www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 4 Issue 5 May 2015, Page No. 12163-12165

Content Dependent Video Retrieval System

Aditi P. Sangale[1], Santosh R. Durugkar[2]

[1]Student , Computer Engineering, S.N.D.C.O.E.R.C, Babhulgaon,Yeola
Savitribai Phule University, Pune, Maharashtra, India
sangle.aditi@gmail.com

[2]H.O.D , Computer Engineering, S.N.D.C.O.E.R.C, Babhulgaon, Yeola
Maharashtra, India
santosh.durugkar@gmail.com

Abstract***: Lecture videos are becoming ubiquitous medium for e-learning process. E-lecturing has evolved more competent popular lectures. The extent of lecture video data on the World Wide Web is increasing fastly. Therefore, a most appropriate method for retrieving video within huge lecture video library is required. These videos consist of textual information on slides as well as in presenter's speech. This paper estimates the virtual utility of mechanically recovered text from both of these sources for lecture video retrieval. This approach gives content based video searching method for getting most relevant results. To implement this system, firstly we have to separate out contents on presentation slides and speaker's speech. For mining textual information written on slides we apply optical character recognition algorithm and to translate speaker's speech into text we will apply automatic speech recognition algorithm. Finally, we will store extracted textual results into database against particular timestamp and unique id by performing automatic video indexing. When user will put a search query, then results will be displayed according to video contents. This technique will be beneficial for the user to search a suitable video within a short period of time.***

Keywords: Lecture videos, Text extraction, video indexing

1. Introduction

In today's environment, because of the consistent scene property of the formed video, suitable results cannot be applied to lecture videos based on visual feature Abstraction. Also for flexible interactions, faculties create lecture videos in double scene format, which shows speaker and his presenting slides simultaneously. As presentation method is used for higher level of understandability of students these lecture videos are rapidly used by the students for e-learning. For this reason, numbers of institutions upload their lecture videos on internet. As World Wide Web is producing a large number of videos, so it's a difficult task to search an appropriate video according to search query. Because when user search a lecture video, results are displayed according to title of video not on the basis of contents. In this case, Sometimes it will be possible like, searched information may be covered within few minutes. Thus ,user may want to view this information within a short period of time without going through a entire video. The problem is that one can't retrieve the accurate information in a large lecture video archive more significantly. Almost each video retrieval search engine or search systems such as You-Tube and other reply based on available textual relative data like video title and its description, etc. Generally, this type of metadata has to be created by a human to confirm improved quality, but the step of creation is slightly time and cost consuming.

The objective of the system is to retrieve a video on the basis of its contents rather than retrieving video according to its title and metadata description in order to provide an accurate result for the search query. For this purpose ,we have to implement a model which captures the various frames from a video lecture. Resulted captured frames are then distinguished according to the duplication property. Video fragmentation is done after particular a time interval within two consecutive frames. It may also happen that a video lecture contains one slide presentation for a few more period of time. So to solve this problem maximum time interval is used in seconds for key frames segmentation. We extract all the text from all the frames for further video retrieval system using optical character recognition (OCR) algorithm. Also we translate all the voice resulting into text using ASR technique. This is also used in the process of video retrieval system. The related 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.

These OCR algorithm is responsible for extracting characters from the textual information as well as ASR algorithm is applied to retrieve the speech information from the video lecture. The OCR and ASR transcript as well as detected slide text line types are assumed for keyword extraction, with the help of which video keywords are accessed for browsing and searching content-based video. The proposed system is evaluated on the basis of performance and the usefulness.

2. Literature Survey

Large number of frames is repeated from one shots and scenes. So, extracted frames contain the proper content with decreasing repetition. H.J.Jeong proposed a method for a video segmentation i.e. with the use of SIFT and adaptive threshold. In this process, slides having similar contents but with different background are compared with each other and frame transition is then calculated. OCR is used for retrieving text from the detected frames. To retrieve text from color images Zhan proposes an algorithm in which multiscale wavelets and distributed information is used to locate the text lines. Then

support vector machine classifier was used to gain text from those previously located frames.

For separating and extracting text from complicated background skeleton based binarization method is developed by H.Yang. By performing differential ratio of text and background color automatic video segmentation is performed by Wang et al. In their proposed system with the help of threshold slide transitions are captured. Grcar et al. applies synchronization method among recorded video and slide files which is being provided by presenters [9].The designed system is opposite to this concept which directly analyzes video independent of any hardware. Tuna et al. offered their analysis for lecture video indexing and search by using global frame differencing metrics [11].For improving OCR results they also have proposed one image transformation technique as global differencing metrics appropriate segmentation results when slides are designed with animations. For lecture video dissection Jeong et al. proposed scale invariant feature transform and adaptive threshold algorithm slides containing similar contents [12].

As speech is another high level semantic feature so we have to consider it for video indexing. ASR resulted from audio signal converted into textual information. Approaches described in [5] and [15] have used special commercial software. The authors of [2] and [6] have proposed English speech recognition for Technology Entertainment and Design corpus. In which they have created a a training dictionary manually for comparing results.

3. System Architecture

The proposed system architecture is designed for retrieving lecture videos according to its contents. As we have to provide a significant video retrieval to students for completing this objective we have designed this system. The designed system composed of four modules as shown in the following figure.1 System Architecture.

In addition to this system, we are also providing some extra facility to user that user can find lecture videos by giving search query in three formats. The first approach is regular search i.e. textual query. Other approaches are search query in image format, search query as small video clip and search by audio in search query format.

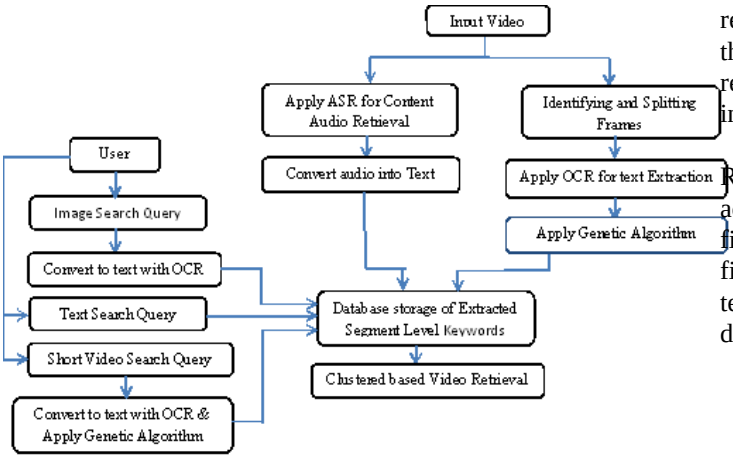

Figure 1. System Architecture

4. Implementation Modules

To implement the proposed system, following modules must be implemented as follows:

4.1 Video Fragmentation

This is the first module of our project in which input video is given and that video is segmented into the number of key frames during certain period of time interval in seconds. In some cases, it may happen that a same presentation slide is displayed for a long period of time then to reduce duplication we will increase the time interval of video segmentation. When this video file is fragmented, it forms number of images and one audio wave file of continuous presentation.

4.2 OCR Module

Optical character recognition module is used to retrieve the text metadata from the extracted key frames of lecture video. When OCR algorithm is applying on the lecture video, it works through certain steps. Tesseract OCR engine is used for translating visual information into text. In this algorithm we have to select video file firstly for processing. In second step it will load images from input path and convert RGB values to gray for every input image file. From gray scale image texts are extracted and those characters are store into database.

As OCR is a very important and primary step so it is mandatory that it must give accurate results. Tesseract OCR results need to pass through genetic algorithm for improving spelling correction. The algorithm uses OCR output as a input character set and that will be compared with standard ELD library. Valid results then continue to display results.

4.3 ASR Module

Automatic speech recognition algorithm extracts speech or voice from lecture video and converts it into textual information and stores it into database. Speech is one of the most important carriers of information in video lectures. Therefore, it is of distinct benefit that this information can be applied for automatic lecture video indexing. Unluckily, most of the existing lecture speech recognition systems in the reviewed work cannot achieve a sufficient recognition result, the Word Error Rates. ASR is aimed to enable computers to recognize speaking voice characters without human intervention.

In the system open source automatic Speech Recognition tool is used which works with steps. Firstly it accepts input audio wave file and then extract sound from input file. After that it will apply speech recognition engine to sound file and set dictation Grammar. The main task is to recognize text from input wave file using speech recognition engine and dictation grammar. At last final results are saved into database.

4.4 Retrieval of Video

After applying the whole procedure of video fragmentation, OCR algorithm, ASR algorithm resulted output information is stored as OCR and ASR results into database. When user will give a search query in form of text that will directly compared with the stored results in database, by matching with the threshold value. If user's query is in image

or small video clip format then both of these are converted into textual query firstly, by applying whole procedure and then will search in database. After searching from database, clusters are formed of whole search results and then results are displayed.

5. Mathematical Model

Set Theory:

Let S be a technique for Retrieval of video from Database.

Such That S= {I, F, O} Where,

I represent the set of inputs:

I= {D, W}

D= Set of Requirements for Retrieval of Video

W= No of Methods for retrieval of Video.

F is the set of functions:

F= {F1, F2, F3}

F1= Apply OCR algorithm and results are stored

F2= Applying ASR and results are stored

F3= Threshold Comparison

O is the set of outputs:

O= {C}

C= Retrieved Video

6. Results and Discussion

The OCR and ASR results need to be accurate so as to increase the performance of the system. The number of matched textual information available from OCR results and Word error rate of ASR results these two factors are used by us for performance evaluation. In proposed system, e-lecturing videos are searched by image or small video clip also. To improve accuracy of OCR results genetic algorithm is used. The result obtained shows that higher accuracy as databases and algorithms used required less computation time

7 . Conclusion

In this project we have designed an algorithm for content based retrieval of video to give more effective and accurate search results to E-learners. The methodology used is more beneficial than the existing one. The main constituents of this process are detected key frames within some time interval and accurate character set given as input to OCR algorithm for text extraction. To exclude spelling correction errors in OCR results a genetic algorithm is used. Both of these algorithms give highly accurate results within less computation time.

The way to future work is to formulate an efficient Clustering algorithm which can retrieve clusters according to different subjects containing number of lecture videos and display results according to relevancy with search query keyword.

References

[1] Haojin Yang, Christoph Meinel and Member IEEE,"Content based Lecture Video Retrieval Using Speech and Text Information", IEEE Transactios on Learning technologies,2014,pp. 142-154.

[2] 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, pp. 232–235.

[3] 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

[4] Haubold and J. R. Kender, "Augmented segmentation and visualization for presentation videos," in Proc. 13th Annu. ACM Int. Conf. Multimedia, 2005, pp. 51–60.

[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 an what to do when it isn't," in Proc. 8th Int. Conf. Multimoda Interfaces,2006.

[7] T.-C. Pong, F. Wang, and C.-W. Ngo, "Structuring low-quality Video taped lectures for cross-reference browsing by video text analysis," J. Pattern Recog., vol. 41, no. 10, pp. 3257–3269, 2008.

[8] J. Adcock, M. Cooper, L. Denoue, and H. Pirsiavash, "Talkminer:A lecture webcast search engine," in Proc. ACM Int. Conf Multimedia, 2010, pp. 241–250.

[9] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," in Proc. Int. Conf. Comput. Vis. Pattern Recog., 2010, pp. 2963–2970.

[10] 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.

[11] T. Tuna, J. Subhlok, L. Barker, V. Varghese, O. Johnson, and S. Shah. (2012), "Development and evaluation of indexed captioned searchable videos for stem coursework," in Proc. 43rd ACM Tech. Symp. Comput. Sci. Educ., pp. 129–134.

[12] H. J. Jeong, T.-E. Kim, and M. H. Kim.(2012), "An accurate lecture video segmentation method by using sift and adaptive threshold," in Proc. 10th Int. Conf. Advances Mobile Comput., pp. 285–288.

[13] G. Salton and C. Buckley, "Term-weighting approaches In automatic text retrieval," Inf. Process. Manage., vol. 24, pp. 513–523, 1988.

[14] H. Yang, B. Quehl, and H. Sack. (2012), "A framework for improved video text detection and recognition," Multimedia ToolsAppl., pp. 1–29, [Online]. Available: http://dx.doi.org/10.1007/s11042-012-1250-6.

[15] 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. 145–156, Jul. 2008.

[16] 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, pp. 9–12.

[17] Kruatrachue, B. Somguntar, K Sirbioon,K. "Thai OCR Error Correction Using Genetic Algorithm", IEEE 2002