Volume 4, Issue 11, November 2014 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

A Review on Circumscribe Based Video Retrieval

Aditi P. Sangale*
Computer Engg Dept, S.N.D.C.O.E.R.C
Yeola, Dist-Nasik, Savitribai Phule University of Pune
Pune, Maharashtra, India Pune, Maharashtra, India

Prof. S. R. Durugkar
Head, Computer Engg Dept, S.N.D.C.O.E.R.C
Yeola, Dist-Nasik, Savitribai Phule University of Pune
Pune, Maharashtra, India Pune, Maharashtra, India

Abstract— E-lecturing system is become more and more popular from last some years .World Wide Web (WWW) is speedily introducing huge amount of lecture videos. So, to archive lecture videos urgently more efficient and effective method is required. This paper presents an approach for video indexing and content based video search. Firstly, we apply video segmentation and key frame detection on a large lecture videos. As text is a high level of metadata, to extract textual metadata from key frame segments optical character recognition (OCR) algorithm is used while to extract audio signal automatic speech recognition technology (ASR) is used. Finally, we have to store OCR and ASR results combinely into a database with a particular timestamp and according to the user's query .The performance & efficiency of a system are improved by applying different effective techniques.

Keywords— Lecture videos, Video Segmentation, Content based video search, Key frames, Video indexing

I. INTRODUCTION

With rapid development of E-learning system through World Wide Web (WWW), large numbers of students are taking an advantage of this system. As most of the institutions and colleges upload and publish their video lectures on the internet to make available a study material to the students. Any student can access this video lectures at anytime and anywhere independent of time and location. This multimedia information is very useful to understand any concept completely in effective manner. to retrieve multimedia information one has to describe, store and organize it in proper manner also it should be indexed in appropriate way so as user can get access in efficiently and quickly. Runtime video is an important kind of multimedia information. Videos have some properties as follows: 1) Qualitative contents are available than images. 2) Huge amount of raw data. Because of these characteristics video retrieval is somehow difficult. In previous years , stored video databases are contains small amount of videos so the video can be searched manually but nowadays a large number of videos are available related to a single concept so to index and store videos properly is a need for efficient and quick access of lecture video. Video indexing is the process of numbering videos according to the priorities as per search query for giving results to the user within a short period of time.

Generally, most of the search system search video according to their metadata description i.e. title of video, author and other related information. But, sometimes it happens that a large video contains a short amount of information that user want to study, at that time, if retrieval system have displayed such a search results then user have to go through a complete video for studying that concept but resulting inefficient and inappropriate output. It's quite hard to search the video on the Basis of its Content and Voice information. In general whenever user type any query for searching any video, most probable all the videos or results are according to the title of the video. Means it uses the name of video for comparing with the search text. This whole process doesn't give accurate results. As results are displayed according to the title of the video so it is inconvenient for the user to search and go through the whole video and study that video. For example, if user wants to study conditional statements in C language then it will search video according to name of video but it is inconvenient for the user to study that video if it contains only short information about user's query. So, it gives inappropriate results. For quick access browsing content based video indexing and retrieval is preferred

II. RELATED WORK

In multimedia domain, information retrieval is a very interested area for research. In case of content based video lecture retrieval text and speech are two important semantic features. In traditional systems; video can be recorded because of continuous scene compositions related to lecture video frames. At that time a single video camera captures a full video lecture shooting combining slide presentation as well as speakers lecture video. In this process a varying factors may affect and degrades the video quality. Recently, lecture videos are produced by using multi scenes format, in which speaker's lecture and his slides presentation are captured simultaneously by using TASK technique. In this system, main part of the video lecture i.e. speaker's voice captured by video camera and his slide presentation is recorded with the frame grabber tool. In this case, no any extra synchronization is required as both parts are going to capture at same time. To retrieve video lectures effectively, one have to store those videos with proper indexing. Tuna et al. presented their view for lecture video indexing and search by using global frame differencing metrics. Jeong et al. developed lecture video segmentation method using Scale Invariant Feature Transform (SIFT) feature and the adaptive threshold feature.

III. CONTENT BASED VIDEO RETRIEVAL

To reduce time complexity at the time of video retrieval when it searches a video based on its metadata description i.e. Title of video. In this approach, to retrieve video according to its contents then firstly we have to consider its two main semantic features like visual screen (Text) and Audio tracks (Speech). So, for this system, one has to open up contents of the lecture video and collect a large amount of metadata. To retrieve and index these videos firstly we have to segment a large video into short key frames by considering time interval. Most of the times it happens that a same slide is displayed for a long period so to avoid duplication a selected time interval is short. As text is a high level semantic feature which is most important for the content based video retrieval system. For character recognition OCR i.e. optical character recognition algorithm is used while to synchronize another high level semantic feature like speech ASR i.e. Automatic speech recognition technique is used.

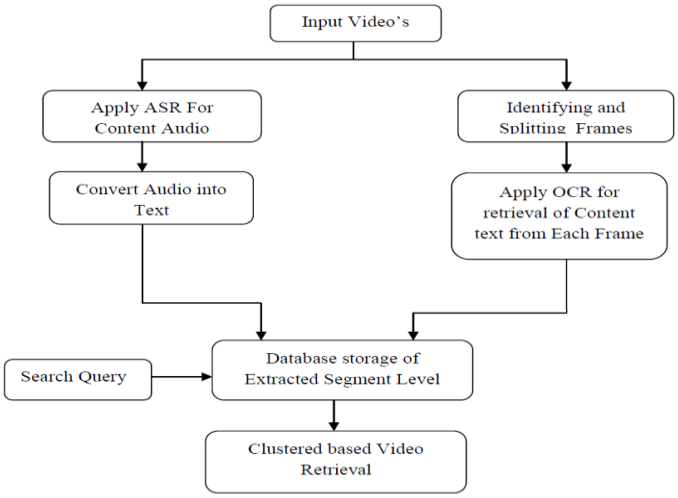

Fig. 1 System Architecture

As shown in fig. 3.1 whole process is categorized into the four steps as follows

A. VIDEO SEGMENTATION

Content based video search starts from the process of video segmentation in which one have to classify a moving objects of a lecture video sequence. This process initially segments the first image frame and then it continuously tracks moving objects, segment it and store it subsequently in the image frames. Key-frames are nothing more but images extracted from lecture video data which represents the contents of the video data in abstract form. If key frames have extracted properly, then those are very effective for showing visual abstract of video contents and also for fast video browsing it is useful. To perform segmentation many approaches are available, many of these uses the concept of global pixel level differencing metrics. But the drawback of this method is salt and pepper noise degrades the accuracy of results.

While observing contents of slides it is realized that most of the contents are tables, figures and those can be considered as a connected components. So, connected components analysis rejects the salt and pepper noise in video segmentation and improves an accuracy of segmentation results. In our approach, segmentation consists of two steps:

In first step, a complete video is analysed and it captures every adjacent frames. Analysis time interval selected is of three seconds by considering efficiency and accuracy of video segmentation results. As minimum time interval for the analysis is set for tree seconds means frames whose duration is smaller than three seconds are discarded in our system. After tis canny edge maps are created and differential binary images are created from those edge maps. Connected component analysis is performed on those differential binary images and for the video segmentation number of CCs is used as a threshold. In this process results are too redundant so second segmentation step is performed.

In the second step, actual slide transition has captured. Then title region and content region of the slide is first analysed. By analysing a large amount of lecture video from database content distribution of frequently used slide styles s is selected. By applying second segmentation steps redundancy is reduced

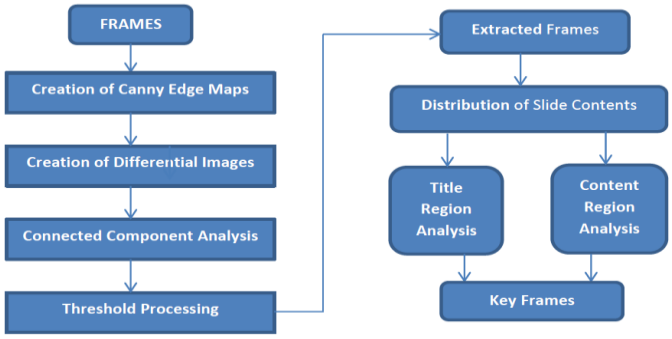

Fig. 2 Video Segmentation

B. VIDEO OCR FOR CHARACTERS

Text is high level semantic feature of the content based video retrieval. An OCR is a system which can load an image, performs pre-processing on that image, performs feature extraction on image, and calculates the "distances" among the extracted image features and predefined known feature vectors stored in the image model library. According to degree of similarity between loaded image and image model it recognizes image.

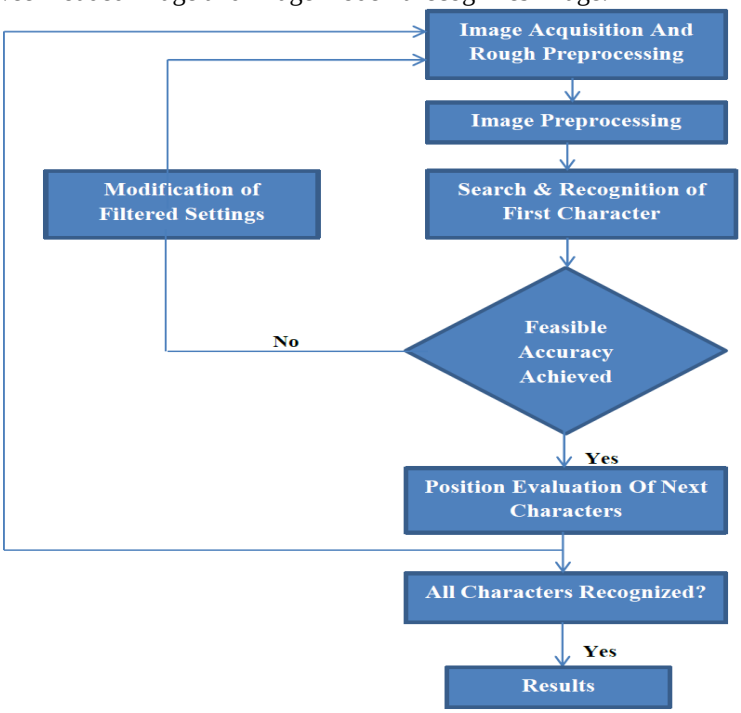

Fig 3 Common algorithmic steps for OCR

In the pre-processing step, images make appropriate for applying feature extraction algorithms. Some feature extraction algorithms calculate every pixel of image. Pre-processing performs multiple operations like thresholding, binarizing, filtering, edge detection, segmentation and also makes images suitable for computation. In the next step of classification, each loaded character image is assigned to one of the possible image model by which corresponds to the matching stage of object recognition system. The final result is calculated on the basis of similarity measuring metrics. The following figure sows the common optical character recognition algorithmic steps containing the flow of steps in the form of flowchart. It contains steps like pre-processing, recognition, comparison, spatial adjustment, feature extraction etc.

In our approach, we developed a new localization verification Scheme for text detection. In the detection stage, to quickly localize candidate text region having low rejection rate an edge-based multi-scale text detector is used. Sometimes, for the subsequent text area verification, an image entropy-based adaptive refinement algorithm is used. Ten to remove the separated non text region blocks stroke width transform based verification procedure is used. Also to identify special non text region patterns correctly Support Vector Machine classifier SVM is used to improve pattern detection accuracy.

C. VIDEO ASR FOR AUDIO

Automatic Speech Recognition ASR is technology which allows a computer to identify the words spoken by a person and then convert it to textual data. The ultimate goal of ASR is to allow a computer to recognize that speech in real-time, with full accuracy, all words that are spoken by an individual, without any affection of vocabulary size, noise, speaker characteristics or accent. The objective of an ASR system is to convert a speech signal into a text message transcription accurately and efficiently without considering speaker's environment or the device used for recording the speech. This process starts with speaker's decision about what to say and actually speaks a sentence. In speech, it may contain a sequence of words with some pauses. As speech signal is giving as an input to the ASR software, it produces a speech wave form. After that software decodes speech signal into the sentence. First it converts the audio signal into a sequence of vectors. After that it generates a valid sequence of representations by using a syntactic decoder.

In our approach, recorded audio file is segmented firstly into smaller pieces and inappropriate segments are separated out. A list of all used words in transcript file is created in the intermediate step. For obtaining phonetic dictionary, pronunciation of every word must be represented phonetically. Manually collected speech and segments is time consuming.

D. BROWSING VIDEO SEARCH

For Searching content based lecture video according to the users query we have to classify a video into key frames. After applying OCR and ASR algorithms on those key frames OCR and ASR results are stored into database

with a unique identifier with a particular timestamp. When user want to search a video, user will fire a search query ten according to that query OCR and ASR results are searched if the contents are found then video search is successful.

IV. CONCLUSION

In this paper we have presented a content based video retrieval and indexing of lecture video. In order to perform this we apply optical character recognition on visual screen and automatic speech recognition on audio tracks for extracting content based video search. To improve the quality i.e. accuracy and efficiency of OCR and ASR results different effective algorithms and techniques have explained in the paper. In a large lecture videos novel indexing features have been developed by using those metadata and a user study has been conducted. To reduce the solidity and consistency problems different algorithmic techniques are used.

REFERENCES

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

[2] 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.

[3] "*Automatic Speech Recognition – A Brief History of the Technology Development" B.H. Juang# & Lawrence R. Rabiner* # Georgia Institute of Technology, Atlanta * Rutgers University and the University of California, Santa Barbara*

[4] International Journal of UbiComp (IJU), Vol.3, No.2, April 2012 DOI:10.5121/iju.2012.3202 13 "*Content Based Video Retrieval Systems*" B V Patel1 and B B Meshram2, Chembur, Mumbai, INDIA

[5] "*Optical Character Recognition Using Optimisation Algorithms*" Kirill Safronov Institute for Process Control and Robotics (IPR) University of Karlsruhe Karlsruhe, Germany

[6] "*Automated Content Based Video Retrieval* "Sr.an Zagorac, Ainhoa Llorente, Suzanne Little, Haiming Liu, Stefan RÄuger Knowledge Media Institute, the Open University Walton Hall, Milton Keynes, MK7 6AA, UK

[7] IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews, Vol. 41, No. 6, November 2011 797 "*A Survey On Visual Content-Based Video Indexing And Retrieval*" Weiming Hu, *Senior Member, Ieee*, Nianhua Xie, Li Li, Xianglin Zeng, And Stephen Maybank

[8] IEEE Transactions On Systems, Man, And Cybernetics, Vol. Unknown, No. Unknown, Unknown 2007 1 "*Automatic Video Classification: A Survey Of The Literature*" Darin Brezeale And Diane J. Cook, *Senior Member, Ieee*

[9] E. Leeuwis, M. Federico, and M. Cettolo, "*Language modelling and transcription of the ted corpus lectures*," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., 2003, pp. 232–235.

[10] 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.

[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.

