Text Retrieval Video Frames For Content-Based Search

[1]D.S.Thosar, [2]Khedkar Kapil, [3]Dharam Ashwini, [4]Shaikh Najir

[1]Professor Department of comp. , SVIT CHINCHOLI, Nashik, India

[2]B.E, Department of comp. , SVIT CHINCHOLI, Nashik, India

[3]B.E, Department of comp. , SVIT CHINCHOLI, Nashik, India

[4]B.E, Department of comp. , SVIT CHINCHOLI, Nashik, India

Abstract - ***In the last decade e-lecturing has become most popular. The amount of lecture video data on the World Wide Web (WWW) is viable briskly. Therefore, a more effictive method for video retrieval in WWW or within large lecture video archives is urgently needed. This paper presents an approach for automated video indexing and video search in huge lecture video archives. Firstly, we apply automatic video segmentation and key-frame detection to offer a visual guideline for the video content navigation. latter, we extract textual metadata by applying video Optical Character Recognition (OCR) technology on key-frames and Automatic Speech Recognition (ASR) on lecture audio tracks. The OCR and ASR transcript as well as detected slide text line types are accept for keyword extraction, by which both video- and segment-level keywords are extracted for content-based video browsing and search. The performance and the use of proposed indexing functionalities is proven by decision.***

Key Words: *automatic video indexing, content-based video search, lecture video archives*

1. INTRODUCTION

The final segmentation result is strongly dependent on the quality of the OCR result. It might be less efficient and imply redundancies, when only poor OCR result is obtained. Tuna et al. presented their approach for lecture video indexing and search. They segment lecture videos into key frames by using global frame differencing metrics. Then standard OCR software is applied for gathering textual metadata from slide streams, in which they utilize some image transformation techniques to improve the OCR result. They developed a new video player, in which the indexing, search and captioning processes are integrated. Similar to, the used global differencing metrics cannot give a sufficient segmentation result when animations or content build-ups are used in the slides. In that case, many redundant segments will be created. Moreover, the used image transformations might be still not efficient enough for recognizing frames with complex content and background distributions. Making use of text detection and segmentation procedures could achieve much better results rather than applying image transformations. Jeong et al. proposed a lecture video segmentation method using Scale Invariant Feature Transform (SIFT) feature and the adaptive threshold in. In their work SIFT feature is applied to measure slides with similar content.

An adaptive threshold selection algorithm is used to detect slide transitions. In their evaluation, this approach achieved promising results for processing one-scene lecture videos. Recently, collaborative tagging has become a popular functionality in lecture video portals.

Sack and Waitelonis and Moritz et al. apply tagging data for lecture video retrieval and video search. Beyond the keyword based tagging, Yu et al. proposed an approach to annotate lecture video resources by using Linked Data. Their framework enables users to semantically annotate videos using vocabularies defined in the Linked Data cloud. Then those semantically linked educational resources are further adopted in the video browsing and video recommendation procedures. However, the effort and cost needed by the user annotation-based approach cannot satisfy the requirements for processing large amounts of web video data with a rapid increasing speed. Here, the automatic analysis is no doubt much more suitable. Nevertheless, using Linked

Data to further automatically annotate the extracted textual metadata opens a future research direction. ASR provides speech-to-text information on spoken languages, which is thus well suited for content-based lecture video retrieval. The studies described in and are based on out-of-the-box commercial speech recognition software. Concerning such commercial software, to achieve satisfying results for a special working domain an adaption process is often required, but the custom extension is rarely possible. The authors of and focus on English speech recognition for Technology Entertainment and Design (TED) lecture videos and webcasts. In their system, the training dictionary is created manually, which is thus hard to be extended or optimized periodically. Glass et al. proposed a solution for improving ASR results of English lectures by collecting new speech data from the rough lecture audio data.

1.1 Existing System

Several content-based video search engines have been proposed no long ago. Adcock et al. suggested a lecture webcast search system , in which they applied a slide frame segmental to extract lecture slide images. The system retrieved more than 37.00 lecture videos from different resources such as YouTube, Berkeley Webcast, The search marks are created based on the global metadata obtained from the video hosting website and texts extracted from slide videos

by using a standard OCR engine. Since they do not assign text detection and text segmentation process, the OCR recognition accuracy of their approach is therefore lower than our system's.

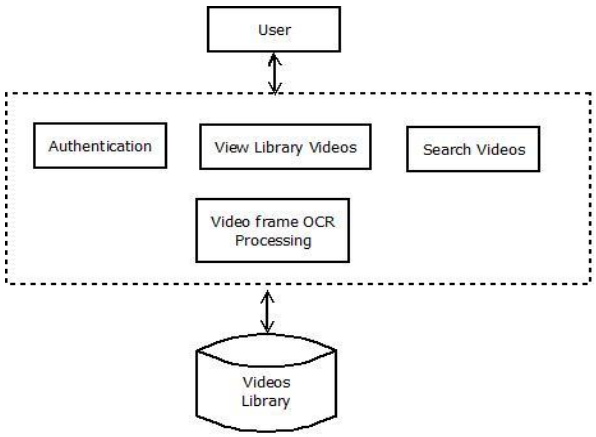
moreever, by applying the text detection process we are able to extract the structured text line such as title, subtitle, key-point, etc., that enables a more soft search function. In the CONTENTUS project, a content based semantic multimedia retrieval system has been developed. After the digitization of media data, a lot analysis techniques, e.g., OCR, ASR, video segmentation, automatic speaker recognition, etc., have been applied for metaditate generation. An entity detection algorithm and an open knowledge base are used to extract entities from the textual metadata. As mentioned before, searching through the recognition results as a degree of confidence, we have to deal with the solidity and the consistency problem.But the reviewed content-based video search systems did not consider this issue.

Most of the lecture speech recognition systems have low recognition rate, the WERs of audio lectures are approximately 40-85 percent. The bad of recognition results limit the further indexing efficiency. Therefore, how to continuously better ASR accuracy for lecture videos is still an unsolved problem.

1.2 Proposed System

In Proposed System user will be given option to register and login into the system based on the specified credentials. With the help of Add Video option, user will be specified to add a local video to the application library of the required format with textual data. The video frames of selected video (of small size with textual data) will be processed using OCR to identify keywords associated to the video and weights of keywords will be analyzed to store it along with the video.

The video frames of selected video (of small size with textual data) will be processed using OCR to identify keywords associated to the video and weights of keywords will be analyzed to store it along with the video. The added videos into the system will be displayed to the user here. The user can download them to his machine if required. The user will enter keyword to be searched from keyword set identified by OCR processing for each video. Videos having keyword set similar to search keywords will be displayed to the user for downloading.



2. SYSTEM ARCHITECTURE

In this System user will be given option to register and login into the system based on the specified credentials. With the help of Add Video option, user will be specified to add a local video to the application library of the required format with textual data. The video frames of selected video (of small size with textual data) will be processed using OCR to identify keywords associated to the video and weights of keywords will be analyzed to store it along with the video.

The video frames of selected video (of small size with textual data) will be processed using OCR to identify keywords associated to the video and weights of keywords will be analyzed to store it along with the video. The added videos into the system will be displayed to the user here. The user can download them to his machine if required. The user will enter keyword to be searched from keyword set identified by OCR processing for each video. Videos having keyword set similar to search keywords will be displayed to the user for downloading.

3.CONCLUSIONS

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.