Clustering and Retrieval of Video Using Speech & Text Information

[1]Patil Pooja , [2]Kharjul Priyanka, [3]Gholap Manali, [4]Dawange Sonali
[1,2,3,4]BE Student, Computer Department,
Matoshri Collage Of Engineering & Search Center, Nashik, India

Abstract—Now a day's e-lecturing has become more popular on the World Wide Web (WWW) the use of lecture video data is increase rapidly. So, there is need of efficient method for video retrieval in WWW. In our proposed system we are presenting technique for automated video indexing and video search in large videos. First of all, we automatically segment the video and key frame detection for the video. As well as, we apply video Optical Character Recognition (OCR) technology on key-frames and Automatic Speech Recognition (ASR) on lecture audio tracks for text extraction and audio recognition respectively. for content-based video searching by using both video and segment level keywords extracted, after successful completion it is prove that proposed system performance is effective.

Index Terms—Lecture videos, automatic video segmentation, content-based video search.

I. INTRODUCTION (HEADING 1)

Digital video has become a popular storage and exchange medium due to the rapid development in recording technology, improved video compression techniques and high-speed networks in the last few years. Therefore, audio visual recordings are used more and more frequently in e-lecturing systems. 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. As a result, there huge increase in the amount of data on the Web. Therefore, for a user it is difficult to find desired videos without a search function within a video. Even when the user found related video data, it is still difficult most of the time for user to decide whether a video is useful by only looking at the title and other global metadata which are often brief and high level. Moreover, the requested information may be covered in only a few minutes, the user only wants to find the piece of information whichever he requires without viewing the complete video. The problem becomes how to retrieve the appropriate information in a large lecture video archive more efficiently. We apply automatic video segmentation and key frame detection to offer a visual guideline to navigate the video content. Simultaneously, we extract text data by applying video Optical Character Recognition (OCR) technology on key-frames and Automatic Speech Recognition (ASR) on lecture audio tracks.

II. RELATED WORK

Retrieval of video using speech and text information which is fetching from the multiple videos and resulting in the creation of clusters. We have to implement a model which captures the various frames from a video. All the captures frames are then classified according to the duplication property. Then we fetch all the text from all the frames for further video retrieval system. Also we fetch all the voice resulting into text using ASR technique is also used in the process of video retrieval system.

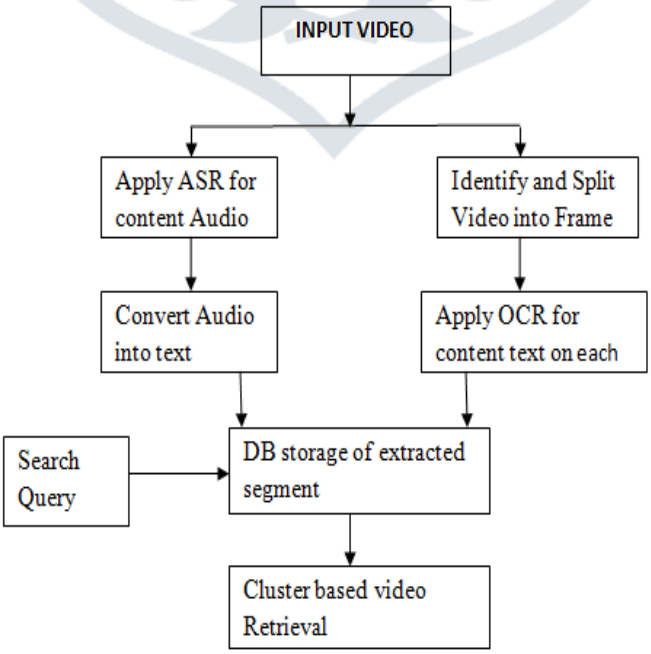

Figure 1: BLOCK DIAGRAM

In this system, Administrator gives the input video to the proposed system. After that apply ASR for Content Audio Retrieval and also identify and splitting the video into frames. After applying the ASR convert audio into text and also extract the segment level keywords and stored into the database. Users search the input query onto the database. If user query found then provide the result as a clustered based video.

III .FLOW OF SYSTEM

In this chapter we will present four processes for retrieving data from video.

3.1 VIDEO SEGMENTATION

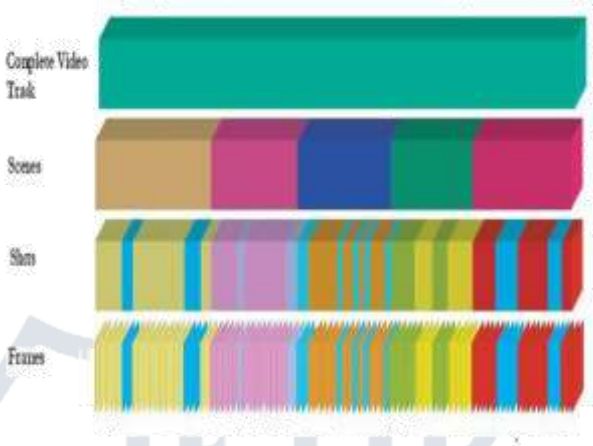

Figure 3: Video segmentation

In our proposed system complete video divided into the number of frames. Segmentation of Video done with the help of step by step process of video segmentation, first of all the complete video is converted into scenes, then those scenes are converted into shots and finally shots are converted into various frames.

In video segmentation, shots boundary detection method is done. Usually, in shot boundary detection method first extracted all visual features from frames. By using these extracted visual features measure similarities between frames. Finally, detect shot boundaries between frames that are dissimilar. Global and local features are used for boundary detection and classification.

3.2 KEY FRAME EXTRACTION

If there are redundancies present among the frames in the same shot, hence certain frames that best reflect the shot contents are selected as key frames. The features used for key frame extraction include colors (particularly the color histogram), edges, shapes, optical flow.

3.3 OCR FOR CHARACTER RECOGNITION

Texts in the video are closely related to the video contents, which provide important information for the Retrieval task. In our system, we developed OCR system for gathering video text.

In our praposed system we can used tesseract-OCR DLL(Dynamic Link Library) file for text recognition from input video.

3.4 ASR FOR AUDIO SPEECH RECOGNITION

In addition to video OCR, ASR can provide speech-to-text information from videos. Most lecture speech recognition systems cannot achieve a sufficient recognition rate.

Steps of ASR Algorithm:

1. Extract wav file from input video.
2. Feature extraction from wav file
3. Convert wav file into text
4. Match converted text with standard dictionary for more accuracy
5. Print resulted text with more accuracy

3.5 CLUSTERING

Video contains huge amounts of data which needs to be organized and compressed in an efficient manner (e.g., one hundred hours of video contains about 10million frame srequiring about TeraBytes of data . Recent work in digital video retrieval has stressed on a hierarchical representation of video for ease of understanding, representing, browsing, and indexing. During the parsing process, video clips are segmented into scenes. Scenes are further segmented into shots which are each represented in terms of a few key frames.

4 CONCLUSIONS

In this paper, we presented an approach for content-based video indexing and retrieval in large video archives. We apply resources for video to extracting content-based metadata automatically. This paper covers the following tasks: Video segmentation including shot boundary detection, key frame ex-traction, scene segmentation and audio segmentation, extraction of features. In our system we developed OCR and ASR for text and audio recognition respectively.

REFERENCES

1] Haojin Yang and Christoph Meinel, Member," Content Based Lecture Video Retrieval Using Speech and Video Text Information", IEEE transactions on learning technologies, vol. 7, no. 2, April-june 2014.

2] C. Meinel, F. Moritz, and M. Siebert, "Community tags in teleteaching environments," in Proc. 2nd Int. Conf. e-Educ., e-Bus., and e-Manage. And E-Learn., 2011.

3] Ground truth data. (2013). [Online].
Available:http://www.yanghaojin.com/research/videoOCR.html.

4] 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.

5] W. Hurst, 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