ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 5, Issue 6, June 2016

# Text Query Based Indexing For Efficient Video Retrieval Using Hashing

1

**Manasi A. Kabade, Prof. U. A. Jogalekar**

***Abstract*- In recent years, e-lecturing is becoming more and more popular. Different research organizations put their e-lectures over World Wide Web. With help of which students or researchers can have access to these lecture videos without any restrictions of time and space. As a result of which, video data over World Wide Web is increasing very rapidly. Thus retrieval and access of video from large video archives is nearly inefficient. Text in the e-lecture videos provide appropriate clue that can be used as search index for video retrieval. We have proposed more efficient technique of hashing for faster retrieval of query text in videos. For query text retrieval open source OCR package can be used. After storing extracted text from videos, we are using hashing technique over texts to generate hash codes. This secures the structural information of each video frame with its extracted text and this will take minimum time to search videos from large databases and retrieve videos faster. As a result, we can retrieve lecture videos efficiently from big video datasets.**

***Index Terms*- Content-based video search, video segmentation, Key-frames, Automatic Speech Recognition(ASR), Optimal Character Recognition (OCR), Hashing.**

## I. Introduction

With the fast development of Internet, and wireless sensor networks, the need of video indexing and retrieval is rapidly increasing. The textual information in lecture videos provides an important clue for search and indexing of video because texts have semantic information which is more related to the contents of videos as compared to images. However, the detection of texts in lecture video is a main issue. The reason is that the texts in videos have different font type and text color and different backgrounds. However, detection of text and their location is the first step of text information extraction from e-lecture videos, and also in research they are under high attention. In today's world, most of the Universities and researchers from different organizations try to upload videos of lectures and seminars, so that anyone can have free access over World Wide Web (www) without limitation of their location and time. This result into huge amount of e-lecture Video data on www and which is increasing very fast and continuously. Video retrieval is most of the times focused on extraction of visual features that can't be connected to video recordings. Motion changes of the camera may affect the shape, size and the edge splendor. Each slide of lecture video, can be incompletely hindered when the presenter moves between camera and slide. Additionally, motion recognition process may get influenced by progressions of camera center which is starting with one point then onto the next. This can be minimized by displaying only solitary video of the presenter which is synchronized with respective slide recording.

Digital video is used as a source for recording Meetings, seminars, research lectures etc. which is very easy to make available online as well as the progressive development in recording e-lectures makes it available widely over internet. While recording the videos information is added such as authors, genre, topic names etc, by video producers. But, when user when searches for required video, most of the times he is not satisfied about the information retrieved back from that particular video. Sometimes, even after watching whole video, the user may not get appropriate piece of information after watching those lengthy videos. Because these video search systems like YouTube, vimeo replies the users with the global metadata data, e.g. title, headings, and brief description, etc.

This global metadata provided may be irrelevant or not sufficient most of the times. Therefore there is huge need of some better technique, which will retrieve maximum amount of information without playing whole lengthy videos. We can use textual data that is indexed automatically. In videos, segments of videos serve as a layout for the video. In this way video document can be divided into a group of key frames, the text identification can be executed on each key segment, and the resulted articles will be further used as a differentiating factor over segments. Particularly, the metadata in slides is always in structured format that can be used as a powerful and more adaptable video perusing.

We extract overall metadata from video slides using OCR which is open source. Also ASR is used to extract text from speech of lecturer from e-lecture videos. And then we have designed an approach with hashing architecture over text extracted from videos, which in turn generate hash codes of text, so that the video retrieval process could be easily handled and the user could get time saving and efficient functionalities of video indexing.

## II. RELATED WORK

Yang et al. have proposed a technique for content-based lecture video retrieval [16] in large databases of lecture video. Most of the OCR-based lecture video indexing approaches that are existing, they make use of the stroke width value and the geometrical information of text detected and the bounding boxes for extracting structure of lecture slides from the slide video key-frames. They have

***Manuscript received Jun, 2016*.**

***Manasi A. Kabade**, Department of Computer Engineering, Savitribai Phule Pune University, Smt. Kashibai Navale College of Engineering, Pune, India, 9503875661.*

***Prof. U. A. Jogaleker,** Department of Computer Engineering, Savitribai Phule Pune University, Smt. Kashibai Navale College of Engineering, Pune, India*

adapted and integrated engines of visual analysis in an automated framework of the lecture video portal. They make use of visual resource of lecture videos for extracting content-based global information from e-lecture videos. Searching videos and extracting text from these data is totally time consuming and it have to go through consistency and solidity problem.

How to leverage on the idea of stroke width for text detection recovery is proposed by Boris et al [14]. They noted the stroke notion and had derived an effective algorithm to compute complexity of it, by generating new image feature. Once recovered, it provides a flexible image feature for searching text in image. For text detection, the SWT combines dense estimation at every image pixel by using stroke width which depends on metadata that is far from pixels. In [10], they proposed a e-lecture video segmentation methods by using Scale Invariant Gimmick Change (Filter) feature and the adaptive threshold. In their proposed work Filter peculiarity is used to calculate slides with substance. Adaptive threshold is utilized to determine slide moves. In their assessment, this terminology achieved guaranteeing results for one-scene lecture video handling.

Communitarian tagging has achieved a well known popularity in gateways of lecture video. Waitelonis and Sack [11] and Moritz et al. [12] apply combination of CTS with an annotation system for multimedia presentations that are synchronized that is able to annotate single multimedia data parts with tags that are defined by user. They developed a system for annotation of synchronized multimedia documents of lecture video recordings. The desktop presentation of lecture is synchronized with a video recording of the lecturer [11 that is used for MPEG-7 metadata automated creation and which enables content-based annotation of single scenes. They used tagging information for recovery of lecture video and for searching video. With the keyword-based tagging, Yu et al. proposed a methodology to expand lecture video assets by using information of Interfaced components. These techniques empower clients to comment videos utilizing vocabularies that are characterized as a cloud of Joined Information. Then for video searching those semantically joined instructive assets are used. On the other hand, the expense and exertion that is required by the client annotation-based methodology are not able to fulfill the prerequisites for preparing a information of web video with a pace that is expanding very fast. They utilized Joined Information to further consequently expand the concentrated textual global metadata that is utilized for comprehensive survey for detection of text [14]. In general, the techniques for detecting text can be categorized in two groups: Firstly, texture-based methods that are used to scan the image at a number of scales, that classify neighbors of image pixels based on a number of text properties, such as high density of edges, high variance of intensity, low gradients above and below text etc. The one of the limitations of these methods include computational complexity due to the need of scanning the image at several scales, problems with integration of information from different scales and lack of precision due to the inherent fact that only small (or sufficiently scaled down) text exhibits the properties required by the algorithm. Also, one more limitation is, if text is slanted then it cannot be detected efficiently by above algorithms.

Grcar et al. presented Videolectures.net in [8] which is an advanced document for media presentations. Wang et al. proposed a methodology for lecture video indexing focused around mechanized video segmentation and OCR examination [14]. The proposed segmentation algorithm in their work is focused around the differential proportion of text and foundation districts. Utilizing thresholds they endeavor to catch the slide move. The last segmentation results are dictated by synchronizing caught slide key-frames and related text books, where the text comparability between them was computed as pointer. Like [14], the creators likewise apply a synchronization handle between the recorded lecture video and the slide document, which must be given by moderators. These two methodologies specifically dissect the video, which is therefore autonomous of any fittings or presentation innovation. The obliged slide group and the synchronization with an outer archive are not needed. Besides, since the animated substance evolvement is regularly connected in the slide, yet has not been considered in [14] and [8], their framework may not work powerfully when those impacts happen in the lecture video.

Weiming et al. in [15] used algorithms to form frames of cluster and then selects frames that are closest to the center of the cluster to set the key frames. Girgensohn and Boreczky select video segments for which they uses the method of hierarchical agglomerative clustering. They extract key-frames using the K-means of fuzzy clustering in subspace of color feature. Gibson uses Gaussian mixture models for the image. For this they uses the comparative method in which, number of components of GMM is equals to the required number of k-means clusters. The clustering-based algorithms use generic clustering algorithm which is main advantage of them, and we can get key-frames with their global characteristics of a video reflected in it. The one of the limitations of these algorithms are as follows: First, they are totally dependent on the results of clustering, but semantic meaningful clusters with successful acquisition is very difficult, especially for larger datasets, and another limitation is, the sequential feature of the video frames cannot be utilized. Usually, complex techniques are used to ensure that one cluster will be having adjacent frames, so one cluster will have only frames that are adjacent to each other.

## III. PROPOSED SYSTEM

### I. *Problem Definition-*

To find efficient method to retrieve e-lecture videos faster from large lecture video database.

### I. *System Overview-*

Video Framing

System accepts lecture videos as input and will create video frames.

Text Extraction

System extracts text from extracted key-frames using Optical Character recognition and ASR techniques. Extracted texts are written into files and used these files at the end to search required video from large dataset.

Hashing

System generates hash values for text extracted using OCR, ASR and for input query entered by user. This hash code is used for retrieving resultant videos from dataset.

### *I. System Architecture -*

We have also proposed a new method to increase the efficiency of the existing system. For this, in training phase, we are using the hashing technique on the text extracted from lecture videos to generate hash codes. Then, we will also apply hashing on query text and will compare it with the hash code of query text. In this way we achieve a faster and efficient indexing for better results.

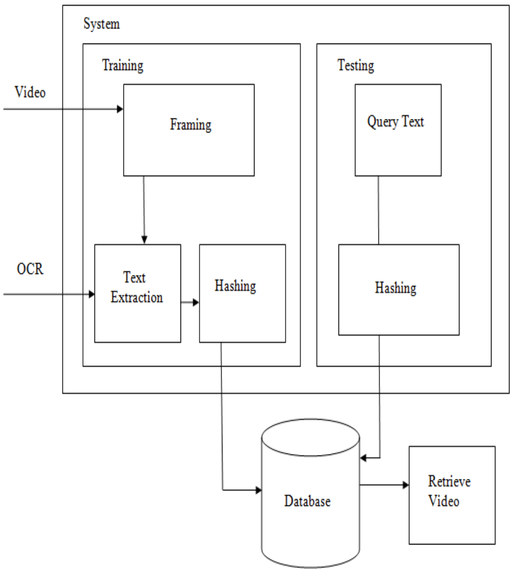

Fig 1: System Architecture

## IV. EXPERIMENTAL SETUP

The system is built using Java framework (version JDK 8) on Windows platform. The NetBeans (version 8) is used as a development tool. The system doesn't require any specific hardware to run; any standard machine is capable of running the application.

## V.RESULT & DISCUSSION

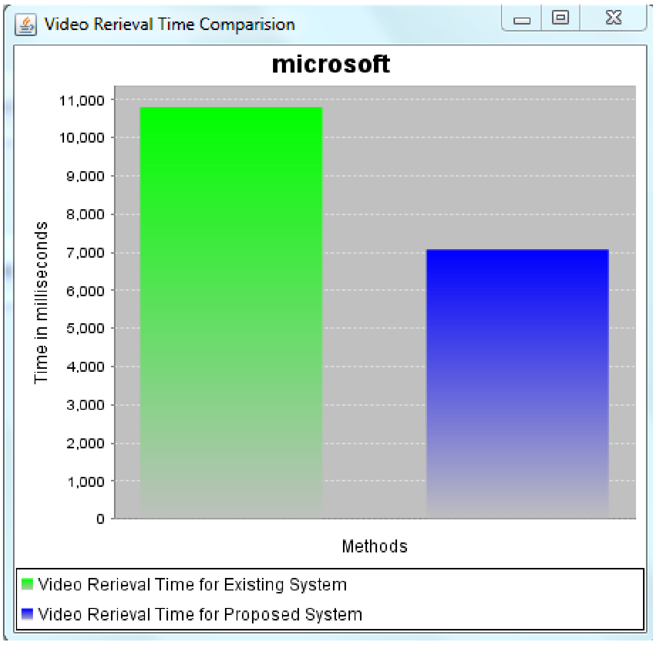

Fig.2: Time Comparison Graph

Above figure shows time required in milliseconds to search query text of required e-lecture video with existing system and proposed system from given input dataset of videos. The limitation of existing system is that it searches for the query text by string matching from extracted texts, which takes more time as compared to the time required for hash code comparison for the same. Thus, video retrieval can be done faster using hashing technique.

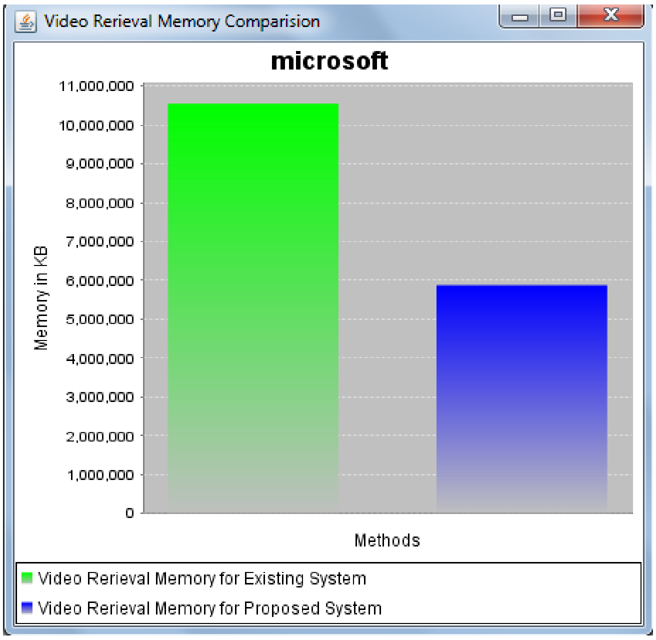

Fig.3: Memory Comparison Graph

Also, Fig3 shows results of memory comparison. Using hashing technique creates message digest which is actually very smaller in size as compared to original texts in videos. Thus, memory required is also very less with proposed system.

## VI. CONCLUSION

New framework is designed for Content-based video retrieval from large e-lecture video archive. The Hashing technique used meets the faster similarity search for text queries. This new framework with hashing technique shows time required is less as compared to text query comparison with string matching. Also the accuracy in terms of number of correctly retrieved videos shows better results.

## VII. ACKNOWLEDGEMENT

I am very thankful to the Savitribai Phule Pune University, Pune. I am highly grateful to my department and college, Department of computer engineering and Smt. Kashibai Navale College of Engineering to providing all the facilities. I am thankful to my esteemed guide Prof. U. A. Jogaleker for expert and precise guidance.

## 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.

[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.

[4] C. Meinel, F. Moritz, and M. Siebert, 'Community tagging in tele-teaching environments,' in Proc. 2nd Int. Conf. e-Educ., e-Bus., e-Manage. and ELearn., 2011.

[5] S. Repp, A. Gross, and C. Meinel, 'Browsing within lecture videos based on the chain index of speech transcription,' IEEE Trans. Learn. Technol., vol. 1, no. 3, pp. 145ˆa156, Jul. 2008.

[6] Boris Epshtein and Eyal Ofek, "Detecting Text in Natural Scenes with Stroke Width Transform".Image operator that seeks to find the value of stroke width for each image pixel,2012.

*[7]* Weiming Hu, Senior Member, IEEE, and Stephen Maybank, 'A Survey on Visual Content-Based Video Indexing and Retrieval' IEEE Trans. on Systems, VOL. 41, NO. 6, Nov 2011.

*[8]* Haojin Yang and Christoph Meinel, 'Content Based Lecture Video Retrieval Using Speech and Video Text Information' IEEE Trans. On Learning Technologies, VOL. 7, NO. 2, APRIL-JUNE 2014.