# Chapter 12

# Conclusion

## 12.1 Summary

This dissertation presents an audio and video recording and transmission system for lectures held with an electronic chalkboard that was developed in conjunction with the E-Chalk project. Related work builds upon standard transmission and archiving methods that were implemented for traditional radio and TV stations. E-Chalk's audio and video systems were created under the assumption that computers can be better utilized to facilitate the creation of multimedia content in a more automated and yet integrated fashion. The result permits distance teaching to be a side effect of the use of the electronic chalkboard in the classroom. Moreover, most of the presented research is general and can also be used in different application areas. The contributions can be summarized as follows:

- The dissertation starts with discussing the underlying architectural ideas of the E-Chalk system for both server and client side. A novel component-oriented software framework is introduced under the name SOPA. SOPA is based on a component-management framework based on the OSGi standard and a component-deployment framework. On top of these two frameworks, SOPA provides a component-assembly mechanism that makes the creation of typical multimedia streaming and processing applications easier. It supports collaborative extension and updating of the system while reducing administrative overhead. SOPA allows to rapidly combine E-Chalk with other streaming applications and to integrate new formats and content types easily.

- The second part presents E-Chalk's audio system and its evolution from a system that aimed to provide a solution to broadcast traditional radio programs over the Internet to a fully integrated component inside E-Chalk. The system was developed while used at different educational institutions in response to feedback and requests from users. One of the major issues concerns the resulting audio quality when automatically recording in classrooms or lecture halls. A system called Active Recording was created that simulates the typical work of an audio technician to help instructors create better speech recordings. The system measures certain critical factors

in the sound equipment, monitors possible malfunctions during recording, and filters out typical audio distortions.

- Then, E-Chalk's video system is described. While traditional video codecs are not suitable for the transmission of chalkboard content, the exclusive replay of vector-based board strokes is suboptimal, too. According to our experience, which is also backed by several studies, the remote student needs to see the image of the instructor because it conveys important contextual information. A mere side-by-side replay of the video and the board content, however, results in an ergonomic issue called the split attention problem. This dissertation presents a solution where the instructor is filmed in front of the electronic board and his or her image is automatically cut out in order to be pasted semi-transparently over the vector-based board image during remote replay. This method solves the split attention problem as well as several layout issues and also eases replay on small devices, such as mobile phones.

- Finally, a generalization of the instructor-extraction method, called SIOX, is presented. It enables semi-automatic segmentation in image and video manipulation programs as well as the improvement of 3D-time-of-flight camera segmentation results. A thorough evaluation of the solution and its implementation in several common open-source applications is presented.

## 12.2 Future Work

Although the presented system already uses methods that go beyond standard computer-based audio and video recording techniques, many problems remain, and the following paragraphs present a small selection of them.

Chapter 5 comes to the conclusion that no matter what current format would be used, the transmission of the overlaid instructor requires at least a DSL or cable connection. However, if only gestures without facial expressions are transmitted, the overall bandwidth (including audio) can be brought down to less than 64 kbit/s. This is achieved by transmitting only the outline of the instructor as shown in Figure 12.1. The idea was inspired by the Italian cartoon figure, "La Linea", by Osvaldo Cavandoli. Of course, the usefulness of such a transmission is still to be evaluated. However, encoding the shape of the instructor as a polyline allows direct application of standard online-character-recognition methods. This would allow for a recognition of some of the instructor's gestures by the computer. Certain shape patterns, for example, could be identified as wiping of the board, and a marker could be set to provide navigational hints for the replay.

Another problem concerns the scalability of the board data. Most standard scaling methods are easily able to scale down images from one screen resolution to another (for example from $1024 \times 768$ pixels to $640 \times 480$ pixels). The resolution of electronic chalkboards, however, is becoming higher and higher, as it is desirable to have a writing resolution that comes very close to that of a real chalkboard. Handheld devices on the other hand are naturally constrained in their display size and resolution. Replaying electronic chalkboard lectures on handheld devices, such as mobile phones or PDAs, therefore requires more and more new scaling strategies. One possibility for proper replay of electronic

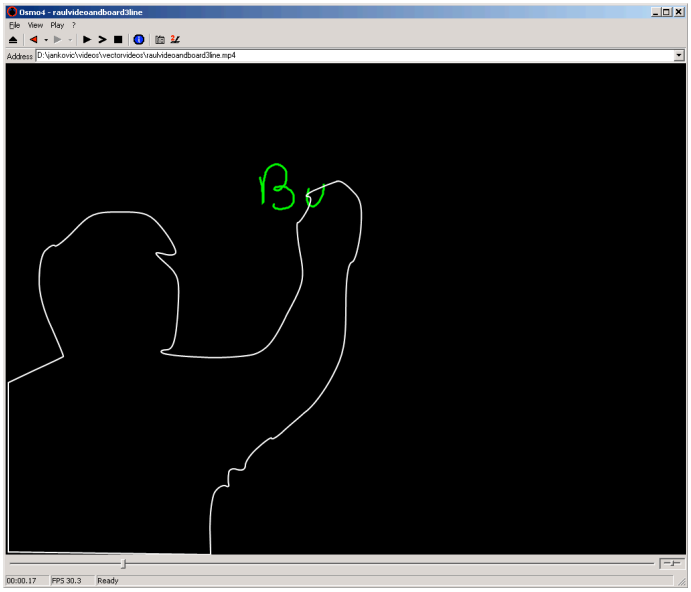


**Figure 12.1:** In order to allow transmission of the overlaid instructor even for low bandwidth connections, the instructor could be transmitted as an outline only. The image shows a preliminary test implementation of the idea in MPEG-4.

chalkboard content on small devices would be to show only a smaller region of interest. In order to determine what region is currently of interest, both the board data and the image of the overlaid instructor could be used. For example, a region of interest could be defined around newly appearing board content or around the instructor's acting hand.

Concerning audio recording, the goal should be to get speech recording quality enhancement without the constraint of having to use a directed microphone and without the need for creating an audio profile. Speech recognition and speaker segmentation methods, as well as physiological models of the human auditory system, could be included to simulate the work of an audio technician during a recording session. The system should be able to decide when to apply a special filter technique out of a set of several possibilities. The system should be capable of handling multiple microphones and inputs to enable switching between classroom questions and the lecturer's voice. One would also like to interface with external studio hardware, such as mixer desks, to enable automatic operation. The system should be able to enhance the quality as far as possible given a certain sound equipment.

## 12.3 Final Note

The word "multimedia" triggers different associations, depending on whose ears are actually perceiving it. A shop assistant in an electronics supply will surely point you to the department where television sets, all kinds of audio receivers and players, and lately also digital photo cameras are sold. Artists and designers often use this word to refer to activities connected to digital content creation, such as building web sites. When school teachers use "multimedia", they usually

mean that their classroom instruction is supported by some kind of audio-visual content.

Computer scientists primarily see the huge amount of information that has to be handled in contrast to regular texts files or binary programs. As a result, most of the research is focused on building hardware and creating algorithms that are able to analyze, process, and retrieve the data masses. With the exception of digital photography, content creation is seldom discussed in scientific articles because multimedia content is mostly assumed to be already given. Moreover, multimedia is hardly ever discussed in the sense of "integrated media" unless, perhaps, in papers on pedagogy or psychology. Mostly, multimedia simply refers to the combination and simultaneous use of different media.

In my opinion, multimedia as a research field must investigate the creation of methods to generate new combinations of different sensory input and output using appropriate devices while optimizing the human perceptibility and the interaction styles. The term "multi" implies more than the addition of an audio track to a sequence of images, or the combination of digitized pictures with a set of text paragraphs to form an electronic book. My personal definition of multimedia is holistic: Content that uses different "media" (types of content) can easily be created, edited, and played back in an integrated fashion, so that the resulting combination forms more than the sum of its parts. I have focused my work in this dissertation bearing in mind this understanding of the double plural "multimedia" and researched solutions that promote this idea.

It is my hope that the work described in this dissertation will be seen as an example and have an impact on the future development of multimedia content creation systems.