

Content Based Video Retrieval

Wagh K. D.¹, Dr. M. U. Kharat²

P.G. Scholor, Department of Computer Engineering, M.E.T.B.K.C.I.O.E, Nashik, India¹

Professor, Computer, M.E.T.B.K.C.I.O.E, Nashik, India²

Abstract: For fast learning and because of availability of fast techniques e-learning and e lecturing concept has more popular. On Internet lots of lectures are available for single concept there an efficient way to perform video retrieval is needed. Video retrieval is a young field which has its genealogy rooted in Artificial Intelligence, Digital Signal Processing, etc, this paper presents a text based video retrieval and video search system. Very first apply automatic video segmentation and key frame detection for video content navigation. Extract textual metadata by applying video OCR technique on key frames and ASR on lecture audio track. This summery will then be used for indexing of video. As there is strong needs for segmenting lecture video into topics for browsing.

Keywords: Content based video retrieval, Optical character recognition, Automatic speech recognition, Segmentation.

I. INTRODUCTION

Now a days number of universities provide facility ti students by recording the lectures, maintaining hundreds of lecture video recordings in a repository, and making them available to remote students over the Internet. A number of universities and research institutions are taking the opportunity to record their lectures and publish them online for students to access independent of time and location. Digital video has become a popular storage and exchange medium due to the amazing growth in recording technology, improved video compression techniques and high speed networks in the last few years. Therefore audiovisual recordings are used more and more frequently in e-lecturing systems. As a result, there has been an enormous increase in the amount of multimedia data on the Web. Hence, for a user it is impossible to find desired videos output without a search function within a video archive. Duplicate contents in video frustrate the user. Universally accepted video retrieval and indexing technique are not well defined or available. These all challenges motivate us to present video retrieval system

This paper presenting Content Based Video Retrieval (CBVR) System it includes various steps: Video Segmentation: Video segmentation is first step of content based video search aiming to segment moving objects in video sequences, segments the video by setting the time, Key frame Selection: Selects the key frame among the extracted frames of the video, frame is nothing but the image or shots, when that image is compressed then that image is nothing but the key frame. Feature Extraction: Features are extracted for the key frame and stored into feature vector. There are two types of features first is spatial and another is temporal. Spatial features are then classified as color, shape and edge; similarly temporal features are also then classified as motion and audio.

Classification and Clustering: Automatic Content Based Retrieval and Semantic Classification of the Video Contents presented a learning framework where construction of a high-level video index is visualized through the synthesis of its set of elemental features. Indexing: For indexing of key frames Hierarchical

Clustering Tree Algorithm (HCT) is used. Matching Similarity: For retrieving the video from database, the retrieval subsystem processes, performs similarity matching operations and finally displays the result to the user. In retrieval stage of a video search system, features of the query which is given by end user is also extracted. After that the similarity between the features of the query video and the stored features of video in database is determined. It means it computes the similarity between two videos, this similarity result is then used to give a distance between the query video given by the user and a candidate match from the feature database

II. RELATED WORK DONE

From in recent years the need for content-based retrieval of image and video information from internet archives has create the attention of almost all researchers. Research efforts have caused to the development of methods that provide access to image and video data. The methods are used to determine the similar things in the visual information content extracted from low level features. These features are clustered for creation of database. This section discussed the related work done with previous studies. Most of the primary work related to Content based video retrieval mainly focus on video segmentation and summarization. To obtain efficient and effective result for video search, very first video has to be segmented and all text information and meta data need to be extracted for indexing as well as tagging purpose. Techniques for searching in lecture videos. [1][2][3][4][5][6] used poor quality cameras for recording a lecture therefore they apply State of Art Lecture recording system such as Tele-Teaching Any Where Solution Kit. Lecture is conducted in two main parts first part is the main scene of lecture is recorded by using a video camera and the second part is capture the speech by desktop of the speaker's computer by using these synchronization problem is occur the main drawback is that the video analysis may introduced error. These problem is solved in next all papers Segmentation of lecture video is most important thing in complete work

a lecture video segmentation method used Scale Invariant Feature Transform (SIFT) feature and the required calculated threshold value. In previous work of [1] SIFT feature is applied to measure slides with similar content. An adaptive threshold selection algorithm is used in their work to detect slide transitions. In their evaluation, this approach attained promising results for processing one scene lecture video.[1]

Now a day, in lecture video portals a collaborative tagging has become a popular functionality, H.Sack and J. Waitelonis applied tagging data for lecture video retrieval and video search. Beyond the keyword based tagging, they used Linked Data to annotate lecture video resources. The 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.[9][10]

Search facilities can be provided in lecture video repositories in two ways. 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 metadata 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.[12]

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. A highly accurate method for video segmentation using SIFT and an Adaptive threshold. Using SIFT, can easily compare two slides, having similar Contents but different backgrounds. And can calculate Frame transition quite accurately by using Adaptive threshold.[11]

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. 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) 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.[13]

The synchronization problem is solved proposed a solution which segments lecture video by analyzing 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[14]

III. ARCHITECTURAL BLOCK DIAGRAM

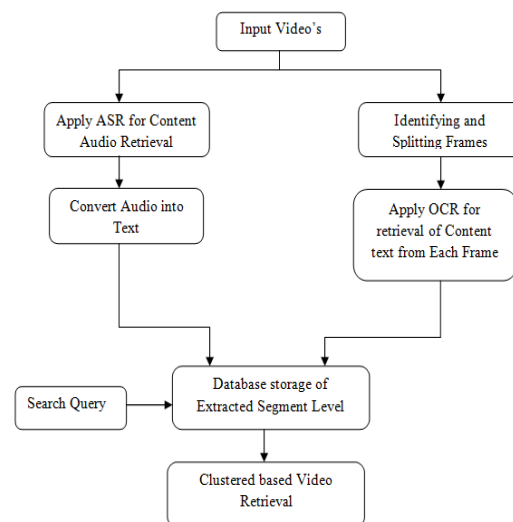


Fig. 1. System Architecture

Fig.3.2 shows the architecture of proposed system The objective of the proposed system is to retrieve a video according to its contents rather than retrieving video according to its title and metadata description, by using this it will provide a desired result for the query given by user . For this develop a model which captures the various frames from a video lecture. All these frames are then

classified according to the duplication property and duplicate frames are discarded. Frames are then compressed using compression techniques. Video segmentation is performed by considering a time interval within two frames. Sometimes it happens that a video lecture contains one slide presentation for a long period of time then maximum time interval is used in seconds. After key frames segmentation, we fetch all the text from all the frames for further video retrieval system using optical character recognition (OCR) algorithm. Also we fetch all the voice resulting into text using ASR technique. This is also used in the process of video retrieval system. The above information (Text and Voice from Video) is used for content based video retrieval system and clustering of video according to their text and voice parameters. The optical character recognition(OCR) algorithm is applied to extract the characters from the textual information also automatic speech recognition(ASR) algorithm is used to retrieve the speech(wave) information from the video lecture. The OCR and ASR transcript and determine slide text line types are adopted for keyword extraction, by which both video and segment-level keywords are extracted for content-based video browsing and search.

In short the architecture diagram shows how video is handled. Very first the input video is uploaded by admin after that video is divided in two parts first is wave signal another is text format, then ASR is applied on speech and get the output as text, on the other hand segmentation is done create the frames after that applied OCR on key frames get the output as text, from both side output is text. Stored this output as a database. When the user search query then preprocessing is done that query is then mapped with database if found then user get the relevant output.

IV. SYSTEM FLOW

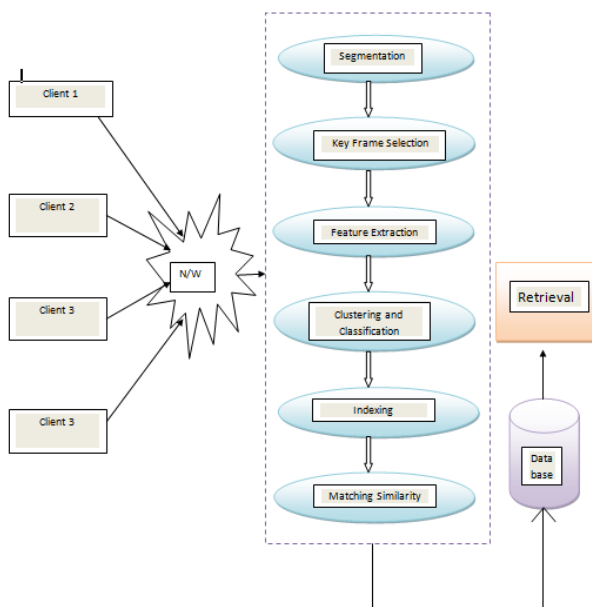


Fig. 2. System Flow Diagram

The above figure shows the flow of system in this system different users are act as client; number of clients firstly

interacting with the server through the network. The server system also then interacted with database warehouse, the video data stored in warehouse. In server side the server system does various operations or actions on video clips these operations are segmentation of a video, key frame extraction from generated key frames, feature extraction from each frame, classification and clustering, indexing and matching similarity by using various algorithms.

These features are mapped with features stored in data warehouse and user gets the final retrieved result from stored database.

A. Video Segmentation

Video segmentation is very first step of the content based video search it aim is to segment moving objects in video sequences. Segmentation is done with the help of high level camera by setting the time to particular time limit and creates the frames , then that frames are converted into compressed images called key frames.

A. Key-Frame Selection

Selects the key frame among the extracted frames of the video, for framing purpose used create a bitmap by setting counter from 0 to stream length, saved that bitmap.

B. Feature Extraction

Features are extracted for the key frame and stored into feature vector. Features are of two types that are **spatial** and **temporal**. Colour, Shape, Edges are spatial features and Motion and Audio are temporal features.

C. Classification and Clustering

For classification purpose use SVM(Support Vector Machine) .The support vector machines associate each set of data points in the multidimensional feature space to one of the classes during training. Clustering is done with k-means clustering algorithm.

D. Indexing

For Indexing purpose use indexing algorithm to store index of the specified frames. Indexing is done with the help of Hierarchical Clustering Tree (HCT) algorithm.

E. Matching Similarity

In retrieval stage of a Video search system, features of the given query video is also extracted. After that the similarity between the features of the query video and the stored feature vectors is determined. That means that computing the similarity between two videos can be transformed into the problem of computing the similarity between two feature vectors [11]. This similarity measure is used to give a distance between the query video and a candidate match from the feature data database

V. CONCLUSION

This work present content based approach to retrieve data in format of text automatically over videos. This paper has gave a brief review of the lecture video segmentation approaches used within the area of E-learning System over a internet. Many of the time it is difficult 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.

ACKNOWLEDGEMENT

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] 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.
- [2] W. Hurst, T. Kreuzer, and M. Wiesenhuber, "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.
- [3] A. Haubold and J. R. Kender, "Augmented segmentation and visualization for presentation videos," in Proc. 13th Annu. ACM Int. Conf. Multimedia, 2005.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] T.-C. Pong, F. Wang, and C.W. Ngo, "Structuring low-quality videotaped lectures for cross-reference browsing by video text analysis," J. Pattern Recog., vol. 41, no. 10, pp. 3257-3269, 2008.
- [8] M. Gracar, D. Mladenic, and P. Kese, "Semi-automatic categorization of videos on video lectures.net," in Proc. Eur. Conf. Mach. Learn. Knowl. Discovery Databases, 2009.
- [9] H. Sack and J. Waitelonis, "Integrating social tagging and document annotation for content-based search in multimedia data," in Proc. 1st Semantic Authoring Annotation Workshop, 2006.
- [10] 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 E-Learn., 2011.
- [11] H. J. Jeong, T.-E. Kim, and M. H. Kim, "An accurate lecture video segmentation method by using sift and adaptive threshold," in Proc. 10th Int. Conf. Advances Mobile Comput, 2012.
- [12] Smeaton, Alan F. "Techniques used and open challenges to the analysis, indexing and retrieval of digital video." *Information Systems* 32.4 (2007): 545-559.
- [13] Epshtein, Boris, Eyal Ofek, and Yonatan Wexler, "Detecting text in natural scenes with stroke width transform." *Computer Vision and Pattern Recognition (CVPR), IEEE Conference 2010*
- [14] 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.