# Automated Hypermedia Authoring for Individualized Learning

**Miloš Liška**[12] and **Vít Rusňák**[1] and **Eva Hladká**[12]

[1]Faculty of Informatics, Masaryk University, Botanická 68a, Brno, Czech Republic
[2]Cesnet z. s. p. o., Zikova 4, Prague, Czech Republic
E-mail: {xliska, xrusnak, eva}@fi.muni.cz

## ABSTRACT

This paper presents an automated system for authoring of hypermedia lecture recordings. We revised our previous work on lectures indexation. Our original method of lectures indexation needed a human interaction and was not feasible for large amount of lectures. We developed a system which utilizes Optical Character Recognition (OCR) to automatically extract the keywords from the lecture slides which are captured together with the lecture video. The keywords together with timestamps of the slides are than used to search through the contents of the lecture and to navigate to its particular sequences.

## I. INTRODUCTION

Lectures recordings proved to be a valuable learning material for students as well as for the teachers over last five years [1]. With increasing number of recorded lectures the automation of the lecture acquisition, processing and publication becomes an essential requirement. An automated system for lectures recording was presented in [1] and is used extensively for lectures video acquisition and processing. However, a raw video does not provide a sufficient learning material. In this paper, we will present an extension of our lectures recording system towards interactive hypermedia authoring.

Nowadays the lecture is usually based on the use of modern computer presentation and projection facilities (PowerPoint and PDF slides or even interactive whiteboard). In addition to the teacher the presentation introduces valuable source of information which is necessary to acquire and embed into the lecture recording.

Each presented lecture recording is about 100 minutes long and there are up to 13 lectures in one particular course. This means that it is almost impossible to search for some concrete subject within the recordings. Solution of this problem is to add textual metadata information (indices) to each lecture recording and a time information pointing to the lecture video. Two years ago we designed and implemented tool for individual indexation [1, 11]. This tool allows each user to make his/her own timestamps and use them in video like comments in a book. But the tool did not solve the problem searching the video sequence corresponding with key word, concrete slide, etc. Moreover manual indexation of the recorded lectures as proposed in [1, 11] is a lengthy, soul-destroying and with large number of recorded lectures completely impossible task.

The paper is further structured as follows: Section 2 briefly presents our current system used for lectures recordings acquisition. In section 3, we discuss processing of the slides to automatically obtain the keywords for the index. Presentation of indexed lecture recordings is then described in section 4. A brief discussion of related work by other parties is given in section 5. Section 6 concludes the paper and gives some ideas on future work and possible enhancements of presented system.

## II. LECTURES ACQUISITION

In the paper we propose automated acquisition, processing and embedding of the presentations into the recorded lecture and especially utilisation of the presentation for automated indexation of the lecture. The presentation is captured using VGA framegrabber or on-screen video converter and converted to a set of static images. Capturing the VGA introduces one of the most important source of information in the lecture recording especially when the lecturer uses devices like Mimio [2] to capture and convert the whiteboard on screen of the presentation computer or the Smart Sympodium pen display [3] to control and draw into the presentation.

Lectures recording, processing and presentation concept described in this paper is inspired by SonicFoundry's commercial lectures authoring product MediaSite [4]. MediaSite consists of a MediaSite Recorder used for lecture recordings acquisition and MediaSite Rich Media Server used for presentation of the lectures.

In our current system we use the ML440 mobile version of MediaSite recorder. The recorder provides a very straightforward way to record a complete lecture. Video is captured either from IEEE 1394 (Firewire) port for digital audio/video or from an Osprey capture card for analog video over S-Video or composite. Osprey card provides inputs for unbalanced audio (line-in) and balanced audio (XLR) as well.

The most important feature of the MediaSite recorder is the VGA frame grabber which provides a VGA input for capturing the presentation (and more generally content of the screen) from the presentation computer. The VGA frame grabber is capable of capturing up to 1600×1200×24-bit SXGA signal. The VGA signal is captured as a static JPEG compressed

image on each change. Processing of VGA signal and the JPEG images means that the ML440 recorder is capable of capturing only about 5 frames per second and thus is the VGA input unsuitable for capturing of any video which might be a part of the lecture presentation. On the other hand lecture presentations are mostly static.

A timestamp relative to the beginning of lecture capturing is stored with each JPEG image (slide) captured from the VGA input and added to a XML description of the lecture recording. These timestamps are then used to synchronize the slides with the lecture video.

```
<SlideSet number="1">
  <PresentationTime>639</PresentationTime>
  <Normal>
    <FileName>Slide_0001.jpg</FileName>
    <Dimensions>
      <Width>500</Width>
      <Height>375</Height>
    </Dimensions>
  </Normal>
  <FullSize>
    <FileName>Slide_0001_Full.jpg</FileName>
    <Dimensions>
      <Width>1024</Width>
      <Height>768</Height>
    </Dimensions>
  </FullSize>
  <ThumbNail>
    <FileName>Slide_0001_Thumb.jpg</FileName>
    <Dimensions>
      <Width>240</Width>
      <Height>180</Height>
    </Dimensions>
  </ThumbNail>
</SlideSet>
```

*Fig. 1: MediaSite XML lecture description stub.*

## III. AUTOMATIC INDEX EXTRACTION

Since the image quality of captured presentation slides is excellent and suitable source for further processing [5], we adopted and utilised Optical Character Recognition (OCR) to extract the indices from the presentation and create references to the lecture video automatically. The quality of presentation slides captured using VGA frame grabber is outstanding in comparison to the scanned ones, however some preprocessing of the slides helps the OCR to recognize the text on them more efficiently.

We use following steps to optimize the acquired slides for the OCR:

1. A greyscale image is more suitable for further automated processing. The slides in JPEG format are converted to greyscale images using the djpeg tool and stored in PNM format.
2. The unpaper [6] tool is used to increase the readability of the images for further OCR processing. unpaper tries to clean the images by removing or reducing of grey areas which were usually a background images in the original presentation. Moreover it proved helpful to reduce the level of noise in the images using the unpaper tool.

For our purposes the OCR must support national and even special (mathematical) character sets and it must be possible to train it to recognize variety of different styled presentations. Moreover we require an OCR tool which can be controlled using a command line interface and can be used for batch processing of the whole lecture.

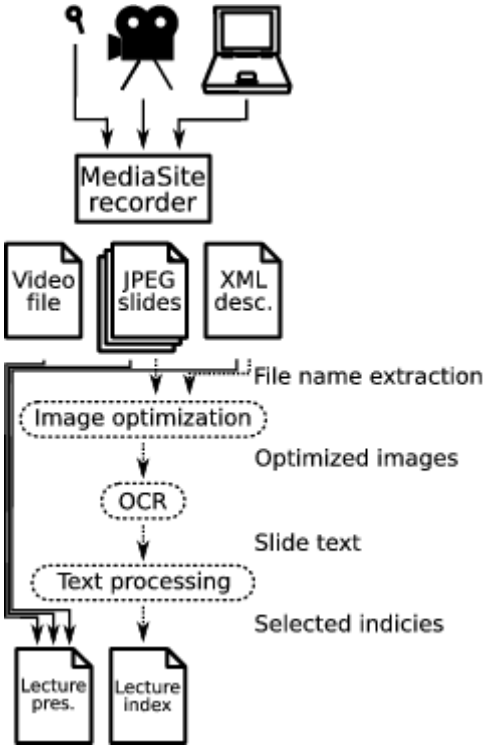


*Fig. 2: MediaSite recorder output processing.*

Among the available OCR tools we chose Jörg Schulenburg's GOCR [7] optical character recognition tool. Just a very limited support for national character sets (Czech language uses a lot of diacritical marks) recognition is included in the GOCR tool. However GOCR is able to learn on supplied texts and new versions are actively developed.

The fruitfulness of Czech characters recognition is extremely low with untrained GOCR. On the other hand it showed up that for most presentations it is sufficient for the GOCR to learn on basic text using sans respectively serif font with Czech national characters. GOCR then creates an internal database with mapping of recognized character shape image and the learned character. Unfortunately the learning process is quite lengthy and requires a human to tell the GOCR what are the recognized characters.

The OCR process extracts a large amount of text from the lecture presentation images. First of all it is necessary to reduce the OCR output by omitting connectives, prepositions and the common and frequented vocabulary. This is done by comparing the

words from extracted text with a dictionary using the regular expressions and deleting all occurrences of the word if a match in the dictionary was found. The remaining words in the scanned text are then treated as the keywords and create a basis for the lecture index.
A timestamp corresponding to the slide is added to all key-words remaining in the text recognized on that particular slide. The key-words together with the timestamps are than stored in a XML index file. The index file is eventually transformed to a static HTML which is used for the lecture recording presentation.
We have done preliminary tests and measurements on indices extraction efficiency. Experiments on several types of presentations with about 40 slides showed that presentations with extensive graphics elements take considerably more time to process than textual slides. A basic OCR on the presentation slides takes the same amount of time for slides in English and in Czech. The difference is that gocr is able to produce 99% correct output on English slides without using learned characters from the internal database. Usage of the internal database is a must for achieving a reasonably correct output with slides in Czech language which considerably lengthens the processing time.
We used a single 3GHz Pentium 4 CPU with 1024MB RAM and fast SATA hard disk node for our preliminary measurements of the slides processing performance. The results are summarized in the following table:

*Tab. 1: Presentation processing performance*

| | **text prevailing** | **graphics prevailing** |
|---|---|---|
| Slides conversion | 5s | 15s |
| Slides optimization | 80s | 132s |
| OCR (no database used) | 47s | 93s |
| OCR (database used) | 192s | 500s |
| Total (no database used) | 132s | 240s |
| Total (database used) | 273s | 647s |

## IV. PRESENTATION OF RECORDED LECTURES

SonicFoundry's MediaSite Rich Media Server is a closed source framework for presentation of recordings acquired using the MediaSite Recorder, providing a database to store multimedia data for the lecture and a XML metadata file linking captured slides to a particular timestamps. When a user requests to watch a particular lecture, Rich Media Server generates a dynamic web page based player.
According to our opinion the capabilities and exploitation of the potential of lecture data recorded by the MediaSite Recorder is quite inadequate. Moreover the server is exclusively oriented on using Microsoft technologies which in some cases prevents especially the Linux (and Mozilla or Mozilla Firefox) users to fully access the web based presentations player.
That is why a platform independent web based lecture recordings player was created supporting video, embedded presentations and indices.
Current version of our player provides a support for embedded lecture video, time synchronization of video and slides, several modes for slides displaying in various resolutions and a bar with slides thumbnails for navigation through recorded lecture.
Web pages with the recording player are generated from the recording metadata XML file using a set of Perl scripts. The original embedding of the video into the web page using a <object> tags was replaced with less powerful but platform and browser independent <embed> tag. Interaction with the player and synchronization of the player with the slides is implemented using purely a JavaScript language which again helps to keep the player browser and operating system independent.

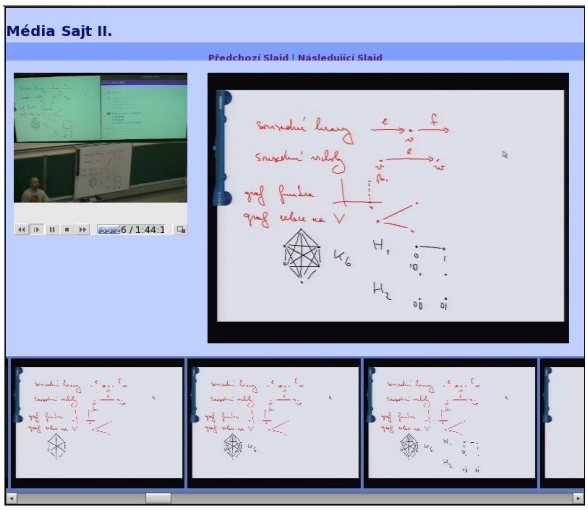


*Fig. 3: Lecture presentation displayed using Firefox and mplayerplug-in on Linux OS.*

We suppose the user to have either Windows Media Player (WMP) plugin or mplayerplug-in [8] installed and available for the web browser depending on the operating system. Both plugins allow playing of video embedded to the web pages and both plugins provide a JavaScript API so that they can be controlled thorough the web browser. Unfortunately the API is different especially in video position handling functions for mplayerplug-in and WMP plugin. Moreover there are differences in the JavaScript API even for various versions of WMP plugin.
To handle this, we generate different static HTML code for a range of supported web browser, embedded video player plugin and operating system combinations.

Lecture index is presented as standalone static HTML page linked from the main player page. The index provides an alphabetically sorted list of keywords extracted from the slides. Each keyword is followed with a link to a JavaScript code which starts the video playback at the time given by the timestamp of the slide where the keyword occurred.
End user is than able to search through the indices, find the concrete subject and directly replay the respective part of a lecture.

## V. RELATED WORK

There are several other projects aimed on presentation processing based lectures indexing.
A Faculty of Information Technology at Brno University of Technology project [9] is aimed on automatic detection and extraction of information from lecture video recorded using a camera which takes both lecturer and the presentation projected on the screen. This project developed techniques to extract the presentation slides from lecture video for further processing. However we believe that the quality of presentation slides captured using a VGA frame grabber is greater and the overall processing of the slides simpler that in before mentioned process. On the other hand an advantage of their method is that it has to deal with just one video source.
A similar project [10] is aimed on analysing lecture recordings capturing a lecturer and presentations displayed using an overhead projector.
Another approach to automatic lectures indexing is based on lecturers speech recognition and processing [12, 13, 14]. The advantage is that it is possible to extract the indices directly from the recorded lecture audio. However, this approach provides much lower accuracy [13, 14] then OCR processing of the lecture presentations.

## VI. CONCLUSIONS AND FUTURE WORK

Proposed automated indexation system is well scaling on large presentations and is a performance modest add-on to our previous system for automated lectures video acquisition and processing.
The success of the indexation depends on two tricky steps in the process of creating the index from the capture slides. The easier one to solve is the GOCR learning which needs a human time. However, it is necessary to undertake the GOCR learning just once for the future lectures recordings. A time save is obvious in comparison with manual indexation of the lectures. More complicated problem introduces the reduction of the text recognized from the presentation to extract just a relevant key-words for the index.
An advantage of static HTML pages generated just once per lecture recording and the fact that the player control is a client side JavaScript is that the whole presentation doesn't impose additional load on the web server. Thus the presentation of indexed lectures scales equally to just serving plain video files to the end users. Moreover, because of that the web based lectures recording player consists only of media data and static web pages, users can download the whole presentation easily as they are used with plain video files.

## ACKNOWLEDGMENTS

Authors of this paper would like to thank to Pavel Šiler for testing of the lectures presentation interface and to Lukáš Hejtmánek for bug fixing.
This project has been kindly supported by the research intent “Optical Network of National Research and its New Application” (MŠM 0021622419).

## REFERENCES

[1] Hladká, E., Liška, M., Matyska, L.: Multimedia Support for Individualized Learning. In Proceedings of the Fifth International Conference on Information Technology Based Higher Education and Training. Istanbul, Turkey: IEEE Catalog Number 04EX898, p. 4-9, 2004.
[2] http://www.mimio.com/
[3] http://www2.smarttech.com/st/en-US/Products/Interactive+Pen+Displays/
[4] http://www.sonicfoundry.com/mediasite/
[5] Daddaoua, N., Odobez, J. M., Vinciarelli, A.: OCR Based Slide Retrieval. In ICDAR '05: Proceedings of the Eighth International Conference on Document Analysis and Recognition, Washington, USA: IEEE Computer Society, p. 945-949, 2005.
[6] http://unpaper.berlios.de/
[7] http://jocr.sourceforge.net/
[8] http://mplayerplug-in.sourceforge.net/
[9] Sumec, S.: Extracting Additional Information from Lecture Recordings. Cesnet z. s. p. o. technical report 11/2006.
[10] Wallick, M., da Vitoria Lobo, N., Shah, M.: Computer Vision Framework for Analyzing Computer and Overhead Projections from Within Video. International Journal of Computers and Their Applications, vol. 8, no. 2., 2001.
[11] Šmíd, P. Jednoduchá indexace videostreamů pro potřeby záznamů přednášek. Bachelaureate thesis (in Czech only). FI MU Brno, Czech Republic, 2004.
[12] Glembek, O., Automatic Lecture Indexing Using Voice Recognition. Msc. thesis. FIT VUT, Brno, Czech Republic 2005.
[13] Hürst, W., “A qualitative study towards using large vocabulary automatic speech recognition to index recorded presentations for search and access over the web”, IADIS International Journal on WWW/Internet, Volume I, Number 1, 2003.
[14] Repp, S., Meinel, Ch., Semantic Indexing for Recorded Educational Lecture Videos. In Proceedings of the 4th annual IEEE international conference on Pervasive Computing and Communications Workshops. 2006.