IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

Analysis of Lecture Videos Retrieval: A Content Based Approach

Kalyani D. Wagh[1], **Dr. M. U. Kharat**[2]

PG Student, Dept of Computer Engineering, METs Institute of Engineering, Nashik[1]

Dept of Computer Engineering, METs Institute of Engineering, Nashik[2]

Abstract: Technology is very fast developed in recent year. Students are interested in fast learning, for that they prefer e-learning technology for study. Numbers of videos are available on internet on single concept. In some cases output videos are useful to user or related with users query, but some videos are not related with users query, to avoid such a type of searching use content based video retrieval system. This system consist text and speech based video retrieval and video search system. First step of content based video retrieval system is segmentation and next is key frame detection. After that extract textual data from each key frame by applying OCR(optical Character Recognition) technology also apply ASR(Automatic Speech Recognition) technology on lecture audio track , the both technology give output in form of text data. These text data is useful for content based video retrieval.

Keywords: Segmentation, OCR, ASR, Indexing, Feature Extraction.

I. INTRODUCTION

E-learning is the most preferred source adopted by students these days maximum students have interest to study with the data available on internet as most of the research institutes and universities refer students to get e-lectures from the experts which are uploaded on internet. The most important things of e-lecture is that any student can access these e-lectures any time and any where as many times he wants, sometimes available videos are useful to user or related with users query, but some videos do not fullfill requirements of users, to avoid such type of in accurate searching use content based video retrieval system in this way CBVR is introduced. These days video compression technology has achieved a higher level therefore audiovisual recording are used in recording of a video. Both technologies provide high quality video recording, it is easy to retrieve such type of video. In multimedia field based video retrieval is very interesting area. Video retrieval is nothing but retrieving of video clips from video database by using content, content involves text, image, metadata of video, etc[4]Content based video retrieval consist of two phases first is database population phase and video retrieval phase. Database population phase consist of:

1. Segmentation.
2. Key Frames selection.
3. Feature extraction.

Video retrieval phase consist of:
1. Similarity measure or matching.
Video is a complex data type consist of audio and video audio is then divided into speech music and sound. Number of audio retrieval techniques are available like Hidden Markov Model, Boolean Search with multi-query using Fuzzy Logic, Automatic Speech Recognition, etc[9] Proper segmentation and feature extraction algorithm gives the proper video retrieval it means to improve the result of content based video retrieval system the most important is to develop efficient segmenting and feature extraction algorithm. In previous work segmentation is done with following steps initially video is parse into number of shots by using canny edge detection algorithm by using this shots key frames are generated ,these key frames are used to extract the textual data. This textual data is then stored. User get output of query from this database, current indexing technique is not well defined yet, so this is a challenge to motivate me to present video retrieval system. Various steps for CBVR very first step is segmentation of video. Aim of segmentation is to segment moving object in video sequence, here segmentation process consist a time slot ,set the time such as 1 second (if key frame having time slot less than 1 sec then that key frame is discarded) for capturing a image then OCR extract the text from each image using OCR algorithm. After segmentation next step is key frame selection key frame is nothing but images which are generated from segmentation process. By applying OCR and ASR extract the textual data and save it into database.

II. LITERATURE SURVEY

Wang et al. proposed is mainly working for video indexing having automated video segmentation technique and OCR algorithm analysis. Their proposed segmentation algorithm is totally depend and on the differential ratio of text and background regions. By Using thresholds values they handle slide transition. The final result of segmentation is observed by synchronization of key-frames and its related books, where the similarity between text was calculated as indicator[3].

DOI 10.17148/IJARCCE.2016.57144

Grcar et al. work for multimedia presentations they introduced .net which is a digital archive of Video Lectures. Similar to [3], the authors also apply a synchronization process between the recorded lecture video and the slide because that time camera quality was poor. Recording of video was done in two parts, the lecture is recorded and then speech is recorded by desktop speaker, which has to be provided by presenters, so extra hardware was required that time. Now a day's camera having best quality compare to previous cameras, digital cameras are available now days so my system avoid these two approaches since it directly analyzes the video, which is not dependent on any hardware or presentation technology. Synchronization are not required. The animated content was not considered in [3] and [4], their system might not work properly when those e_ects occur in the lecture video. In [3], the _nal segmentation result is totaly dependent on the working quality of the OCR result. It might be less efficient, when only poor OCR result is obtained [4].

Subhlok et al. presented their approach for video indexing and video search. They used segmentation for converting lecture videos into images those images were nothing but key frames they applied global frame differencing metrics for segmentation process. Then standard OCR algorithm was applied on each key frame for gathering text data. They work on OCR result improvement for that they used some image transformation techniques. They developed a new video player, in which the indexing, search and captioning processes are integrated. Similar to [3], in this work global di_erencing metrics couldnot give a su_cient result for segmentation when animations is presented in the slides. such a type of cases, many duplicate segments will be created these can be avoided by using image transformations technique, still it was not e_cient for recognizing key frames with complex or composite content and background distributions. They observe that text detection technique and segmentation technique could give much better results as compare to image transformations. [5].

Jeong et al. used Scale Invariant Feature Transform (SIFT) feature and the adaptive threshold for segmentation of lecture video. The SIFT feature is helpful to measure slides with similar content. Detection of slide transition an adaptive threshold selection algorithm is used. In their evaluation, this approach achieved promising results for processing one-scene lecture videos. They work for collaborative tagging. In lecture video portals now days collaborative tagging has become a popular functionality [6].

Sack and Waitelonis and Moritz et al, apply tagging data for retrieval of lecture video and video search. According to the Text or 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 that linked educational resources of videos are further used in the video browsing process and video recommendation procedures. The effort and cost for processing of large amounts of web video data with a rapid increasing speed cannot satisfy the requirements of user. Using Linked Data to further automatically annotate the extracted textual metadata opens a future research direction. They also work on speech data for that they used ASR algorithms. ASR converts all the speech into text, which is then more used for content-based lecture video retrieval. The authors of [3] and [8] mainly focus on English speech recognition for Technology Entertainment and Design (TED) video lectures and webcasts. In their system, the dictionary is created manually, which is hard to be extended or maintained periodically. Glass et al.[12] work on improving ASR results and provide the solution for that they collect English lectures. Inspired by their work, I developed an approach for creating speech data from lecture videos[7][8].

III.ARCHITECTURE OF SYSTEM

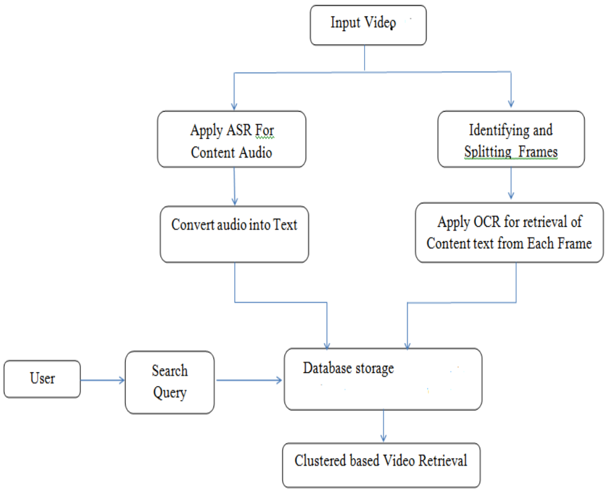
Input Videp
Apply ASR For Content Audio
Identifying and Splitting Frames
Convert audio into Text
Apply OCR for retrieval of Content text from Each Frame
User
Search Query
Database storage
Clustered based Video Retrieval

Fig3.1 architecture of proposed system

video retrieval system consists of an administrator and user, the part of administrator is provide the input video to the system, then the video get divided in two parts here wave file is separated and text file also get separated from complete video, separated part is nothing but key frames. Apply ASR and OCR to each key frame and produce text data. Proposed system having following flow:

. Segmentation
. Key frame selection
. Feature extraction
. Indexing
. Matching similarity.

Fig 3.1 itself show the architecture of proposed system. The important functionalities will be understood by architecture diagram of content based video retrieval (CBVR) system. Video database is maintained by storing the video with extension such as .avi. For the purpose of retrieval develop a model which can capture the various frames (snap shots) from a complete video lecture. All these frames are then classified according to the duplication property and if duplication frames are present then discard it. Next step for proposed system is compress each frame using various compression techniques and create key frames.

3.1 Key frame selection
Select a key frame for further processing, number of key frames get similar. From similar key frames select only one and discard other.

3.2 Feature extraction
There are two ways to handle the feature first is spatial and temporal, color, shapes, edges are spatial feature, spatial feature is nothing but low level feature and motion and audio are temporal feature i.e. semantic level feature.

Here only text and speech are extracted.
Tesseract OCR is applying on key frames and ASR is applying on speech. Steps for both are as follows.

Steps for OCR:

1. Get input path for processing
2. Load images from input path
3. Convert input image into low scale 312 by 256
4. Extract characters
5. Set upper/lower case
6. Resulting text compare with dictionary words.
7. Save character set

ASR algorithm involve following steps:

1. Get input file
2. Extract sound from input file
3. Apply speech recognition engine to sound file
4. Set Dictation Grammar
5. Save text set

3.3 Matching similarity
By OCR and ASR all text gets saved in database. When user search any query using text that text is get matched with database if match is found then user get output.

IV. PERFORMANCE EVALUATION

4.1 Final Results:

The previous method uses various algorithms for segmentation. In my proposed system mainly work on time based segmentation process it provides most relevant result to the user. By using proposed system the time required for searching is less compare to existing system. Here I used cluster based searching so that user can get maximum result for their query.

Following table shows searching result based on time factor.

TABLE I: Time Analysis

Sr No.	Search Query	Video Retrieval Using Established system	Time required for Established system	Video Retrieval Using Existing system	Time required for Existing system
1	level	10	15ms	12	10ms
2	computer	11	15ms	12	13ms
3	story	5	7ms	8	6ms
4	English	7	7ms	10	7ms
5	graphics	3	4ms	5	3ms

4.2 Performance Evaluation:

Performance of the system is evaluated based on the Precision and Recall values. Table-2 shows the how much Precision and Recall is calculated for a given query video.

Precision =:
No. of retrieved videos that are relevant to the query clip
Total no. of retrieved videos

Recall =:
No. of retrieved videos relevant to the query clip_______________________________________
Total no. of relevant videos available in database

TABLE III: Precision recall values

Query Video	Precision	Recall
Level	1	0.92
Computer	0.9	0.733
Story	0.625	0.625
English	0.7	0.63
Graphics	0.6	0.6

4.3 Final Retrieval

After final retrieval of the video from the database, Number of video are fetched according to user's query from database. Following table show the total no of videos retrieved by the system, No of similar videos are available in the database and most matched videos from database with the query video.

Sr No.	Search Query	Most Matched	Total Retrieved by System	Similar Available in Database
1	level	12	12	13
2	computer	11	12	15
3	story	5	8	8
4	English	7	10	11
5	graphics	3	5	5

DOI 10.17148/IJARCCE.2016.57144

Following Graph shows the No. of videos most matched with query video, total videos retrieved by the system and similar videos available in the database.

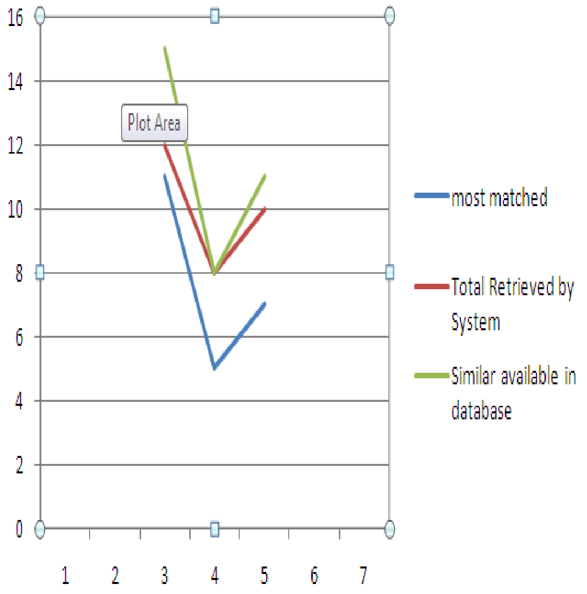

Fig-2. Graph of most matched, retrieved and available video

V. CONCLUSION

Video segmentation, indexing and retrieval is very important in multimedia database management and retrieval. Video contains more information than other media types (text, audio, images).In these system video is search by using text, images, and audio also, it is helpful to all user, this system provide maximum video which is based only on content. Feature extraction is depending on segmentation process. Here segmentation is done with time retrieval so get maximum images for feature extraction. System gets maximum text for storage in database, these is helpful to matching similarity. In matching similarity users query compare with database, it get maximum text in database, and user get a large number of related video.

REFERENCES

[1] Kalyani Wagh and Dr. M. U. Kharat, "Content Based Video Retrieval System," IJARCCA, Jan 2015.

[2] Yang, H., and C. Meinel, Content Based Lecture Video Retrieval Using Speech and Video Text Information. in Proc. of CCS'07, Alexandria, VA,April-June 2014, pp. 142-154.

[3] T.-C. Pong, F. Wang, and C.-W. Ngo, “Structuring low-quality videotaped lectures for cross-reference browsing by video text analysis”, Pattern Recog.vol.41, no.10, pp 3257-3269, 2008.

[4] M. Grcar, D. Mladenic, and P. Kese,Semi-“automatic categorization of videos on videolectures.net”, in Proc. Eur. Conf. Mach. Learn. Knowl. Discovery Databases, 2009, pp 730-733.

[5] T. Tuna, J. Subhlok, L. Barker, V. Varghese, O. Johnson, and S. Shah, Development and evaluation of indexed captioned searchable videos for stem coursework, in Proc. 43rd ACM Tech. Symp. Comput. Sci. Educ., pp. 129-134. [Online]. Available: http:// doi.acm.org/10.1145/2157136.2157177. 2012

[6] 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., pp. 285-288. [Online]. Available: http://doi.acm.org/10.1145/2428955.2429011

[7] 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., 2006.50 Analysis of lecture video Retrieval: A Content Based Approach

[8] 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.

[9] A. Haubold and J. R. Kender, Augmented egmentation and visualization for presentation videos, in Proc.13th Annu. ACM Int. Conf. Multimedia, 2005, pp.51-60.

DOI 10.17148/IJARCCE.2016.57144