ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

# EFFICIENT CONTENT BASED VIDEO RETRIEVAL USING ASR AND OCR TECHNIQUES

**Anil Shravan Gavhane**

[*] Electronics and Telecommunication, Late G.N.Sapkal COE,Anjeneri,Nashik, India

## ABSTRACT

E-lecturing has become more and more popular, in the last decade. On the World Wide Web (WWW) is growing rapidly, the amount of lecture video data. For video retrieval in WWW or within large lecture video archives is urgently needed, a more efficient method. For automated video indexing and video search in large lecture video archives, this paper presents an approach. To offer a visual guideline for the video content navigation, we apply automatic video segmentation and key-frame detection. By applying video Optical Character Recognition (OCR) technology on key-frames and Automatic Speech Recognition (ASR) on lecture audio tracks, we extract textual metadata. For keyword extraction, by which both video- and segment-level keywords are extracted for content-based video browsing and search, the OCR and ASR transcript as well as detected slide text line types are adopted. By evaluation, the performance and the effectiveness of proposed indexing functionalities are proven.

**KEYWORDS**: Lecture videos, automatic video indexing, content-based video search, lecture video archives.

## INTRODUCTION

Improved video compression techniques and high-speed networks in the last few years, DIGITAL video has become a popular storage and exchange medium due to the rapid development in recording technology. In e-lecturing systems, therefore audiovisual recordings are used more and more frequently. To access independent of time and location, a number of universities and research institutions are taking the opportunity to record their lectures and publish them online for students. On the Web, there has been a huge increase in the amount of multimedia data. To find desired videos without a search function within a video archive, for a user it is nearly impossible. To judge whether a video is useful by only glancing at the title and other global metadata which are often brief and high level, even when the user has found related video data, it is still difficult most of the time for him. The user might thus want to find the piece of information he requires without viewing the complete video, the requested information may be covered in only a few minutes. In a large lecture video archive more efficiently, the problem becomes how to retrieve the appropriate information. Most of the video retrieval and video search systems such as You Tube, Bing and Vimeo reply based on available textual metadata such as title, genre, person, and brief description, etc. But the creation step is rather time and cost consuming; this kind of metadata has to be created by a human to ensure a high quality. High level and subjective, the manually provided metadata is typically brief. By using video analysis technologies, beyond the current approaches, the next generations of video retrieval systems apply automatically generated metadata. Much more content-based metadata can thus be generated which will lead to two research questions in the e-lecturing context:

- The learner in searching required lecture content more efficiently, can those metadata assist.
- The important metadata from lecture videos and provide hints to the user, If so, how can we extract.

In summary, the major contributions of this paper are the following:

- By applying appropriate analysis techniques, we extract metadata from visual as well as audio resources of lecture videos automatically. Which can guide both visually and text-oriented users to navigate within lecture video, for evaluation purposes we developed several automatic indexing functionalities in a large lecture video portal. To verify the research hypothesis and to investigate the usability and the effectiveness of proposed video indexing features, we conducted a user study intended.

- For slide video segmentation and apply video OCR to gather text metadata, for visual analysis, we propose a new method. Furthermore, lecture outline is extracted from OCR transcripts by using stroke width and geometric information. search function has been developed based on the structured video text.
- Which fills the gap in open-source ASR domain; we propose a solution for automatic German phonetic dictionary generation. The dictionary software and compiled speech corpus are provided for the further research use.
- We propose a keyword ranking method for multimodal information resources, In order to overcome the solidity and consistency problems of a content-based video search system. We implemented this approach in a large lecture video portal, in order to evaluate the usability.
- By using compiled test data sets as well as opened benchmarks, the developed video analysis methods have been evaluated. All compiled test sets are publicly available from our website for the further research use.

## RELATED WORK

Recording lectures and putting them on the Web for access by students has become a general trend at various universities. To take full gain of the knowledge data base that is built by these documents elaborate search functionality has to be provided that goes beyond search on meta-data level but performs a detailed analysis of the corresponding multimedia documents. In this paper, we present some experiments we did towards setting up a Web-based search engine for audio recordings of presentations. We evaluate standard, state-of-the-art speech recognition software as well as achievable retrieval performance. In addition, we compare the speech retrieval results with a traditional, text-based approach for searching to evaluate the value of speech processing for lecture retrieval.

This paper presents an approach for automated video indexing and video search in large lecture video archives. First of all, we apply video segmentation and key-frame detection to offer a visual guideline for the video content navigation. We further extract textual metadata by applying video Optical Character Recognition (OCR) on detected key frames and by performing Automatic Speech Recognition (ASR) on lecture audio streams, automatically. The OCR and ASR transcript as well as the detected OCR text line types are adopted in the subsequent keyword extraction process, by which both video- and segment-level keywords are extracted respectively. A user study is provided for evaluating the performance and the effectiveness of proposed indexing methods in our lecture video portal. Furthermore, we propose a novel concept for content-based video search by using multimodal information resources.

We investigate methods of segmenting, visualizing, and indexing presentation videos by separately considering audio and visual data. The audio track is segmented by speaker, and augmented with key phrases which are extracted using an Automatic Speech Recognizer (ASR). The video track is segmented by visual dissimilarities and augmented by representative key frames. An interactive user interface combines a visual representation of audio, video, text, and key frames, and allows the user to navigate a presentation video. We also explore clustering and labeling of speaker data and present preliminary results. Transcribing lectures is a challenging task, both in acoustic and in language modeling. In this work, we present our first results on the automatic transcription of lectures from the TED corpus, recently released by ELRA and LDC. In particular, we concentrated our effort on language modeling. Baseline acoustic and language models were developed using respectively 8 hours of TED transcripts and various types of texts: conference proceedings, lecture transcripts, and conversational speech transcripts. Then, adaptation of the language model to single speakers was investigated by exploiting different kinds of information: automatic transcripts of the talk, the title of the talk, the abstract and, finally, the paper. In the last case, a 39.2% WER was achieved. Creating video recordings of events such as lectures or meetings is increasingly inexpensive and easy. However, reviewing the content of such video may be time-consuming and difficult. Our goal is to produce a “comic book” summary, in which a transcript is augmented with key frames that disambiguate and clarify accompanying text. Unlike most previous keyframe extraction systems which rely primarily on visual cues, we present a linguistically-motivated approach that selects key frames that contain salient gestures. Rather than learning gesture salience directly, it is estimated by measuring the contribution of gesture to understanding other discourse phenomena. More specifically, we bootstrap from multimodal coreference resolution to identify gestures that improve performance. We then select key frames that capture these gestures. Our model predicts gesture salience as a hidden variable in a conditional framework, with observable features from both the visual and textual modalities. This approach significantly outperforms competitive baselines that do not use gesture information.

A lot of content for tele-teaching portals was produced in the last decade. But metadata to filter and search through the content was not generated adequately. That is why it is a new challenge to find solutions how to filter the large amount of data with a small base of metadata available. To engage the user community to generate metadata is one option. Due to the small size of tele-teaching user communities this approach needs to be enhanced with automatic methods. This paper describes community rating and tagging as two widely used social web functionalities and motivates their usage in the tele-teaching context. Furthermore approaches to extend and enhance this user-generated meta- data are explained. The potential of other social web features is explained afterwards. Finally the need to activate more users and the opportunity to connect the social with semantic web features in tele-teaching are motivated.

## CONCLUSION AND FUTURE WORK

In this paper, we presented an approach for content-based lecture video indexing and retrieval in large lecture video archives. In order to verify the research hypothesis we apply visual as well as audio resource of lecture videos for extracting content-based metadata automatically. Several novel indexing features have been developed in a large lecture video portal by using those metadata and a user study has been conducted. As the future work, the usability and utility study for the video search function in our lecture video portal will be conducted. Automated annotation for OCR and ASR results using Linked Open Data resources offers the opportunity to enhance the amount of linked educational resources significantly. Therefore more efficient search and recommendation method could be developed in lecture video archives.

## REFERENCES

1. E. Leeuwis, M. Federico, and M. Cettolo, “Language modeling and transcription of the ted corpus lectures,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., 2003, pp. 232–235.
2. D. Lee and G. G. Lee, “A korean spoken document retrieval system for lecture search,” in Proc. ACM Special Interest Group Inf. Retrieval Searching Spontaneous Conversational Speech Workshop, 2008.
3. J. Glass, T. J. Hazen, L. Hetherington, and C. Wang, “Analysis and processing of lecture audio data: Preliminary investigations,” in Proc. HLT-NAACL Workshop Interdisciplinary Approaches Speech Indexing Retrieval, 2004, pp. 9–12.
4. A. Haubold and J. R. Kender, “Augmented segmentation and visualization for presentation videos,” in Proc. 13th Annu. ACM Int. Conf. Multimedia, 2005, pp. 51–60.
5. W. H€urst, T. Kreuzer, and M. Wiesenh€utter, “A qualitative study towards using large vocabulary automatic speech recognition to index recorded presentations for search and access over the web,” in Proc. IADIS Int. Conf. WWW/Internet, 2002, pp. 135–143.
6. C. Munteanu, G. Penn, R. Baecker, and Y. C. Zhang, “Automatic speech recognition for webcasts: How good is good enough and what to do when it isn’t,” in Proc. 8th Int. Conf. Multimodal Interfaces, 2006.

## AUTHOR BIBLIOGRAPHY

**Athor Anil S. Gavhane**
I complited my BE from pune university in electronics and telecommunication.Now student of master in signal processing in same university.