IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834,p- ISSN: 2278-8735.
PP 38-43
www.iosrjournals.org

Video Data Retrieval Based On Text Content Detection

Ms.M.Ayesha Parveen[1], Mr.P.Deepan[2], Ms.S.Jummera Nasrin[3],
Ms.J.ISRATH BEGUM[4]
Department of CSE, Arasu Engineering College, Kumbakonam, India.
E-mail id:ayeshaparveen300@gmail.com[1],deepanp87@gmail.com[2], nasrinjummera@gmail.com[3], israthbegum12@gmail.com[4]

Abstract: *Informational retrieval application plays a major role in web content mining. Now a days,e-lecturing has become added professional popular. The number of lecture video data on the World Wide Web (WWW) is growing rapidly.Upto this development, the video can be retrieved only based on the title or description of the video.In this paper presents, the video will be retrieved based on speech and video text content.We offered an approach for content-based lecture video segmenting and retrieval in large lecture video collection. Video consists of image frames with applicable points mentioned by the text. Subsequently, we extract textual metadata by applying video Optical Character Recognition (OCR) technology on key-frames and Automatic Speech Recognition (ASR) on lecture audio tracks. Hence the proposed system will be more efficient for retrieving the videos and also improves the recognition rate.*

Keywords*:* *Speech Recognition, Lecture video, content based video search, Video Segmentation, OCR, Extraction, Video retrieval.*

I. I.INTRODUCTION

Now a days, the huge amount of professional lecture videos are available in the WWW (World Wide Web) data. The relevant metadata can be automatically gathered from lecture videos by using appropriate analysis techniques. It is very useful to quickly understandable for students by viewing the video rather than reading the text. It is one of the easiest technique of online course learning. We have a lot of videos that are available in web mining.

In a video searching, such a method is difficult to provide the exact user's required video. It is complicated to find the exact video by giving the input because it will display the video which does not related to that input. So, it consumes more time to retrieve the video. As a result, there has been enormous increase in the sum of digital video data on the Web.

Most of all, an e-lecture video consists of image frames with applicable points mentioned by the lecturer. In some e-lectures, one can find a written transcript of the oral presentation. Text is an advanced semantic feature which has often been used for content-based information retrieval.In lecture videos, text from lecture slides serve as an outline for the lecture and are very important for understanding. The video retrieved through the title and description of the video files stored in the data stores. In this paper, we proposed the system, the video will be retrieved based on speech and video text content.

II. RELATED WORK

The main objective of this paper is to retrieve the efficient video using speech recognition based on content text in video.Information retrieval in the digital video-based learning domain is an active and frequently dynamic study of research area. Video texts, oral presentation, manual explanation, video actions, or body language of speakers can act as the source to open up the content of lectures.To retrieve the efficient video using speech recognition based on content text in video.

1. Lecture video retrieval

Haojin Yang et al. presented an approach for content- based lecture video indexing and retrieval in large lecture video analysis [1]. This approach applies video as well as audio resource of lecture videos for extracting content-based metadata automatically. For textual extraction purpose the OCR (Optical Character Recognition) and ASR (Automatic Speech Recognition) techniques are used.

Wang et al. proposed an method for lecture video indexing based on automated video segmentation and OCR (optical character recognition) analysis [2]. This OCR technique is used to scan the text from the frames. The input is given as a speech. The speech can be converted into text by using ASR (Automatic Speech Recognition) technique. The scanned text is matched with the input text. After that, the e-lecture video can be retrieved.

Stephan Repp et al. proposed an sufficient chain indexing method for computer science courses based on their existing recorded videos by using Speech Recognition Engine (SRE) analysis[3]. The index structure and the evaluation of the supplied keywords are presented within the lecture video and the user interface for dynamic browsing of the e-learning content.

Tayfun Tuna et al.proposed an technology for Indexed, Captioned, and Searchable videos and their usage for STEM (Science, Technology, and Engineering& Mathematics) coursework. Indexing and search features were considered very helpful and easily understandable for user. This work points to a new and innovative direction for effective use of videos in STEM courseworkanalysis [4]. The framework developed is freely available to educational institutions.

Hyun Ji Jeong et al. proposed a new method for lecture video segmentation by utilizing SIFT (Scale Invariant Feature Transform) analysis [5]and the adaptive threshold. By using SIFT, we can reliably match two slides whose contents are the same but are visually different. Here, developed a method for selecting the adaptive threshold. Since the adaptive threshold value is automatically adjusted for each local video segment and do not need to set the value heuristically.

2.Content based video search

ASR provides speech-to-text information on spoken languages, which is thus well suited for content-based lecture video retrieval. Here the universal language like as english language only taken as a input of the content based video detection.

III. METHODOLOGY

The proposed system describes that the user giving voice input. The voice can be converted into text. The required video is obtained from database. It can be segmented, the text can be extract from the frames. Whether the extracted text and input text are matched, finally the relevant video can be retrieved.

Proposed Architecture

The architecture of content based video retrieval system are as given below (fig.1)

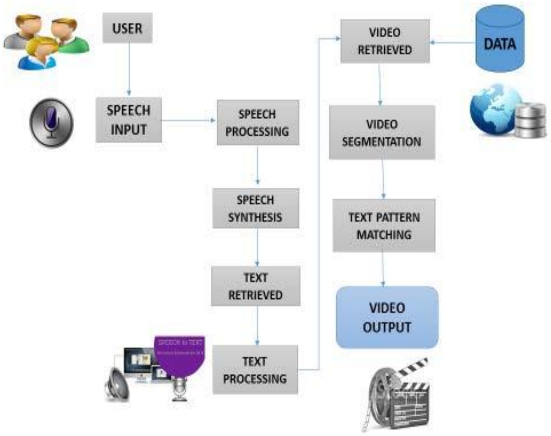

Fig. 1 System Architecture

A. SPEECH TO TEXT ANALYSIS

The easiest way to create notes with your voice is to record an audio note. However, does not convert audio recordings into text nor does it allow you to search for a word mentioned inside the recording. Modern general-purpose speech recognition systems are based on Hidden Markov Models. These are statistical models that output a sequence of symbols or quantities. HMMs are used in speech recognition because a speech signal can be viewed as a piecewise stationary signal or a short-time stationary signal. In a short time-scale (e.g., 10 milliseconds), speech can be approximated as a stationary process. Speech can be thought of as a Markov model for many stochastic purposes.

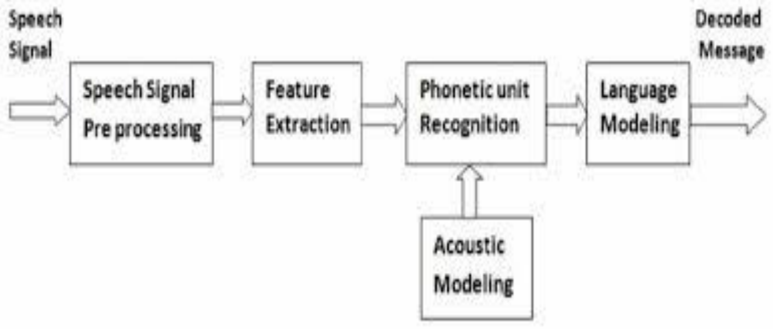

Fig.2 conversion of speech into text.

B. TEXT QUERY PROCESSING

Text is a high-level semantic feature used for the content-based information retrieval. In lecture videos, texts from lecture slides serve as an outline for the lecture and are very important for understanding. While positioning a video, title and description are the most important factors, because there we can find most of the necessary information. The titles should be descriptive and should not contain word or phrase that is not useful. The video retrieved through the title and description of the video files stored in the data stores.

C. VIDEO SEGMENTATION

Our segmentation algorithm consists of two steps: In the first step, the entire slide video is analyzed. For reasons of efficiency, we do not perform the analysis on every video frame; instead, we established a time interval of one second. Therefore, our video segmented (fig.3) considers only one frame per second (fps). We try to capture frames from video by providing appropriate time interval; this provides us key-frames. On these gathered key-frames we apply OCR technique, which will extract textual data. This data is then stored in a text file which will be further utilized in ranking and searching process.

When user provides title/name of video to be searched then the search is directed to text files, searching in text files provides better results. Then we create canny edge maps for adjacent frames and build the pixel differential image from the edge maps. The CC analysis (Connected Component) is subsequently performed on this differential image and the number of CCs is then used as a threshold for the segmentation. In this way, high-frequency image noises can be removed in the frame comparison process by adjusting a valid size of CCs.

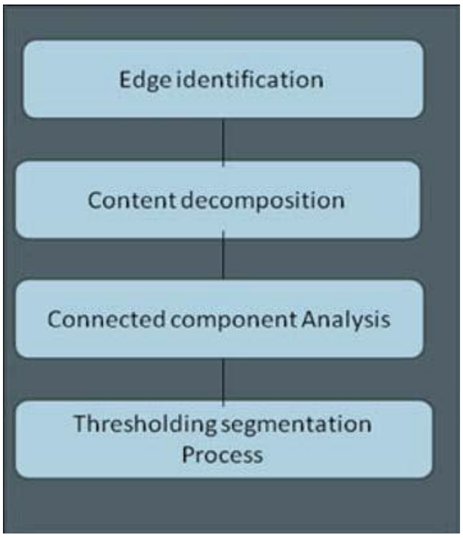
Edge identification
Content decomposition
Connected component Analysis
Thresholding segmentation Process

Fig.3 Video segmentation process block diagram.

D. FEATURE DETECTION AND VIDEO RETRIEVAL

The lecture content-based metadata can be gathered by using OCR and ASR tools. However the recognition results of automatic analysis engines are often error prone and a large amount of irrelevant words are also generated. Therefore we extract keywords from the raw recognition results. Keywords can summarize a document and are widely used for information retrieval in digital libraries. In this work, only nouns and numbers are considered as keyword candidate. The top n words from them will be regarded as keyword. Segment-level as well as video-level keywords are extracted from different information resources such as OCR and ASR transcripts respectively.

For extracting segment-level keywords, we consider each individual lecture video as a document corpus and each video segment as a single document, whereas for obtaining video-level keywords, all lecture videos in the database are processed, and each video is considered as a single document. The lecture content-based metadata is gathered by using OCR and ASR tools. But the recognition results of automatic analysis engines are often error prone and generate a large amount of irrelevant words. Therefore we extract keywords from the raw recognition results.

Keywords summarize a document and are widely used for information retrieval. Only nouns and numbers are considered as keyword candidate. The top n words from them are considered as keyword. Segment-level as well as video-level keywords are extracted from different information resources such as OCR and ASR transcripts respectively. For extracting segment level keywords, we consider each individual lecture video as a document corpus and each video segment as a single document, whereas for obtaining video-level keywords, all lecture videos in the database are processed, and each video is considered as a single document.

To extract segment-level keywords, we first arrange each ASR and OCR word to an appropriate video segment according to the time stamp. Then we extract nouns from the transcripts by using the Stan ford part-of-Speech tagger and a stemming algorithm is subsequently utilized to capture nouns with variant forms. Thus the content based video can be detected and also retrieved from web database.

IV. IV.IMPLEMENTATION

A.ADAPTIVE THRESHOLD ALGORITHM

Adaptive threshold algorithm is used for segmentation of video. The histogram of oriented gradient point is used to segment the video. It is a feature descriptor used in computer vision and image processing for the purpose of object detection.

HOG Feature Extraction

Input: Image with edge detection

Output: Feature extracted gradient value

```
step_x=floor(C/(nwin_x+1));
step_y=floor(L/(nwin_y+1));
Cont=0;
hx = [-1, 0, 1];
hy = -hx';
grad_xr = imfilter(double(Im),hx);
grad_yu=imfilter(double(Im),hy);
angles=atan2(grad_yu,grad_xr);
magnit=((grad_yu.^2)+(grad_xr.^2)).^.5;
Edge_I=imread(['Detected_Edge\',num2str);
HOGfea=HOG(Edge_I);
set(handles.uitable1,'visible','on');
set(handles.text1,'visible','on');
set(handles.uitable1,'data',HOGfea);
imwrite(HOGfea,['Extracted_Feature)
```

Histogram of Oriented Gradient is an algorithm which is used for the Feature Extraction.

B.Multi - SVM Algorithm

Multi SVM(Support Vector Machine) classifier algorithm used for classification technique. The main purpose of this algorithm to classify the text from the required video because we are searched about content based. It is also called as Support Vector Network. In machine learning, the SVM are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis.

Multi Support Vector Machine

Input: Extracted feature frames

Output: Retrieval of video text

```
multisvm(TrainingFea,GroupTrain,TestFea);
u=unique(Group Train);
numClasses=length(u);
C=zeros(length(TestFea(:,1)),1);
```

svmtrain(TrainingFea,G1vAll);
(svm classify(models(k),TestFea(j,:)))

In SVM classifier, the scanned text from the video frame. Thus the extracted text from image frame is matched with the input text from the Automatic Speech Recognition (ASR). Finally predicted the user's expected video output.

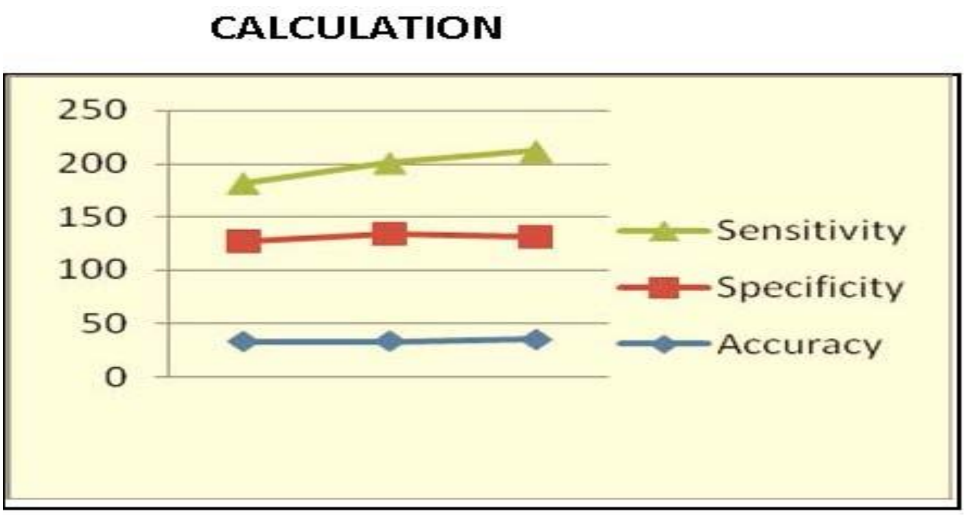

Fig.4 Performance Calculation.

Processing Accuracy can be calculated by means of Recall and Precision. The formula for the Recall and Precision as follows.
Recall =<u>number of correctly retrieved outline words</u>
Number of all outline words ground truth

Precision =<u>number of correctly retrieved outline words</u>
Number of retrieved outline words

Processing Accuracy	Retrieved Video
Recall	0.8700
Precision	0.4350

The above table shows the value of single video. The recall and precision are calculated and valued in the table. The Performance Analysis has shown as below graph.

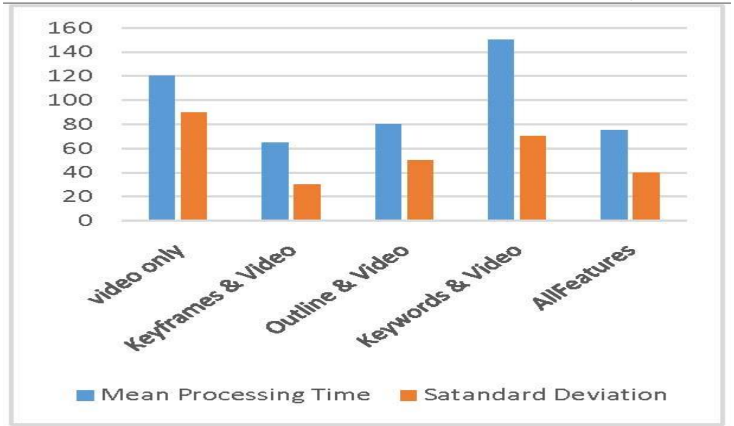

Fig.5 Performance Analysis.

The video retrieval accuracy of the performance analysis method is evaluated. The analysis processing of speed can be determined by mean processing time and standard deviation. The analysis of video output is compared with existing system. Features and keyword analysis are also calculated and are compared.

V. CONCLUSION

In this paper, we presented a new approach for content- based lecture video and retrieval in huge amount of lecture video archives.The proposed system will be more efficient for retrieving the videos and also improves the recognition rate. We offered a concept for content-based lecture video segmenting and retrieval in large lecture video collection. Hidden Markov Model (HMM) is used for speech recognition. The purpose of Edge detection algorithm is to convert into video frames. The process of segmentation can be done by Adaptive Threshold Algorithm using HOG points and OCR (Optical Character Recognition) algorithm is used for scanning the characters from video frames. Finally, Multi SVM(Support Vector Machine) classifier algorithm used for classification technique. Thus the user can be obtained the exact video output.

VI. FUTURE WORK

As the future work, the usability and utility study for the video search function in our lecture video portal will be conducted. Not only the lecture video, it is efficient to retrieve all the multimedia videos and also achieve a high performance speed. In future enhancement, the video will be retrieved content based speech of the user, the speaker of that video.

REFERENCES

[1]. Haojin Yang and Christoph Meinel, “Content Based Lecture Video Retrieval Using Speech and Video Text Information”, IEEE transactions on learning technologies, 2014.

[2]. T.-C. P. F. Wang, C-W. Ngo, “Structuring low-quality videotaped lectures for cross-reference browsing by video text analysis,” Journal of Pattern Recognition, vol. 41, no. 10, pp. 3257–3269, 2008.

[3]. S. Repp, A. Gross, and C. Meinel, “Browsing within lecture videosbased on the chain index of speech transcription,” IEEE Transaction Learning Technology volume 1, no. 3, pp. 145–156, Jul. 2008.

[4]. 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 Proceeding 43rd ACM Technical Symposium Computer Science Education.

[5]. H. J. Jeong, T.-E. Kim, and M. H. Kim.(2012), “An accurate lecture video segmentation method by using SIFT and adaptive threshold,”in Proceeding 10th International Conference Advances Mobile Computing pp. 285–288.

[6]. MS.G.Vigneshwari, Mrs.A.Noble Mary Juliet,” Optimized Searching of Video Based On Speech and Video Text Content”, in the process of 2015 International Conference on Soft-Computing and Network Security (ICSNS -2015), Feb. 25 – 27, 2015, Coimbatore, INDIA.

[7]. Holub.A, Moreels.P, Perona.P,”Unsupervised Clustering For Google Searches Of Celebrity Images”, in Process IEEE International Automatic Face& Gesture Recognition.,2012.

[8]. B. Epshtein, E. Ofek, and Y. Wexler, “Detecting text in natural scenes with stroke width transform,” in Proc. of International Conference on Computer Vision and Pattern Recognition, 2010, pp. 2963–2970.

[9]. C. M. F. Moritz, M. Siebert, “Community tagging in tele-teaching environments,” in Proc. of 2nd International Conference on e-Education, e-Business, e-Management and E-Learning, 2011.

[10]. M. Grcar, D. Mladenic, and P. Kese, “Semi-automatic catego- rization of videos on videolectures.net,” in Machine Learning and Knowledge Discovery in Databases, European Conference, ECML PKDD2009,Bled,Slovenia,September7-11,2009,Proceedings,PartII,ser. Lecture Notes in Computer Science, W. L. Buntine, M. Grobel- nik, D. Mladenic, and J. Shawe-Taylor, Eds., vol. 5782. Springer, 2009, pp. 730–733.

[11]. D. Lee and G. G. Lee, “A korean spoken document retrieval system for lecture search,” in Proc. of the SSCS speech search workshop at SIGIR, 2008.

[12]. Youhao Yu, “Research on Speech Recognition Technology and Its Application”, 2012 International Conference on Computer Science and Electronics Engineering, 978-0-7695-4647-6/12, 2012 IEEE.

[13]. D. Lee and G. G. Lee, “A korean spoken document retrieval system for lecture search,” in Proc. of the SSCS speech search workshop at SIGIR, 2008.

[14]. E. Leeuwis, M. Federico, and M. Cettolo, “Language modeling and transcription of the ted corpus lectures,” in Proc. of the IEEE ICASSP. IEEE, 2003, pp. 232–235.

[15]. C.Meinel, F.Moritz, and M.Siebert,“Community tagging in tele- teaching environments,” in Proceeding 2nd International Conference eEducation, business, e-Management and E-Learning 2011.