shown in Figure 2, the video pops up as a small additional window during lectures replay. The importance of the additional video is also supported by the fact that several other lecture recording systems (compare for example [11]) have also implemented this functionality and the use of an additional instructor or classroom video is also widely discussed in empirical studies. Not only does an additional video provide non-verbal information on the confidence of the speaker at certain critical points, like irony [5], several experimental studies (for an overview refer to [13]) have also provided evidence that showing the lecturer’s gestures has a positive effect on learning. For example, Fey [6] has reported that students are better motivated when watching lecture recordings with slides and video in contrast to watching a replay that only contains slides and audio. Glowalla [10] also shows in a comparative study that students usually prefer lecture recordings with video images over those without.

3 Split Attention

The video of the instructor conveys non-verbal information that several empirical studies have shown to be of value for the student. There are, however, several reasons against showing a video of the lecturer next to slides or the blackboard visualization. The video shows the instructor together with the board content: In other words the board content is actually transmitted redundantly. On low resolution devices, the main concern is that the instructor video takes up a significant amount of space. The bigger the video, the better can non-verbal information be transmitted. Ultimately, the video must have the size of the board to convey every bit of information. Let alone layout constraints, as the board resolution increases because electronic whiteboards become better, it is more and more impossible to transmit the video side-by-side with the whiteboard content. Even though there still might be solutions for these layout issues, a more heavily discussed topic is the issue of “split attention”.

In a typical E-Chalk lecture with instructor video there are two areas of the screen competing for the viewer’s attention: the video window showing the instructor, and the board or slides window. Glowalla [10] tracked the eye movements of students while watching a lecture recording
that contains slides and an instructor video. His measurements show that a student spends about 70 percent of the time watching the instructor video and only about 20 percent of the time watching the slides. The remaining 10 percent of the eye focus was lost for activities unrelated to lecture content. When the lecture replay only consists of slides and audio, students spend about 60 percent of the time looking at the slide. Of course, there is no other spot to focus attention on in the lecture recording. The remaining 40 percent, however, were lost in distraction. The results are not directly transferable to electronic whiteboard-based lecture replays because the slides consist of static images and the whiteboard window shows a dynamic replay [15]. However, motion is known to attract human attention more than static data (see for example [12]) it is therefore likely that the eyes of the viewer will focus more often on the board content, even when a video is presented. Although the applicability of Glowalla’s study to chalkboard lectures is to be proven yet, the example shows, that on a typical computer screen two areas of the screen may well be competing for attention. Furthermore, it makes sense to assume that alternating between different visual attractors causes cognitive overhead. Cooper [4] already discussed this issue and provided evidence that “Students presented a split source of information will need to expend a portion of their cognitive resources mentally integrating the different sources of information. This reduces the cognitive resources available for learning.”

The conclusion leads to an enhanced solution for the transmission of non-verbal communication of the instructor in relation to the electronic whiteboard content: The instructor is filmed acting in front of the board with a video camera and is then separated by a video segmentation approach. The image of the instructor may then be overlaid on the board, creating the impression that the lecturer is working directly on the screen of the remote student. Facial expressions and gestures of the instructor appear in direct correspondence to the board content. Pasting the instructor on the board also reduces bandwidth and resolution requirements. Moreover, the image of the lecturer can be made opaque or semi-transparent. This enables the student to look through the lecturer. In the digital world, the instructor does not occlude any board content, even if he or she is standing right in front of it. In other words, the digitization of the lecture scenario solves another “layout” problem that occurs in the real world (where it is impossible to solve).

4 Related Approaches

The importance of transmitting gestures and facial expressions is not specific to remote chalkboard lecturing. In a computer-supported collaborative-work scenario people first work together on a drawing and then want to discuss it by pointing to specific details of the sketch. For this reason, several projects have begun to develop means to present gestures in their corresponding context.

Two early projects of this kind were called VideoDraw [23] and VideoWhiteboard [22]. On both ends of the transmission a person can draw atop a monitor using whiteboard pens. The drawings together with the arms of the drawer were captured using an analog camera, so that each side sees the picture of the remote monitor overlaid on their own drawings. Polarizing filters were used to omit video feedback. The VideoWhiteboard uses the same idea, but people are able to work on a large upright frosted-glass screen and a projector is used to display the remote view. Both projects are based on analog technology without any involvement of the computer. Modern approaches include a solution presented in [17] that uses chroma keying for segmenting the hands of the acting person and then overlaying it on a shared drawing workspace. In order to use chroma keying, people have to gesture in front of a solid blue surface and not in front of their drawing. This has been reported to produce confusion in several situations. LIDS [2] captures the image of a person working in front of a shared display with digital cameras. The image is then transformed via background subtraction into a frame containing the whiteboard strokes and a digital shadow of the person.

The VideoArms project by [21] works with touch-sensitive surfaces and a web camera. After a short calibration, the software extracts skin colors and overlays the extracted pixels semi-transparently over the image of the display. This combined picture is then transmitted live to remote locations. The system allows multi-party communication. Tang [20] presents an evaluation of the VideoArms project. They argue that the key problem is still a technical one: “VideoArms’ images were not clear and crisp enough for participants. [...] The colour segmentation technique used was not perfect, producing on-screen artifacts or holes and sometimes confusing users”. In summary, the
presented approaches either tried to work around object extraction, or the technical requirements for the segmentation made the systems suboptimal. It is therefore important that the lecturer extraction approach is both easily used in classroom and/or after a session and technical requirements do not disturb the classroom lecture.

5 Instructor Extraction

In E-Chalk, the principal scenario is that of an instructor using an electronic whiteboard in front of the classroom. The camera records the instructor acting in front of the board, such that exactly the screen showing the board content is recorded. With a zoom camera this is easily possible from a non-disturbing distance (for example from the back of the classroom) and lens distortion is negligible. In this article, it is assumed that the instructor operates using an electronic whiteboard with a rear projection (for example a StarBoard) rather than a front projection. The reason for this is that when a person acts in front of the board and a front projector is used, the board content is also projected onto the person. This makes a segmentation very difficult and the projected board artifacts disturb the appearance of the lecturer in the final result. Once setup, the camera does not require operation by a camera person. In order to ease segmentation, light changes and (automatic) camera adjustments should be inhibited as much as possible.

The approach presented here is based on the following assumptions: The hardware is setup as described above, the colors of the instructor image are overall different from those in the rest of the image, and during the first few seconds after the start of the recording, there is only one instructor and he or she moves in front of the camera.

The core idea behind the approach presented here is based on the notion of a color signature. A color signature models an image or part of an image by its representative colors. The idea behind using color signatures for segmentation is to provide a means for abstraction that sorts out individual outliers caused by noise and small error.

We have published a generalized version of the algorithm under the name SIOX (Simple Interactive Object Extraction) [19]. It can be used for various segmentation tasks and has been implemented as a low-interaction still-image segmentation engine into the open-source image manipulation programs GIMP and Inkscape. Further details on the following algorithm and an evaluation of the robustness of the approach including benchmark results can be found in [7].

5.1 Gathering Background Samples

The input is a sequence of digitized YUV or RGB video frames either from a recorded video or directly from a camera. The first step of the algorithm is to convert each frame to the CIELAB color space [3]. In practice, the Euclidean distance between two colors in this space better approximates a perceptually uniform measure for color differences than in any other color space, like YUV, HSI, or RGB. It is hard to get a background image for direct subtraction. The instructor can paste images or even animations onto the board and when the instructor scrolls a page of board con-
tent upwards, the entire screen is updated. However, the instructor sometimes stands still producing less changes than the background noise. The idea is thus to extract only a representative subset of the background that does not contain any foreground for further processing. A simple model is used that is described in [7]. Figure 3 shows an example of the output of the motion statistics processing.

5.2 Building a Model of the Background

A color signature is a set of representative colors, not necessarily a subset of the input colors. While the set of background samples from section 5.1 typically consists of a few hundreds of thousands of colors, the following clustering reduces the background sample to its representative colors, usually about a few hundreds. The known background sample is clustered into equally sized clusters because in CIELAB space specifying a cluster size means to specify a certain perceptual accuracy. To do this efficiently, we use a modified k-d tree clustering described in [7]. Given a certain pixel, all that has to be done is to traverse the tree to find out whether it belongs to one of the known background clusters or not. Figure 4 shows a sample color signature. A biggest connected component search is to be performed. The biggest connected component is considered to be the instructor, and all other connected components (mostly noise and other moving or newly introduced objects) are eliminated from the output image. Figure 5 shows two sample frames of a video where the instructor has been extracted as described here.

6 Results

The resulting segmented video is scaled to fit the board resolution (usually $1024 \times 768$) and is pasted over the board content at the receiving end of the transmission or lecture replay. Since a static camera is used that captures the same rectangular area that is covered by the whiteboard, the instructor can be directly pasted without any further transformations. The current Java-based prototype implementation processes a $640 \times 480$ video at 25 frames per second.

Two examples of lectures that contain an extracted and overlaid instructor can be seen in Figure 6. Our experiments show that this approach is feasible and also esthetically appealing. The superimposed lecturer helps the student to better associate the lecturer’s gestures with the board contents. Pasting the instructor on the board also reduces space and resolution requirements. This makes it also possible to replay a chalkboard lecture on mobile devices (see Figure 7).

As the algorithm focuses on the background it provides rotation and scaling invariant tracking of the biggest moving object. The tracking still works when the instructor turns around or when he leaves the scene and a student comes up to work on the board. Once initialized, the instructor does not disappear, even if he or she stands absolutely still for several seconds (which is actually very unusual). Reflections on the board display are mostly classified as background and small moving objects never make up the biggest connected component. For the background reconstruction process to collect representative background pixels it is not necessary to record a few seconds without the instructor. The only requirement is, that for the first few seconds of initialization the lecturer keeps moving and does not occlude background objects that differ significantly from those in the other background regions.

Figure 6. The final result: The instructor is extracted from the original video (left) and pasted semi-transparently over the vector based board content (right).

Figure 7. Lecture replay using the video capabilities of small devices. Left: A Symbian OS based mobile phone. The resolution is $176 \times 144$ pixels. Right: A video iPod.
7 Conclusion

This article proposes to change the way interactive-whiteboard lecture webcasts are to be transmitted. The standard side-by-side replay of video and blackboard content causes technical and cognitive problems. We propose to cut the lecturer out of the video stream and paste it on the rendered image of the board and realized an algorithm that does this. In the result, the board content can be transmitted as a crisp image while facial expressions and body gestures of the instructor appear in direct correspondence to it without causing a split attention effect. The remote lecture becomes more human-centered.

References


