# Survey of Content Based Lecture Video Retrieval

Dipali Patil[#1], Mrs. M. A. Potey[*2]

[#] *Department of Computer Engg. Savitribai Phule Pune University*
*DYPCOE, Akurdi, India*

***Abstract:*** **In the last few years, e-lecturing has become more and more popular because video provide rich source of information. The amount of lecture video data on the internet is growing exponentially. Thus, a more efficient method for video retrieval in internet or within large lecture video archives is urgently needed. This paper presents a text based video retrieval and Video search system using Optimal Character Recognition (OCR). First, we convert the video into key-frames and extract the Text using OCR. Following step is to produce a summary presenting key points of the video, by making use of meradata of text and audio extracted from the Video. This summary will then be used for grouping and Indexing of videos. In this paper, we discuss various lecture video segmentation approaches. As there is strong needs for segmenting lecture videos into topic units in order to organize the videos for browsing and to provide search capability.**

***Keywords*** **- Content based Video Retrieval, Lecture Video Segmentation and Optical Character Recognition (OCR)**

## I. INTRODUCTION

Many universities offer distance learning by recording classroom lectures, maintaining hundreds of lecture video recordings in a repository, and making them available to remote students over the Internet [1]. Digital video has become a popular storage and exchange medium due to rapid development in recording technology. Videotaping of lecture video is more common e-Learning. A number of universities and research institutes are taking the opportunity to record their lectures and publish them online for students. As a result, there has been a enormous increase in the amount of multimedia data on the Web. Hence, for a user it is nearly impossible to find desired videos without a search function within a video archive.

Content-based video retrieval have a wide range of applications such as quick browsing of video folders, analysis of visual electronic commerce (such as analysis of interest trends of users' selections and orderings, analysis of correlations between advertisements and their effects), remote instruction, digital museums, news event analysis, intelligent management of web videos (useful video search and harmful video tracing), and video surveillance.

As the amount of video data is generated exponentially, it is often burdensome for students to search through a full video or across many videos, in order to find specific portions of their interest. For example, one repository (NPTEL [2]) has a course on Networking, containing around 60 recorded lecture videos, each of nearly 90 minutes duration. If a user wants to find the portion where TCP protocol is discussed, the user has to manually go through titles of all the lectures in the course and first decide which videos might contain the preferred explanation. Then the user has to individually browse through each of the chosen videos to find the portion where TCP protocol is discussed. The problem is exacerbated if the user is not familiar with the area, or if the topic is very specific or particular, as the user may not be able to decide the videos to be scrutinized. A large amount of textual metadata will be created by using Optical Character Recognition (OCR) und Automatic Speech Recognition (ASR) method, which provides the content of lecture videos. For content-based video retrieval and search, the search indices are created from different information resources, including manual annotations, observations, comments, OCR and ASR keywords, metadata, etc.

The rest of the paper is structured as follows. Section II describes work related to video retrieval. Section III gives lecture video segmentation approaches that can be used for same purpose. Later in the section, we have reviewed some text retrieval techniques like OCR and strategies for indexing and retrieval. In section VI, is briefly reviewed conclusion.

## II. LITERATURE SURVEY

Most of the prior work related to Content based video retrieval mainly focus on video segmentation and summarization. To obtain efficient result for video search, video has to be segmented and all text information need to be extracted for indexing and tagging purpose. Techniques for searching in lecture videos.

### *A. Techniques for video retrieval*

Search facilities can be provided in lecture video repositories in two ways [6]. They are:

Meta Data Based*:* Meta data is textual data that is applied to a piece of multimedia content in order to describe it. These methods make the use of Meta data, such as video title, video description, user feedback and comments, to identify video results matching given set of keywords. This kind of Meta data based approach may be able to identify videos that contain the keywords but they cannot locate where those keywords appear in the video time line.

Content-based: Lecture videos typically contain the following contents: (i) Lecturer Speech: Portion of video that shows the instructor talking, (ii) Slides: Portion of video that shows the current slide of the presentation, and (iii) Lecturer Notes: Portion of video that shows the board/paper on which instructor is writing. Content based approaches extract meta-data from appropriate portions of the video and create an index that can be used for searching within the video. This

techniques are difficult to automate and time-consuming to do manually.

### B. Existing Lecture video Repositories

NPTEL [2], freevideolectures.com [4] and MIT Open Courseware [5] are some of the existing lecture video repositories. We investigated support for search and browsing features available in those repositories. Unfortunately, many of the repositories are not providing search functionality for their users. Some repositories have manual transcriptions (subtitles) for lecture videos but they are not making use of them to provide search features.

TABLE I
LECTURE VIDEO REPOSITORIES COMPARISON

| Repository | Search | Navigation Feature |
|---|---|---|
| NPTEL | No | No |
| Freelecturevideos.com | Meta-data | No |
| Videolectures.com | Meta-data | Slide Synchronization |
| MIT Open Course Ware | Content | Speech-transcript |

### C. Lecture Video Retrieval

Tuna et al. [7] presented their approach for lecture video indexing and search. First, they segment lecture videos into key frames by using global frame differencing metrics. Then standard OCR engine is applied for gathering textual metadata from slide, in which they apply some image transformation techniques to improve the OCR result. Jeong et al. [8] proposed a lecture video segmentation method using Scale Invariant Feature Transform (SIFT) feature and the adaptive threshold. 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 attained promising results for processing one-scene lecture video.

Recently, collaborative tagging has become a popular functionality in lecture video portals. Sack and Waitelonis [9] and Moritz et al. [10] apply tagging data for lecture video retrieval and video search. Automatic Speech Recognition (ASR) provides speech-to-text information on spoken languages, which is thus well suited for content-based lecture video retrieval. Leeuwis et al. [11] and Munteanu et al. [12] focus on English speech recognition for Technology Entertainment and Design (TED) lecture videos and webcasts. In this, the training dictionary is created manually, which is thus hard to be enhanced or optimized periodically. In this way, OCR and ASR are used to obtain transcript.

### D. Content Based Video Retrieval

Several content-based video search engines have been proposed in recent times. Adcock et al. [13] presented a lecture webcast search system in which they applied a slide frame segmented to extract lecture slide images the system retrieved more than 36,000 lecture videos from different resources such as YouTube, Berkeley, etc. The search indices 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. In the CONTENTUS [14] project, a content based semantic multimedia retrieval system has been developed. After the digitization of mass media data, several analysis techniques, e.g., OCR, ASR, video segmentation, automated speaker speech recognition, etc., have been applied for metadata generation.

## III. Lecture Video Segmentation Approaches

In multimedia-based learning systems, there is need of segmenting of lecture videos and organizing them into topic and subtopics. Basic problem in any lecture video is to give semantic query and effectively retrieve relevant contents form long video. Effective and efficient search capability for the students can be provided if proper browsing facility is provided.

H. J. Jeong et al. [20] proposed a highly accurate method for video segmentation using SIFT and an Adaptive threshold. Using SIFT, we can easily compare two slides, having similar Contents but different backgrounds. And we can calculate Frame transition quite accurately by using Adaptive threshold.

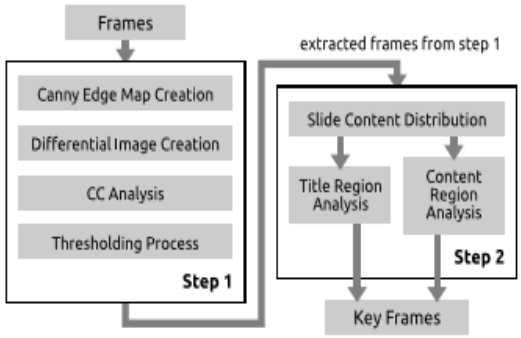


Fig. 1 Workflow of slide video segmentation [19]

### A. Lecture Video Segmentation by Automatically Analysing the Synchronized Slides

Xiaoyin Che et al. [15] proposed a solution which segments lecture video by analysing its supplementary synchronized slides. The slides content arises automatically from OCR (Optical Character Recognition). Then partition the slides into different subtopics by examining their logical relevance. As the slides are synchronized with the lecture video, the subtopics of the slides indicate the segments of the video. Then OCR results are the inputs for the whole procedure.

*1) Global Segmentation:* The aim of global segmentation is to segment the lecture or presentation by its main structure. Xiaoyin et al. [15] attempt to figure out all possible 'border' and generate segments based on them. There are tag-page, split-page and section-page found in lecture video. A tag-page in fact is an outline of the whole slides, with a special title such as 'Overview', 'Titles', 'Topics' or 'Outline', and its content containing most or all the subtopics. Split-page is other kind of widely used 'border' by the lecturer or presenter. Not like tag-page or split-page above, section-page is more

than a border. A section-page includes all the features a common slide may have, definitions, explaining algorithms or pictures.

*2) Partial Segmentation:* Partial segmentation process is used to explore the logical correlation among several neighbouring slides. Under partial segmentation, PLS (Partially Logical Segment) will be found out, by which some slides with continuous or relevant content can be gathered together. Compared to GLS, PLS is less convincing, but still reasonable. Partial segmentation process moves in 2 steps: index-page and virtual index-page. The content of an index-page is a preview of a sequence of following common slides, so it is natural to combine all these slides together as a PLS. The method of index-page searching is also similar to the tag-page, by discovering the similarity of text-lines in the potential index-page and titles from following slides. Virtual index-page derives from a series of continuous slides sharing some words in their titles. In this case, those slides are very likely describing similar topics, and can be packaged as a whole.

*3) Default time Segmentation:* A time segmentation procedure is reserved to apply for the rest of the presentation except the GLSs and PLSs, by which all segments will not be too long. The length of TS (Time Segment) depends on the average length of logical segments in the same presentation, or else, if there is neither GLS nor PLS found, a time segment should not longer than 1/4 of the whole presentation. The annotation text of a TS adopts the title of the slide in this segment with longest duration.

### B. *Lecture Video Segmentation based on Spontaneous Speech Recognition*

Natsuo Yamamoto et al. [16] proposed a segmentation method of continuous lecture speech into topics. A lecture includes several topics but it is difficult to judge their boundaries. To solve this problem, transcriptions obtained by spontaneous speech recognition of a lecture speech is associated with the textbook used in the lecture. In this work, 2-pass search strategy is adopted. At the 1st-pass, a word graph is generated using the lexical tree search with bigram language model. Then, at the 2nd-pass, the best sentence is searched in the word graph.

Using the acoustic score computed at the 1st-pass and trigram language model. In order to absorb the acoustic mismatch between training speech and real lecture speech, MLLR (Multiple linear regression) adaptation technique is employed. At first, the input lecture speech is recognized using a basic acoustic model and the language model. Then the transcription obtained from the speech recognition is used for MLLR adaptation.

### C. *Lecture Video Segmentation based on Text: A Method Combining Multiple Linguistic Features*

Ming Lin et al. [17] make use of the transcribed speech text extracted from the audio track of video to segment lecture videos into topics. Approach utilizes features such as noun phrases and combines multiple content-based and discourse-based features.

Approach utilizes the idea of sliding window in terms of method of finding boundaries. Move a sliding window (e.g. 120 words) across the text by certain interval (e.g. 20 words). Then compare the similarity between two neighbouring windows (one gap), and then draw a similarity graph for all the comparison or gaps. The gap with lowest values (most dissimilar) are identified as possible topic boundaries.

Algorithm takes the transcript text as input, and uses GATE to handle tokenization, sentence splitting, and part-of-speech (POS) tagging. The POS tagger in GATE is a modified version of the Brill tagger, which produces a part-of-speech tag as an annotation on each word or symbol. Porter's stemmer was used for suffix stripping.

TABLE III
LECTURE VIDEO SEGMENTATION APPROACHES

| Approach | Author | Description |
|---|---|---|
| Synchronization of lecture slides | Xiaoyin Che et al. [15] | Segments lecture video by analysing its supplementary synchronized slides. |
| Spontaneous speech recognition | Natsuo Yamamoto et al. [16] | Transcriptions obtained byspeech recognition of is associated with the textbook used in the lecture. |
| Text transcript | Ming Lin et. al. [17] | A Method Combining Multiple Linguistic Features. |

## IV. TEXT DETECTION USING OCR

OCR was initially developed for high contrast data images, taken from metal and other surfaces with uneven roughness and reflectivity.

Content-based gathering within video data requires textual metadata that has to be provided manually by the users or that has to be extracted by automated analysis. For this purpose, techniques from common OCR focusing on high-resolution scans of printed (text) documents have to be improved and adapted to be also applicable for video OCR. In video OCR, video frames containing visible textual information have to be identified first. Then, the text has to be separated from its background, and geometrical transformations have to be applied before common OCR algorithms can process the text successfully [19].

Texts in the lecture slides are closely related to the lecture content, can thus provide important information for the retrieval task. In the detection stage, an edge-based multi-scale text detector is used to quickly localize candidate text regions with a low rejection rate. Then Stroke Width Transform (SWT) [18]-based verification procedures are applied to remove the non-text blocks. The video text images are converted into a suitable format for standard OCR engines.

A vigorous approach to retrieve text from a colour image was given by Y. Zhan et al. [21]. The proposed algorithm uses the multiscale Wavelet features and the structural information to locate the text lines. Then a Support Vector Machine (SVM) classifier was used to get the exact text from those previously located text lines. H. Yang et al. [22] has developed a Skeleton-Based binarization method to separate and extract text from complex backgrounds. These can be processed by standard OCR software.

## V. VIDEO RETRIEVAL METHODS

Keywords can summarize a document and are widely used for information retrieval in digital libraries. Keywords generated from Optimal Character Recognition (OCR) and Automatic Speech Recognition (ASR) summarizes the document or Video. These keywords are used for information retrieval from Video archives.

J. Fan et al. [3] has proposed a new Framework, called Class View for more advanced content-based video retrieval. The important concept they have proposed is, a hierarchical video classification technique to minimize the difference between low level visual features and high level visual concepts. In conventional retrieval, the Euclidean distance between the database and the query is calculated. Short distance indicates that there are more similarities between query frame and database frame. Using this, it is easier to group and retrieve videos.

## VI. CONCLUSION

In this paper, we presented a content based approach to retrieve textual data automatically over videos. This paper has presented a brief review of the lecture video segmentation approaches used within the area of E-learning System. Sometimes, it is cumbersome for students to search through a full video or across many videos, to find particular portions of their immediate interest. In order to remove this difficulty, video segmentation and tagging method is used to extract topics from video for indexing.

## ACKNOWLEDGMENT

We express our thanks to publishers, researchers for making their resource available & teachers for their guidance. We also thank the college authority for providing the required infrastructure and support. Last but not the least we would like to extend a heartfelt gratitude to friends and family members for their support.

## REFERENCES

[1] Glass, James R., et al. "Recent progress in the MIT spoken lecture processing project." *INTERSPEECH*. 2007.
[2] NPTEL http://www.nptel.iitm.ac.in/
[3] Smoliar, Stephen W., and HongJiang Zhang. "Content-based video indexing and retrieval." *IEEE multimedia* 1.2 (1994): 62-72.
[4] freevideolectures.com http://www.freevideolectures.com/
[5] MITOPENCOURSEWARE http://ocw.mit.edu/courses/audio-video-courses/
[6] Smeaton, Alan F. "Techniques used and open challenges to the analysis, indexing and retrieval of digital video." *Information Systems* 32.4 (2007): 545-559.
[7] Tuna, Tailfin, et al. "Development and evaluation of indexed captioned searchable videos for STEM coursework." *Proceedings of the 43rd ACM technical symposium on Computer Science Education*. ACM, 2012.
[8] Jeong, Hyun Ji, Tak-Eun Kim, and Myoung Ho Kim. "An accurate lecture video segmentation method by using sift and adaptive threshold." *Proceedings of the 10th International Conference on Advances in Mobile Computing & Multimedia*. ACM, 2012.
[9] Sack, Harald, and Jörg Waitelonis. "Integrating social tagging and document annotation for content-based search in multimedia data." *Semantic Authoring and Annotation Workshop (SAAW)*. 2006.
[10] Moritz, F., M. Siebert, and C. Meinel. "Community tagging in tele-teaching environments." *2nd International Conference on e-Education, e-Business, e-Management and E-Learning (to appear)*. 2011.
[11] Leeuwis, Erwin, Marcello Federico, and Mauro Cettolo. "Language modelling and transcription of the TED corpus lectures." *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP'03). 2003 IEEE International Conference on*. Vol. 1. IEEE, 2003.
[12] Munteanu, Cosmin, et al. "Automatic speech recognition for webcasts: how good is good enough and what to do when it isn't." *Proceedings of the 8th international conference on Multimodal interfaces*. ACM, 2006.
[13] Adcock, John, et al. "Talk miner: a lecture webcast search engine. *"Proceedings of the international conference on Multimedia*. ACM, 2010.
[14] Nandzik, Jan, et al. "CONTENTUS—technologies for next generation multimedia libraries." *Multimedia tools and applications* 63.2 (2013): 287-329.
[15] Che, Xiaoyin, Haojin Yang, and Christoph Meinel. "Lecture video segmentation by automatically analysing the synchronized slides." *Proceedings of the 21st ACM international conference on Multimedia*. ACM, 2013.
[16] Yamamoto, Natsuo, Jun Ogata, and Yasuo Ariki. "Topic segmentation and retrieval system for lecture videos based on spontaneous speech recognition."*INTERSPEECH*. 2003.
[17] Lin, Ming, et al. "Segmentation of lecture videos based on text: a method combining multiple linguistic features." *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*. IEEE, 2004.
[18] Epshtein, Boris, Eyal Ofek, and Yonatan Wexler. "Detecting text in natural scenes with stroke width transform." *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010.
[19] Yang, H., and C. Meinel. "Content Based Lecture Video Retrieval Using Speech and Video Text Information." (2014): 1-1.
[20] Jeong, Hyun Ji, Tak-Eun Kim, and Myoung Ho Kim. "An accurate lecture video segmentation method by using sift and adaptive threshold." *Proceedings of the 10th International Conference on Advances in Mobile Computing & Multimedia*. ACM, 2012.
[21] Zhan, Yaowen, Weiqiang Wang, and Wen Gao. "A robust split-and-merge text segmentation approach for images." *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*. Vol. 2. IEEE, 2006.
[22] Yang, Haojin, Bernhard Quehl, and Harald Sack. "A framework for improved video text detection and recognition." *Multimedia Tools and Applications* 69.1 (2014): 217-245.