

PACIFY BASED VIDEO RETRIEVAL SYSTEM

Gaikwad Harshada¹, Sonawane Jayshri², Gore Kapil³, Avhad Madhuri⁴

^{1,2,3,4}*B.E Scholar, Department Of Computer Science, S.N.D College Of Engineering And Research Center, Yeola.*

University Of Pune, Maharashtra, India

gaikwadharshda210@gmail.com

sonawanejayshri1995@gmail.com

kapilgore9194@gmail.com

madhuriavhad18@rediffmail.com

Guided By:-

Prof: Chandgude A.S.

Department Of Computer Science, S.N.D College Of Engineering And Research Center, Yeola. University Of Pune, Maharashtra, India

Abstract

Recently, Video is becoming a catholic medium for e-learning. As per the popularity of online video information over the World Wide Web (WWW) is mostly dependent on user-assigned tags or specification, which is the system by which we can access such videos. However, this system have limitations for retrieval and frequently we want access to the content (pacify) of the video itself is directly matched against a user's query except manually assigned tags or specifications. In e-lecturing videos it contains visual and aural medium: slides of presentation and speech. in this system, we are going to retrieve the text from the videos automatically. To abstract visible information, we apply video content analysis to detect slides and optical character recognition to obtain their text. We abstract textual metadata by applying video Optical Character Recognition (OCR) technology on key-frames and Automatic Speech Recognition (ASR) on lecture audio. The ASR and OCR translate and discover slide text line types are accept for keywords abstraction, in which video and fragment-level keywords are abstracted for video searching on the basis of contents.

.Key Words: Video fragmentation, Frame Abstraction, video indexing, and etc.

1. INTRODUCTION

E-learning is famous or leading used today by peoples in different fields like learning or research such as, for presenting the explanation of practical in medical field, for learning new technologies, to gives training for employees located in various sectors of the multinational companies etc. So the use of video lectures is increasing in recent years and most of the institutes and universities are uploading their video lectures on the World Wide Web (WWW). Therefore this information stored in large extent on internet. The available search engine is search the video on the basis of available textual metadata such as title of that video, writer etc. From this multimedia, data to gathered or hold the required information is a difficult task without the appropriate retrieval system. also this data may not get conveniently because of limited metadata. Therefore, this searching techniques are time wasting belong to manual and also it is inconvenient.

So, we present advanced automatic video indexing and video search from large lecture video storage. We choose outline of lecture from imaged resources and construct videos depend on contents. For this intention, we have to create a system which captures the separate frames from a video. Generated captured frames are then differentiating

according to the duplicated property. Video splitting is done after appropriate time interval between two successive frames. It may appear that a video lecture contains one slide presentation for a maximum period of time. So that to solve this issue minimum time interval is used in seconds for key-frames fragmentation. We select all the text from all the frames for new video retrieval system applying optical character recognition (OCR) algorithm And also we converts voice or audio resulting into text using Automatic Speech Recognition (ASR) algorithm.

The associated information (Text and Voice from Video) is used for pacify (content) based video retrieval system and clustering of video belongs to their text and audio contents. These OCR technique is responsible for abstracting characters from the video data as well as ASR technique is used to fetch the speech information from the video .The OCR and ASR techniques translate as well as detect slide text line types are assumed for keyword abstraction, with the help of which video keywords are accessed for browsing and searching content-based video.

In this survey paper we proposed the system which is examined on the basis of performance and the usability. we advocate a different keyword categorization system for

different structure of information resources. In order to check the usability, we tried this technique in a huge lecture video portal

2. LITERATURE SURVEY

Madhav Gitte, Harshal Bawaskar, Sourabh Sethi, Ajinkya Shinde [1] establishes a content based video retrieval system which used for multimodal characteristics to retrieve the videos from multimedia data warehouses. The system is not so complicated but which retrieves the information based on only feature extraction algorithm and then clustering method is applied it is giving the efficient result in case of video. Erwin Leeuwisl, Murcello Federico and Mauro Cettolo [2] are concentrating on English speech recognition in Tran's language English database (TED) corpus. He establishes the language model by lecture translation. Wolfgang Hurst, Thorsten Kreuzer, Marc Wiesenhütter [3] proposed a method which retrieves audio information from video by creating the discrete ASR technique. Large vocabulary automatic speech recognition (LVASR) is used which hold all type of voice signal such as low quality signal, noisy signal, fast speaker word determine which improves word accuracy than other methods for audio retrieval. Xiaoqiang Xiao, Jasha Droppo and Alex Acero [4] have advocated the new technology to be use with the ASR which is more effective compared to conventional HMM based ASR system. The author has established an IR system which determines the sub words units into the valid word. Matthew Cooper [5] has analyzed the results from both the OCR and ASR algorithm. He has presents that there is mega aberration in texts retrieved from ASR and OCR method. Also he has authenticate that the OCR method is gives more correct results for text abstraction than the ASR method. The reason is that ASR has many sources of error like pronunciation of the speaker, phonetic errors etc. and the errors take place in this will precisely affect the retrieval performance which not occur in OCR.

Alexander G. Hauptmann, Rong Jin, and Tobun D. Ng [6] present the distinct video retrieval method for the multimedia video data, produce and bind all the metadata. The author had also proves the results for retrieval by using all technique and distinguished it. The author has shows that the result is more effective when retrieval is done by using improve OCR technique with preprocessing is done efficiently, the probabilistic system is used and if the speech recognition technique is also used with it. Tiecheng Liu and John R. Kender [7] had established a new software tool for the content retrieval of video which is based on its low level features. The model especially works on real time content retrieval and key frame retrieval from videos which have handwritten slide scene type. The author has used a rule based model for tracking the border between frames. As the handwritten slides are there for retrieval so it has a consequential bundle of frames to be created for every event which has more suitable in content. So it becomes tough to recognize the borders or shots between frames rather than to find in other scene type like previously prepared slides or notes. But the retrieval is only done with the content and no further processing is done for convenient of the search. The

ASR technique provides the information for transforming the audio information into text forms. The module is depending on efficient information retrieval method used to retrieve and forming the securing of text information. In our survey paper, we are first of all doing the fragmentation of frames as a preprocessing work and then placed OCR and ASR technique for creating the metadata.

3. EXISTING SYSTEM

Prefer to existing system, it was a difficult thing to provide the users with accurate results in a minimum time. Mostly, when user requests any video, the results came in the output will have associated similar names of video for which it was searched. Generally it works on names of files stored in the database with necessary data and likewise results are fetched and displayed. But in this assurance of accuracy is not yet provided. In this system, results are appearing as per the title of the video hence it becomes impossible to view each and every video having similar type of titles. For example, if user wants a video for study java programming then it will search video according to the name of video but it is unsuitable way to study the video if it contains only short information as per user's request. Hence, it may give unnecessary and random results.

In this system video retrieval depend on visual characteristics abstraction which is difficult to concern lecture recording due to the identical format of lecture videos. For example, the quality of video may depend on many factors like brightness of light, motion of camera or person hence a single shot can be record number of times.

4. PROPOSED SYSTEM

In this system, we proposed the architecture is to achieve a video depends on its contents instead of retrieving video depends on its title and metadata specification. So that it will give an appropriate result for the user's requirement. Because of this, we have to prepare a model which catches the different fragment from a video lecture. All the catch fragments then categorized according to the replication property. Video fragmentation is performed by considering a time interspace in between two frames. In some conditions it occurs that a video lecture provide single slide show for a large time span then maximum time span is presented in few seconds. Simultaneously, key frames fragmentation occurs in which we retrieve all the text from all the fragments for some time interval video fetch system using optical character recognition (OCR) techniques. Also we abstract text from voice of video using ASR technique. All these things are used for pacify based video retrieval system and clustering of video depends on their textual and audio factors.

The OCR technique is used to abstract the textual information from the literal information as well as ASR technique is used to abstract the speech information from the video lecture. Also Software for dictionary also for finding words from voice collected by ASR. The OCR technique is used to abstract the textual information from the literal

information as well as ASR technique is used to abstract the speech information from the video lecture. The OCR and ASR noted as well as found text line types from slides are designed for abstracting some necessary words, by which video as well as fragment-level keywords are abstracted corresponding to the content video browsing. The characteristic of indexing and clustering provides a more ease to use this system for video browsing.

5. SYSTEM ARCHITECTURE

This system architecture is shows for retrieving lecture videos corresponding to its contents. As we have to afford

an important video retrieval to all users for achieving their need, we have created this system. The system having four module as shown in the following figure System Architecture. In addition to this system, we are also accommodate some additional services to user that user can find lecture videos by offering search query in three formats. The first format is regular search i.e. textual query. Another formats are search query in image format, search query as small video clip and search by audio in search query format.

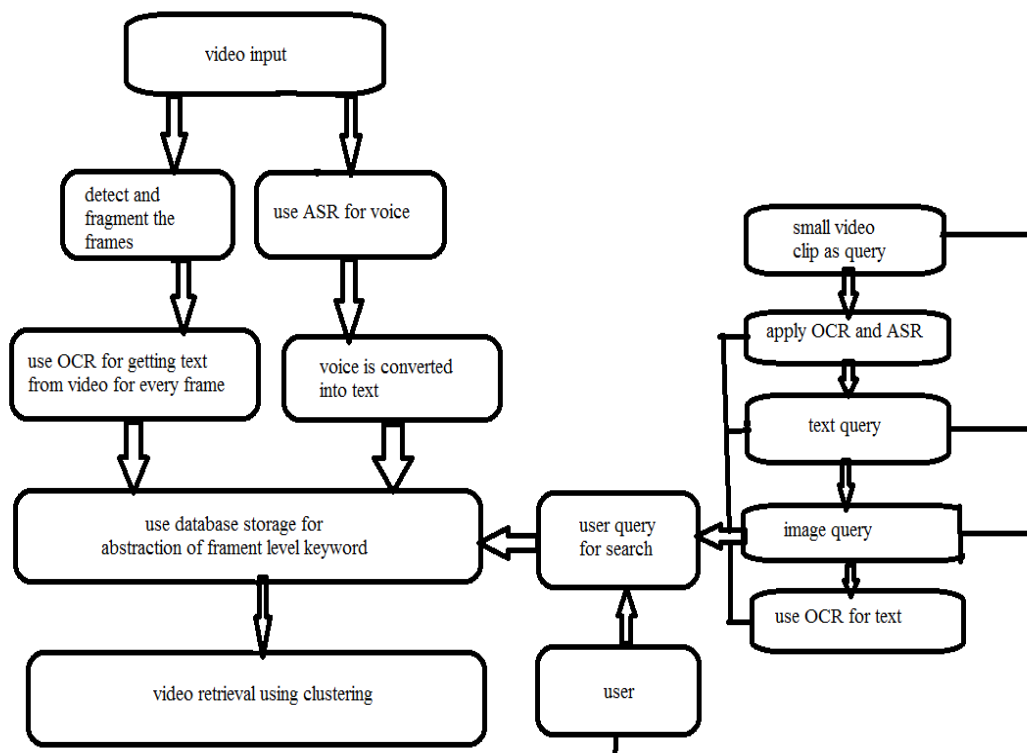


Fig -1: System architecture

6. MODULES FOR IMPLEMENTATION

6.1 Frame Abstraction-

This is the first module of our system in which input video is given and that video is separated into the number of key frames during particular time of interval in seconds. Sometimes, it appear that a same key frame is viewing for a large period of time then to minimize duplication we will maximize the time interval of video fragmentation.

6.2 OCR Module-

Optical character recognition module is used to extract the text metadata from the extracted key frames of lecture. When OCR technique is applying on the lecture video it has to adopt some steps. At first the raw signal from the combined video signal is sampled to generate a binary

image, from which the subtitles can be extracted. When using the signal from the camera it is also need with some spatial adaption to ensure that the lines of text are horizontal in the image, which is a precondition for our OCR technique. Next the binary image is pre filtered to remove noise and enhance characteristic features. After the filtering the image can be split into individual sentences, arguments and characters. Every feature and characteristics whether it is semantic they are spotted and matched with an earlier existing database. After selecting the most likely letters in a given word, the word can be analyze with a dictionary lookup, to confirm if the letter mixture is likely.

6.3 ASR Module

Automatic speech recognition retrieves speech or audio from video and transforms it into textual data and stored it into database. Speech is one of the important factor of

information in video. ASR is used to enable computers to recognize speaking voice characters. In the model open source automatic Speech Recognition tool is used which works with following steps. Firstly it receives input audio file and then extract sound from input file. After that it will apply speech recognition engine to audio file and set dictation Grammar. The important work is to recognize text from input audio file using speech recognition engine and dictation grammar. At last step appropriate results are stored into database.

6.4 Retrieval OF Video

After applying the whole process of video segmentation, OCR technique, and ASR technique resulted output data is saved as OCR and ASR results into database. When user will entered a search query in the form of text that will directly check with the saved results in database.

7. CONCLUSION

In this paper, we have established an effective information retrieval model from video lecture data. The existing systems are doing retrieval depend on different technique which supplies the WER nearly 71%. But as here we have used OCR with ASR technique with preprocessing technique, which decreases word error rate much and also the clustering method used after data retrieval, which is depend on ontology of the terms, supplies much co-related and appropriate metadata for the file and linked information for the videos. So the video lecture searching becomes efficient and easy.

8. REFERENCES

- [1]. Madhav Gitte¹, Harshal Bawaskar², Sourabh Sethi³, Ajinkya Shinde⁴ “Content based video retrieval system”, International Journal of Research in Engineering and Technology, Volume: 03 Issue: 06 | Jun-2014.
- [2]. 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.
- [3]. Wolfgang Hürst, Thorsten Kreuzer, Marc Wiesenhütter , “A qualitative study towards using large vocabulary automatic speech recognition to index recorded presentations for search and access over the web”
- [4]. Xiaoqiang Xiao, Jasha Droppo and Alex Acero” Information retrieval methods for automatic speech recognition”, IEEE international conference on Acoustic Speech and Signal Processing, 2010.
- [5]. Matthew Cooper, “Presentation Video Retrieval using Automatically Recovered Slide and Spoken Text”, Multimedia content and mobile devices, SPIE proceedings vol. 8667, 2013.
- [6]. Alexander G. Hauptmann, Rong Jin, and Tobun D. Ng, “Video Retrieval using Speech and Image Information”, Electronic Imaging Conference (EI'03), Storage Retrieval for Multimedia Databases, Santa Clara, CA, January 20-24, 2003
- [7]. Tiecheng Liu and John R. Kender, “Rule-based semantic summarization of instructional videos”, IEEE International conference on Image Processing, Vol 1, 2002
- [8]. Duc Phuong Nguyen, Martin Guggisberg, Helmar Burkhart “Multimedia Information and Mobile-Learning”, Eighth IEEE International Symposium on Multimedia (ISM'06), 2006
- [9]. Journal article – Nianhua Xie, Li Li, Xianglin Zeng, and Stephen Maybank —A Survey on Visual Content-Based Video Indexing and Retrieval IEEE Transactions On Systems, Man, And Cybernetics— Part C: Applications And Reviews, Vol. 41, No. 6, November 2011.

BIOGRAPHIES



Gaikwad Harshada is a Final year Graduate student, Pursuing her bachelor degree in computer engineering at S.N.D college of Engineering, yeola dist:nashik



Sonawane Jayshri is a Final year Graduate student, Pursuing her bachelor degree in computer engineering at S.N.D college of Engineering, yeola dist:nashik



Gore Kapil is a Final year Graduate student, Pursuing his bachelor degree in computer engineering at S.N.D college of Engineering, yeola dist:nashik



Avhad Madhuri is a Final year Graduate student, Pursuing her bachelor degree in computer engineering at S.N.D college of Engineering, yeola dist:nashik