An Efficient Approach of Video Indexing and Retrieving For Search Engines

Prof. S .N .Badhane[1], Mr. Naval Somani[2], Miss. Swati Matale[3], Mrs. Chaitali Mahale[4], Miss. Pooja Tare[5]
[1, 2, 3, 4, 5] Department of Information Technology
[1, 2, 3, 4, 5] PVG's COE Nashik, Maharashtra, India

Abstract- *E-lecturing has become more popular in the last decade. Amount of lecture video data on the Internet is growing rapidly. Therefore, a more efficient method for retrieval video on Internet or within large lecture video archives is promptly needed. This system presents an approach for automatic video indexing and video search in huge lecture video archives. Firstly, we apply automatic video segmentation and key-frame detection to offer a visual help for the video content navigation. Secondly, we extract textual Meta data by applying Optical character recognition (OCR) technology on key-frames and Automatic Speech Recognition (ASR) on lecture sound tracks. The OCR and ASR transcript as well as detected slide text line types are used for keyword extraction, by which both video- and segment-level keywords are extracted for content-based video browsing and searching. The performance and the effectiveness of proposed indexing functionalities are proven by performance evaluation.*

Keywords- Automatic speech recognition, automatic video indexing, content-based video search, lecture video archives, Optical character recognition.

I. INTRODUCTION

Retrieval of video, using speech and text information, which is fetched from multiple videos and resulting in the creation of clusters. We have implemented a model which captures various frames from a video. All the captured frames are then classified according to the duplication property. Then we fetched all the text from all the frames for further video retrieval system. Also we evoke all the voice resulting into text using ASR technique is also used in the process of video retrieval system. 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 likely all the videos or result are according to the name of the video. Means it uses the name of video for comparing the search text. This whole process doesn't give accurate results. Therefore, for a user it is nearly impossible to find desired videos without a search function within a video archive. Even when the user has found related video data, it is still difficult most of the time for him to justice whether a video is useful by only glancing at the title and other global metadata which are often brief and high level.

Moreover, the requested information may be covered in only a few minutes, the user might thus want to find the piece of information he requires without viewing the complete video. The problem becomes how to retrieve the correct information in a large video archive more efficiently. Most of the video retrieval and video search systems such as YouTube, Bing and Vimeo reply based on available textual metadata such as title, genre, person, and brief description, etc. Furthermore, the manually provided metadata is typically brief, high level and subjective.

Some features can achieve relatively good performance, but their feature dimensions are usually too high, or the implementation of the algorithm is difficult. Feature extraction is very crucial step in retrieval system to describe the video with minimum number of descriptors. Therefore, beyond the current approaches, the next generation of video retrieval systems apply automatically generated metadata by using video analysis technologies.

II. SYSTEM ARCHITECTURE

In this system, Administrator gives the input video to the proposed system. After that we apply ASR for Content Audio Retrieval and also identify and split the video into frames. After applying the ASR convert audio into text and also extract the segment level keywords and store into the database. User searches the input query onto the database. If user query found then provide the result as a clustered based video.

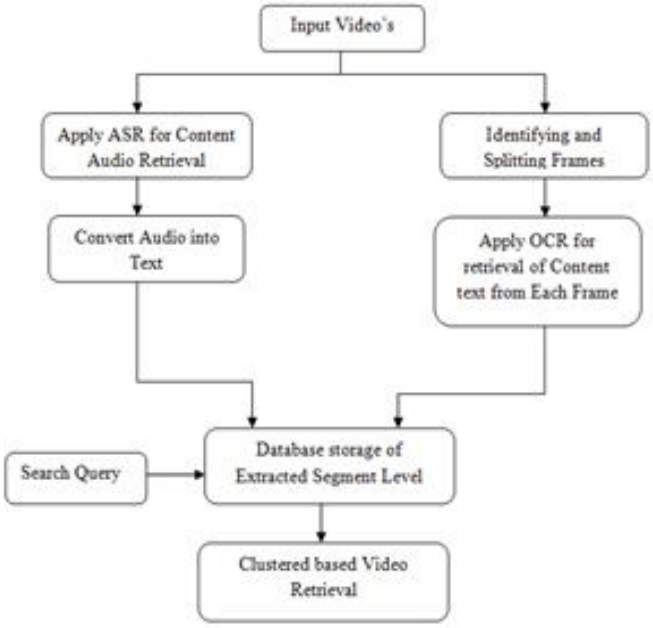
III. IMPORTANT CONCEPTS

A. OCR-Tesseract Engine
B. Speech Recognition Engine
C. FFmpeg tool
D. Mediadet Class
E. ELD Dictionary

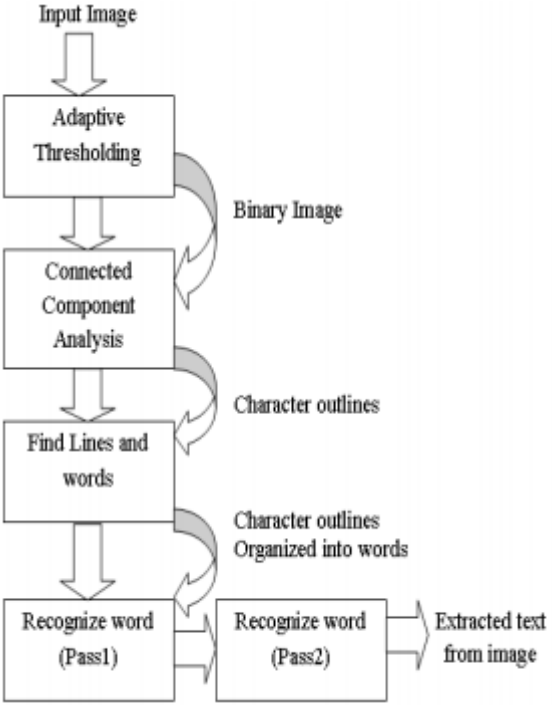
A. OCR-Tesseract Engine:-

Tesseract is a character recognition engine developed on HP-UX at HP between 1985 and 1994 to run in a desktop scanner. It is made open source in 2005.The first phase is a connected component analysis (CC Analysis) in which outline of the components are stored. This was a computationally expensive design decision at the time, but had a big advantage: by checking of the nesting of outlines, and the value of child and grandchild outlines, it is simple to detect inverse text and recognize it as efficiently as black-on-white text.

Tesseract was the first OCR engine able to handle white-on-black text so efficiently. At this stage, outlines are collected together, purely by grouping, into Blobs. Blobs are placed into text lines, and the lines and regions are tested for fixed pitch or Proportional text. Text lines are fragmented into words differently, according to kind of character spacing. Fixed pitch text is separated immediately by character cells. Proportional text is chopped into words using fixed space and fuzzy space. Recognition then proceeds as a two phase



process. In the first phase, an approach is made to recognize each word in return. Each word that is desirable is passed to an adaptive classifier as training data. The adaptive classifier then gets a chance to more precisely recognize text lower on the page. Since the adaptive classifier may have learned something important too late to make a contribution near top of the page, a second phase is done over the page, where words that were not recognized good enough are recognized again. A final phase resolves fuzzy spaces, and checks different hypothesis for the x-height to plot small cap text.

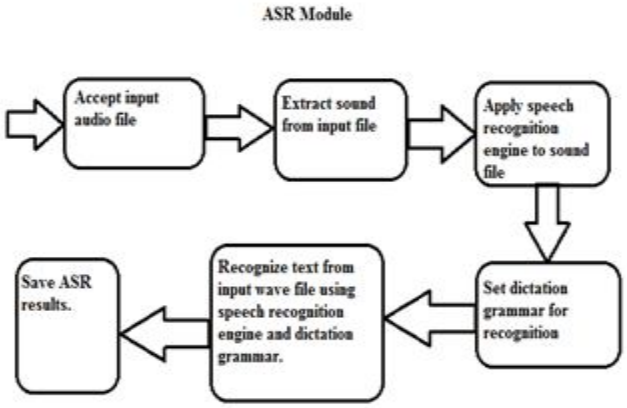


B. Speech Recognition Engine:-

Extracts text from wav file. The goal of an ASR system is to properly and efficiently convert a speech signal into a text message transcript of the spoken words in absence of the speaker, environment or the device used to record the voice (i.e. the microphone). This process start when a speaker decides what to say and speaks a sentence.

The software then produces a sound wave form, which consist the words of the sentence as well as the extra sounds and pauses in the spoken input wav file. Next, the software tries to decode the speech into the best alternative of the sentence. First it converts the speech signal into a fragments of vectors which are measured throughout the time span of the speech signal. Then, using a syntactic decoder it generates a proper sequence of representations.

The ultimate goal of ASR approach is to allow a computer to recognize dynamically, with 100% accuracy, all words that are smartly spoken by any individual, independent of vocabulary length, unwanted noise, speaker characteristics or accent. Today, if the system is trained to learn a person speaker's voice, then much larger dictionaries are possible and accuracy can be greater than 90%.



C. Mediadet class:-

It extracts still images from a video file. Also extracts some other metadata like frame rate, video format, etc. Uses Dexter Library for the same.

D. FFMPEG tool:-

FFmpeg is a command line tool .If we extract audio from video in .wav format then it needs to be merged in c#.net environment. Ffmpeg isa freeware software Project That produces library and programs to handle Multimedia data. FFmpeg has libavcodec, an audio/video codec library used by many other projects,libavformat- an audio/video container mux and demuxlibraries, and the ffmpeg command line Program -to transcode multimedia files. FFmpeg is formed under the GNU Lesser General Public License 2.1+ or GNU General Public License 2+. FFmpeg is developed on Linux, but can be compiled under most OS, including Mac OS X, Microsoft Windows, Android, iOS, as well as AmigaOS. Almost many computing platforms and microprocessor instruction set architectures are too supported, like x86 (IA-32 and x86-64), ARM, DEC Alpha and SPARC. The name of project is inspired by the MPEG video standard group, along with "FF" for "fast forward". The logo uses zigzag form that indicate show MPEG video codecs handle entropy encoding.

E. ELD Dictionary:-

It is very important component, as it provides training data to both Tesseract and ASR .It provides word dictionary, language grammar for the system.

IV. RESULTS

In following table 1 shows Precision and Recall and value of F1 measure. The bar graph however shows accuracy. Evaluation of system is done by precision recall and f1 measure.

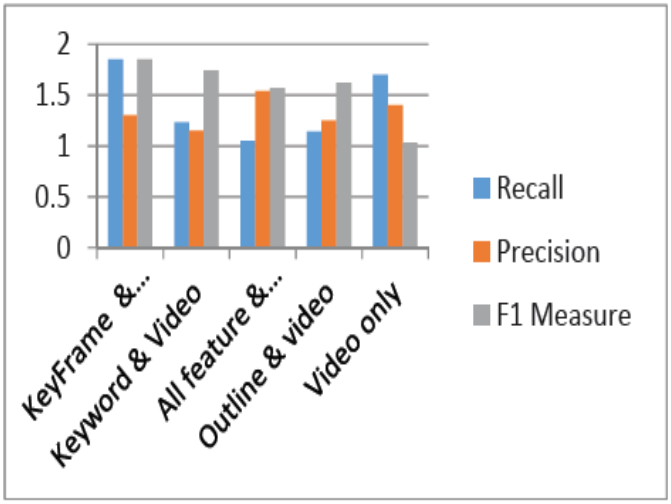
Table of existing system

Setup	Recall	Precision	F1 Measure
Keyframes and video	0.99	1	0.99
Keyword and video	0.99	1	0.99
All Features and video	0.96	0.99	0.97
Outline and video	0.87	0.95	0.91

Table of developed system

Setup	Recall	Precision	F1 Measure
Keyframes and video	1.85	1.3	1.85
Keyword and video	1.23	1.15	1.74
All Features and video	1.05	1.54	1.57
Outline and video	1.14	1.25	1.62
Video only	1.7	1.4	1.03

Bar graph of comparison



V. CONCLUSION

In this paper, we have presented an effective, robust system for indexing and analyzing lecture videos. We have developed and evaluated a novel slide video segmenter and a text localization and verification scheme. Furthermore, we have developed a new method for slide structure analysis using the geometrical information and stroke width of detected text lines.

Further expansion of this system is to support various file formats, as we can't restrict anyone to upload only .avi file. Individual access to "search by image" and "search by sound" options, as if there is no text written on any frame or no sound at all. Support to other official languages other than English (UK), though our official language is English, we have to support at least 4-5 regional languages.

REFERENCES

[1] Haojin Yang and Christoph Meinel, Member, IEEE, “Content Based Video Retrieval Using Audio and Video Text Information”, IEEE transactions on learning technologies, vol. 7, no. 2, april-June 2014

[2] Alexander G. Hauptmann, Rong Jin,Tobun Dorbin Ng,” Multi-modal Information Retrieval from Broadcast Video using OCR and Speech Recognition”

[3] laxmikant Kate, Prof M.M Waghmare “An efficient method for lecture retrieving”.

[4] Haojin Yang, Maria Siebert, Patrick Lühne, Harald Sack, Christoph Meinel”E-Lecture Video Indexing and Analysis by using OCR Technology”Hasso Plattner Institute (HPI), University of Potsdam.

[5] https://github.com/tesseract-ocr

[6] https://tesseractocr.googlecode.com/files/TesseractOSCON.pdf

[7] www.microsoft co.in/asrengine/wrapcode