IJARCCE

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

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 12, December 2015

Multimedia based information retrieval approach for lecture video indexing based on video segmentation and Optical Character Recognition

Surabhi Pagar[1], Gorakshanath Gagare[2]

M.E. Student, Department of Computer Engineering, SVIT, Nasik, India [1]

Assistant Professor, Department of Computer Engineering, SVIT, Nasik, India [2]

Abstract: Recently advanced development in multimedia technology allows the capturing and storing of video data with highly expensive computers. Further, the new opportunities offered by the information technology have made a vast collection of video data publicly available. But still, without proper search techniques these all data are hardly usable. Users are not well satisfied with the video retrieval systems which provide analogue VCR functionality. For example, a user analyses a football video will ask for specific events likewise goals. Content-based search and retrieval of video data becomes a challenging and important problem now a day. So, the needs for tools that can be used to manipulate the video content as in traditional databases manage numeric and text content is significant. Therefore, a more efficient way towards video retrieval in World Wide Web or within large lecture video archives is urgently required. This research presents an approach for automated video content retrieving over large lecture video archives. First of all, It apply automatic video segmentation and key-frame detection to offer a visual guideline for the video content navigation. Subsequently, It extract textual metadata by applying video Optical Character Recognition (OCR) technology algorithm on key-frames and Automatic Speech Recognition on lecture audio tracks content of the video. Proposed algorithm Multimedia question information retrieval is to provide a multimedia data such as image and video for extracted words in OCR method. The multimedia search diversification method helps to collect the appropriate answer based on words. It provides relevant information i.e. text, audio, video to the user for more effectiveness.

Keywords: Questing Answering, multimedia search, reranking, video segmentation, video browsing, video retrieval, video structure analysis, OCR, API.

I. INTRODUCTION

Digital video has become a very popular storage and exchange way of medium as there is rapid development in recording technology and so much improved video compression techniques and high-speed networks in the last few years. That is why audiovisual recordings are used more and more widely in e-lecturing systems. There are number of universities and research institutions which are taking this useful opportunity to record their lectures and publish them online for students to access independent of time and location. So that they can study from anywhere and anytime. Because of this reason there has been a huge increase in the amount of multimedia data on the World Wide Web. Due to this for a user it is almost impossible to find desired videos without having a search function in a video archive. Even when the user has found related video data, it is still difficult most of the time to compare whether a video is useful or not just by only looking at the title and other global metadata provided which are often brief and a high level document.

Moreover, the requested information may be covered in only a few minutes, the user generally might thus want to get the piece of information he requires without viewing at the complete video. The problem then arises like how to retrieve the appropriate information in a large lecture video archive more efficiently. Most of the video retrieval and video search systems such as You- Tube, Bing and Vimeo reply based on available textual metadata such as title, genre, person, and detail description. Generally, this kind of metadata has to be created by a human to ensure a high quality, but the creation step is rather most time as well as cost consuming. So the proposed system is an Multimedia based information retrieval approach for lecture video indexing based on automated video segmentation and OCR analysis.

The system consists of six components like Video segmentation, audio to text conversion, keyword extraction, query analysis for multimedia search, answer pattern selection and multimedia data selection and presentation. The system automatically determines keywords in video and search multimedia information should be added for textual answer by collecting data from web to enrich the information [1].

A. Objectives:

1) To propose Model for automated multimedia information retrieval from video keywords.
2) To extract audio from video.
3) To convert video to images for OCR algorithm.
4) To extract keyword from images using OCR algorithm.
5) To extract keywords from video and audio by applying appropriate OCR algorithm.

IJARCCE

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

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 12, December 2015

II. LITERATURE SURVEY

In the last decade e-lecturing has become more and more popular. The amount of lecture video data on the World Wide Web(WWW) is growing rapidly. Therefore, a more efficient method for video retrieval in WWW or within large lecture video archives is urgently needed. This paper presents an approach for automated video indexing and video search in large lecture video archives. First of all, they apply automatic video segmentation and key frame detection to offer a visual guideline for the video content navigation. Subsequently, extract textual metadata by applying video Optical Character Recognition (OCR) technology on key-frames and Automatic Speech Recognition (ASR) on lecture audio tracks. The OCR and ASR transcript as well as detected 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. The performance and the effectiveness of proposed indexing functionalities is proven by evaluation [1].Authors presented a new system for automatic transcription of lectures. The system combines a number of novel features, including deep neural network acoustic models using multi-level adaptive networks to incorporate out-of-domain information, and factored recurrent neural network language models. It demonstrates that the system achieves large improvements on the TED lecture transcription task from the 2012 IWSLT evaluation. results are currently the best reported on this task, showing an relative WER reduction of more than 16 percent compared to the closest competing system from the evaluation [2].Authors introduced a Korean spoken document retrieval system for lecture search. It automatically build a general inverted index table from spoken document transcriptions, and extract additional information from textbooks or slide notes related to the lecture. It integrates these two sources for a search process. The speech corpus used in our system is from a high school mathematics lecture videos. Experimental results showed that the contents information is slightly beneficial for the lecture spoken document retrieval [3].Recording lectures and putting them on the Web for access by students has become a general trend at various universities. To take full gain of the knowledge data base that is built by these documents elaborate search functionality has to be provided that goes beyond search on meta-data level but performs a detailed analysis of the corresponding multimedia documents. In this paper, presented some experiments towards setting up a Web-based search engine for audio recordings of presentations. Authors evaluate standard, state-of-the-art speech recognition software as well as achievable retrieval performance. In addition, They compare the speech retrieval results with a traditional, text-based approach for searching to evaluate the value of speech processing for lecture retrieval[4].There is also investigation of methods of segmenting, visualizing, and indexing presentation videos by separately considering audio and visual data. The audio track is segmented by speaker, and augmented with key phrases which are extracted using an Automatic Speech Recognizer (ASR). The video track is segmented by visual dissimilarities and augmented by representative key frames. An interactive user interface combines a visual representation of audio, video, text, and key frames, and allows the user to navigate a presentation video. Authors also explore clustering and labeling of speaker data and present preliminary results[5].James Glass, Timothy J. Hazen, Lee Hetherington, and Chao Wang, presented in paper a report on their recent efforts to collect a corpus of spoken lecture material that will enable research directed towards fast, accurate, and easy access to lecture content. Thus far, collected a corpus of 270 hours of speech from a variety of undergraduate courses and seminars. Authors report on an initial analysis of the spontaneous speech phenomena present in these data and the vocabulary usage patterns across three courses. Finally, examine language model perplexities trained from written and spoken materials, and describe an initial recognition experiment on one course [6].

PROPOSED SYSTEM:

To propose a Multimedia based information retrieval approach for lecture video indexing based on automated video segmentation and OCR analysis. In contrast with existing system the Propose system automatically determines keywords in video and search multimedia information should be added for textual answer by collecting data from web to enrich the information.

A. Architecture Of Proposed System

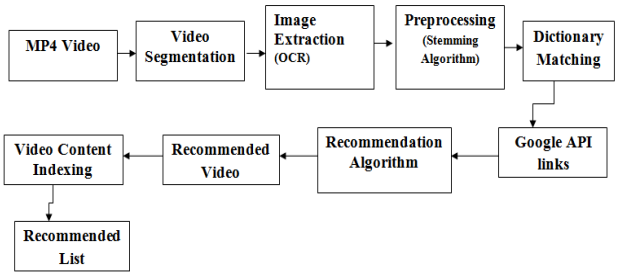

Fig. 1 Proposed System

The system model comprises of distinct entities:

1) **Slide Video Segmentation**: Video browsing can be achieved by segmenting video into representative key frames. The selected key frames can provide a visual guideline for navigation in the lecture video portal. Moreover, video segmentation and key-frame selection is also often adopted as a pre-processing for other analysis tasks such as video OCR, visual concept detection, etc. In the next step, the entire slide video is analyzed. It will try to capture every knowledge change between adjacent frames, for which it established an analysis interval of three seconds by taking both accuracy and efficiency into account. This means that segments with duration smaller than three seconds may be discarded in this system. Since there are very few topic segments shorter than three seconds, this setting is therefore not critical. In next step each slides stores as image in the backend database for OCR algorithm processing.

2) **OCR Algorithm**: Texts in the lecture slides are closely related to the lecture content can thus provide important

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

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 12, December 2015

information for the retrieval task. In the proposed framework, developed a novel video OCR system for gathering video text. Propose system uses OCR algorithm to extract text from images and store extracted text in database.

3) **Information Pre-Processing**: From extracted text irrelevant keywords are remove by comparing predefine keyword list and stop words also remove by using Stemming algorithm. These extracted keywords are used for multimedia query search process.

- Multimedia Query Search
- Query Analysis

Query analysis helps to find the informative keyword for searching corresponding media data using multimedia search engines. The main objective of this process is to find the stem word which is considered as the informative keyword. Here system using an algorithm called stemming algorithm. Stemming algorithm is generally used to remove the stop words which can be applied as follows
– The first step is to consider the given query and initialize the empty variable of string data type
– Split the query based on the space between them and pass them into array list of string type.
– Initialize for loop and remove the stop words i.e., a, and, an, in, be, for and soon by passing the words in the array list .continue the process until length of the array list.
– Pass the remaining words into empty string variable initialized in step 1.
– Finally use the obtained words as informative keyword for search and vertically collect the media data.

4) **Recommendation and Indexing**: The next component of our model is Recommendation and Indexing. In this module the given question is judged whether it requires any media data or it requires only textual answer. Here system will categorize mainly into four types such as text, text and image, text and video, text and image and video based on the given query. For those question which require media data such images and videos. For this system will vertically collect media data by using multimedia search engines such as YouTube for videos and Google images API (Application programming interface).

B. **Algorithms**
The first one is Porter Stemmer algorithm which Conflates inflected/derived words to a stem (root).Input is Training set T with n words and second is Optical Character Recognition algorithm. Input is segmented video key frames.

1) **The Porter Stemmer Algorithm**:

1. Gets rid of plurals and -ed or -ing suffixes.
2. Turns terminal y to i when there is another vowel in the stem.
3. Maps double suffixes to single ones: -ization, -ational, etc.
4. Deals with suffixes, -full, -ness etc.
5. Takes off -ant, -ence, etc.
6. Removes a final -e.

2) **Optical Character Recognition Algorithm:**

1. Initialize application.
2. Fetch handwritten Character.
3. Crop fetched image.
4. Downsample cropped image.
5. Check if network is trained if yes then check whether data is empty or not.
6. Data is not empty then recognize character.
7. Display recognized character

III.CONCLUSION

A more efficient method for video retrieval in WWW or within large lecture video archives is urgently needed. This research presents an approach for automated video content retrieving large lecture video archives. First of all, we apply automatic video segmentation and key-frame detection to offer a visual guideline for the video content navigation. Subsequently, we extract textual metadata by applying video Optical Character Recognition (OCR) technology on key-frames and Automatic Speech Recognition on lecture audio tracks of the video. Propose algorithm is to provide a multimedia data such as image and video for extracted words in OCR method. The multimedia search diversification method helps to collect the appropriate answer based on words. It provides relevant information to the user for more effectiveness.

REFERENCES

[1] Haojin Yang and Christoph Meinel," Content Based Lecture Video Retrieval Using Speech and Video Text Information" ,IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES, VOL.7,NO.2,APRIL-JUNE.
[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. 232235.
[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] W. Heurst, T. Kreuzer, and M. Wiesenhutter, "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. 135143.
[5] A. Haubold and J. R. Kender," Augmented segmentation and visualization for presentation videos," in Proc. 13th Annu. ACM Int. Conf. Multimedia, 2005, pp. 5160.
[6] James Glass, Timothy J. Hazen, Lee Hetherington, and Chao Wang,"Analysis and Processing of Lecture Audio Data: Preliminary Investigations",in Proc. HLT-NAACL Workshop Interdisciplinary Approaches Speech Indexing Retrieval, 2004, pp. 912.

DOI 10.17148/IJARCCE.2015.41273