Chapter 1

Introduction

This thesis describes my research on improving the production of distance lectures directly in the classroom. The work is conducted with the concrete example of E-Chalk, a software system for recording and transmitting electronic whiteboard lectures over the Internet. Several tasks that formerly required technical personnel are now performed by the computer. A novel combination of video and board content is proposed that solves well-known ergonomic problems and achieves a higher degree of media integration. Although developed inside the scope of a concrete system, the results are general and the realized methods may be used in a variety of applications. In essence, this dissertation shows that computers can be utilized much better to automate the creation of multimedia content than the current state of the art by making the following contributions:

- The architectural base for E-Chalk is provided by a novel multimedia software framework. Based on state-of-the-art solutions for component-based software development, the system simplifies the implementation and the configuration of multimedia processing applications. It facilitates the integration of diverse multimedia content already on the architecture level.

- An Internet audio broadcasting system is presented that makes studioless voice recording easier by automating several tasks usually handled by technicians. The solution described in this document measures the quality of the sound hardware used, monitors possible hardware malfunctions, prevents common user mistakes, and provides online sound-enhancement mechanisms.

- Using a novel segmentation algorithm, the lecturer image is extracted from a video and then pasted on top of the vector-graphics image of the board. This allows to transmit the facial expressions and gestures of the instructor in direct correspondence to vector-based handwritten board content. The resulting lecture allows to watch the instructor acting in front of the board without occluding any board content.

- Finally, the instructor video segmentation approach is generalized. The method can be used for interactive object extraction in generic image or video processing software. This thesis presents benchmark results and shows how the approach has been implemented in several popular open-source applications.

1.1 Motivation

Not even three decades ago, texts were regularly produced with mechanical typewriters. While it was difficult at best to write a typo-free business letter, the creation of a book or a professional article from scratch involved numerous steps that were usually handled in collaboration with several professionals, such as the "markup persons". Today, word processors take over almost all of these steps automatically. It is possible to create text documents from scratch without being a typography expert or having to relying on the help of one. Images, tables, or mathematical formulas can be easily integrated into the document. Spell and grammar checkers even automatically underline suspicious words while typing. What took several people in the past is one step today. But that's not everything. A text can contain clickable references to other documents, or interactive forms can be filled out directly on the screen. In other words, the computer even extended the definition of text in several ways, making new types of documents possible.

However, when creating documents that contain sound, video, or other data directly encoded for human sensory perception, the abilities of the computer are more comparable to those of the typewriter. It is very difficult to create a decent audio recording in one step and without professional help. Combining several types of content, for example, video and vector data, is not easily possible for a non-expert and usually involves several steps. Although many computer manufacturers have recently begun to advertise their products with slogans about video and audio processing capabilities, the computer is almost exclusively used for archiving, transmitting, and playing back multimedia content. The integrated creation of content that has become commonplace in the text domain is not yet available for multimedia data. One of the reasons for this is that computers have just started to be able to display, store, and transmit the huge amounts of data required for encoding information that can directly be perceived by the human senses, and the full potential of computer-based multimedia content creation has yet to be explored.

When the automated creation of audiovisual content is a pressing need, as in the live broadcasting and recording of university lectures, most technicians and engineers do not only rely on traditional video and audio hardware, the work flows are also still similar to the work flows established decades ago. Lectures are usually just captured with a camera, and the use of the computer, for example for digital storage and Internet broadcasting, is in most cases just an additional step after manual recording, digitization, and cutting. As a result, tremendous effort and money has to be put into the creation of distance lectures, and/or the resulting presentation quality is far from optimal. The main reason for this is that it is not yet well understood how multimedia content creation can be automated.

Using the example of E-Chalk, this thesis provides evidence that computers are able to facilitate the creation of multimedia content substantially and that we have not reached the final step in recording automation yet. Furthermore, this dissertation shows how computers can provide new ways of presenting multimedia content without manual editing, which go well beyond the capabilities of mere audio and video replay.

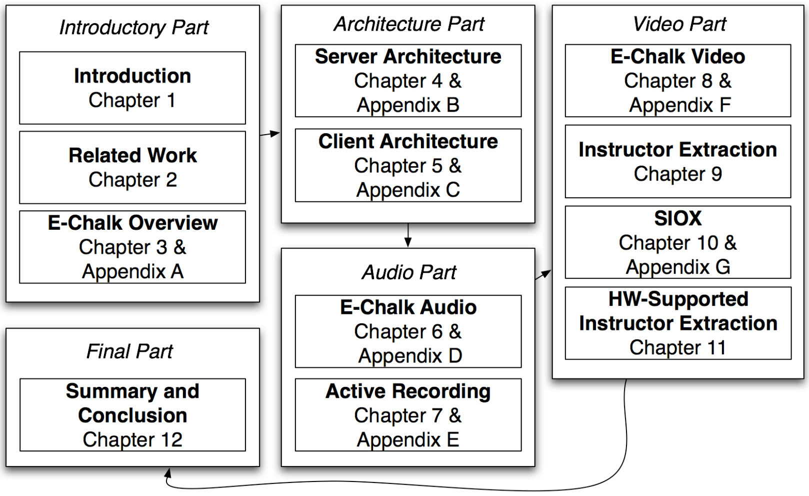

Figure 1.1: Conceptual overview of the structure of this dissertation.

1.2 Overview of this Document

This dissertation is structured as shown in Figure 1.1. Chapter 2 starts with the review of related work on the automated creation of distance-education courses from the classroom. Then the philosophy of the E-Chalk system as well as an overview of the entire project is presented in Chapter 3. Chapter 4 introduces software-architectural considerations for the design of multimedia systems for classroom teaching and introduces the component-based framework that is the base of the E-Chalk server system. Technical considerations for the transmission and playback of E-Chalk lectures on the client side are presented in Chapter 5. Based on the observations presented in the previous chapters, Chapter 6 describes audio capturing, transmission, and archiving. Chapter 7 then describes the Active Recording approach to facilitate the production of voice recordings by simulating several tasks usually handled by audio technicians. Video capturing, transmission, and archiving is described in Chapter 8. Chapter 9 continues with the extraction of the lecturer image out of a video in order to paste it on top of the vector-graphics image of the board. SIOX, a generalization of the method used for the instructor video segmentation is presented in Chapter 10. Chapter 11 finally suggests how SIOX might be combined with range sensors to provide another solution for the extraction of the lecturer image from a video. In the end, Chapter 12 concludes this thesis with a summary and a brief presentation of future work. Appendix A provides an overview of the components of the E-Chalk system along with all contributors. Appendices B to G provide further technical details on the component framework and E-Chalk's audio and video subsystems, as well as the details of the benchmark described in Chapter 10.