SAND SEARCH ENGINE

[1]Ashwini Sonawane, [2]Deepshri Mali, [3]Neha Shaikh, [4]S. Shriya, [5]A.R.Gaidhani
[1234]Student, Dept. of Computer Engg., Sandip Institute of Engineering & Management Nasik, India
[5]Dept. of Computer Engg., Sandip Institute of Engineering & Management Nasik, India

Abstract—The rapid pace with which digital information, particularly video, is being generated, has necessitated the development of tools for efficient search of this media. The SAND search engine is web based search engine which crawls the web for video contents. As the amount of video on WWW is growing rapidly so in order to optimize the search of precise method for video retrieval is must. The paper presents an approach for video retrieval using its contents such as image, text, audio. To achieve this objective we apply various algorithm like frame extraction, video optical character recognition (OCR) on key frames and automatic speech recognition on audio tracks (ASR). By identifying the limitations of the current system, the SAND search engine retrieve videos by its contents optimizing the search experiences of the user.
Keywords—CBVR, Frame exraction, OCR, ASR

I. INTRODUCTION

With the advent of recording technologies, improved video compression techniques and high speed networks the storage and exchange of digital video has increased tremendously in the last few years. The present scenario witnesses that users are often shoved with huge amount of videos and are still unsatisfied. Search engines that retrieve videos using titles often neglect the contents of the video and focus merely on the titles. SAND search engine retrieves videos based on titles as well as its contents. Contents based means that the search will analyze the actual contents of the video. As video is the sequence of frames its contents can have various characteristics such as text, voice, image among which text is the key feature to focus on. To enable users to quickly locate their interested contents in huge amount of videos, the extraction of key feature is required. The text can be extracted through automatic speech recognition and optical character recognition. This approach of content based retrieval will provide the users with a wide range of videos of their interest thus enhancing the search experience of the users.

II. EXISTING SYSTEM

In existing system whenever user type any query for searching any video, most probable all the videos or result are according to the name of the video. Means it uses the name of video for comparing with the search text. This whole process doesn't give accurate results. By surveying the various video search engines, we found a loophole in the way the search engine retrieve videos. For example, when a user types "cricket" in the search box then the resulting videos whose name matches with the keyword appears. Additional videos related to cricket are not displayed and hence many times user is not satisfied. As there are many videos related to cricket but the problem is that, their names do not contain the keyword "cricket" and so the search engine fails to display them in result. The scenarios are illustrated by following figure.

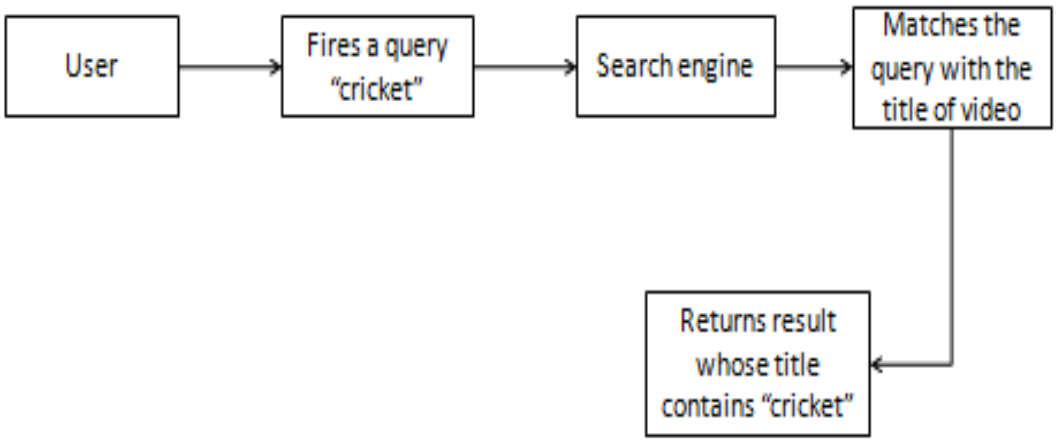


Figure 1. Block Diagram of Existing System

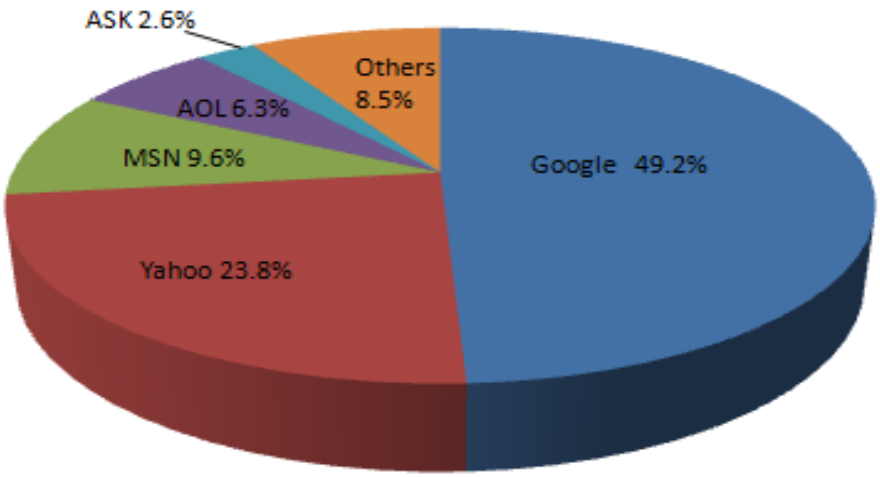


Figure 2. Statistical representation of current video search engines.

III. PROPOSED SYSTEM

The proposed SAND search engine enables retrieval of video using text and speech information which is fetch from multiple videos. The figure given below depicts the system architecture. There are 2 entities i.e. users and admin. The role of the admin is uploading the video and managing the database. The video is splitted into frames using frame extraction algorithm. Delay of some seconds is given to each frame and the frames are extracted after the specified delay. To retrieve textual representation from video OCR and ASR algorithms are applied. Firstly, the frame extraction algorithm is applied to extract the individual frame from the sequence of frames and OCR algorithm is applied. Secondly, the audio tracks are used to derive the .wav from the video on which ASR algorithm is applied to extract text. Eventually when the textual representation is obtained the search query is matched with the available data and all the videos are retrieved in which the users are interested in.

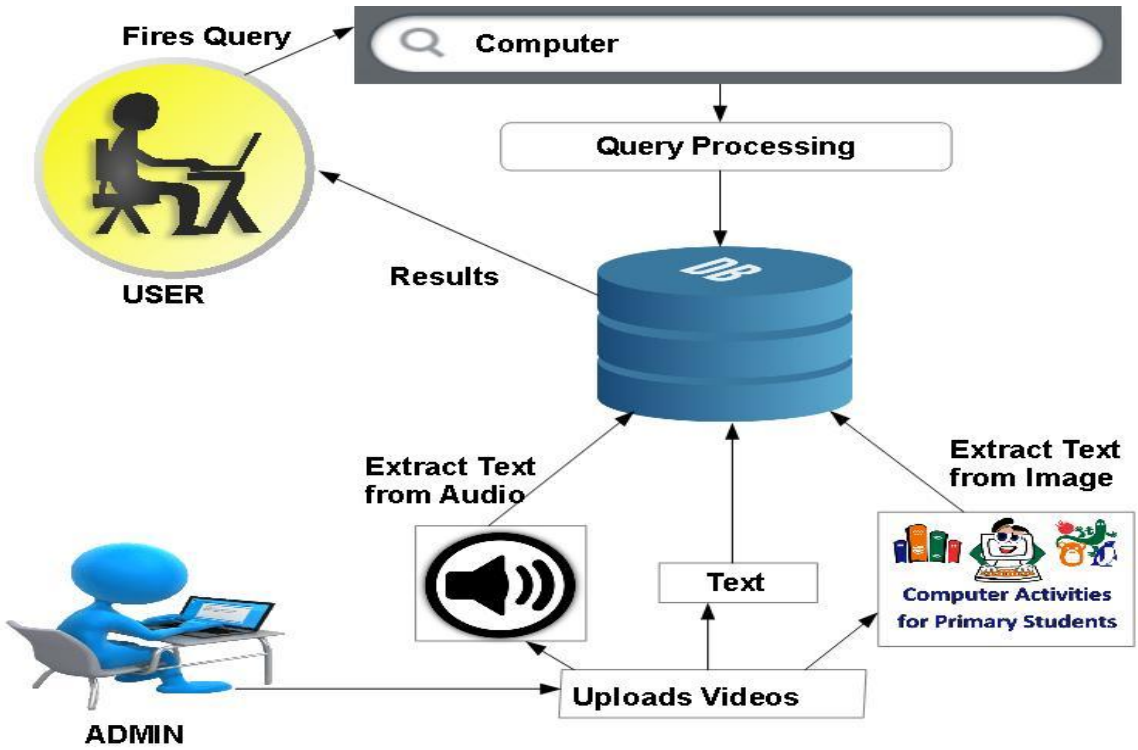


Figure 3. System Architecture

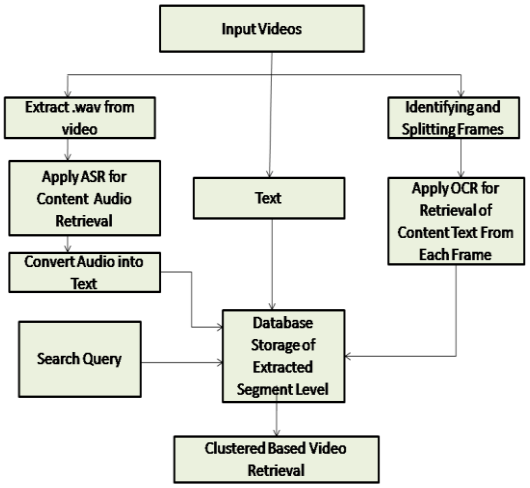


Figure 4. Execution flow of system

A. Video Fragmentation Module

This is the first module of our project in which video is given as an input and that video is splitted into the number of key frames within certain amount of time interval in seconds. In some cases, it may happen that the same frame is displayed for a long period of time then to reduce redundancy we will increase the time interval of video segmentation. When this video is fragmented, it forms number of images and one audio wave file of continuous video.

B. OCR Module

Text is a rich feature of the content based video retrieval process. An OCR is a system that loads an image, performs pre-processing and feature extraction on image and calculates the distances among the extracted image features and predefined known feature vectors stored in the image model library. It recognizes image according to degree of similarity between loaded image and the image model. The following figure represents the common optical character recognition algorithmic steps in the form of a flowchart.

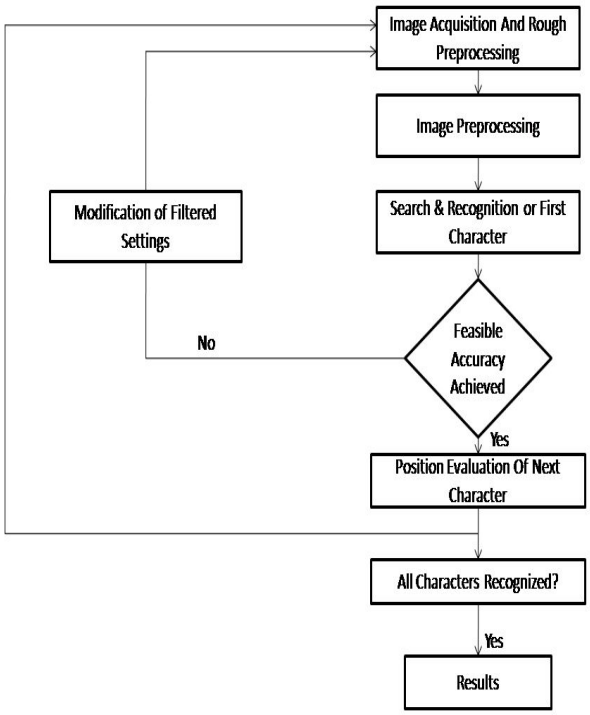


Figure 5. OCR Flowchart

In the pre-processing step, images are made suitable for applying feature extraction algorithms. Some feature extraction algorithms calculate every pixel of the image. Multiple operations such as binarizing thresholding, segmentation, filtering, and edge detection are performed in pre-processing and it also makes images suitable for computation. The next step is classification in which each loaded character image is assigned to one of the possible image model which corresponds to the matching stage of object recognition system. The final result is calculated on the basis of similarity measuring metrics.

C. ASR Module

The automatic speech recognition (ASR) algorithm extracts voice or speech from multimedia files and converts it into meaningful textual information and stores it into the database. Voice plays a crucial role in carrying the information in videos. Therefore, it is of prominent advantage that this information can be applied for automatic video indexing. ASR is aimed to enable computers to recognize speaking voice characters without human intervention. For converting the audio information into text, the ASR takes the text information from the phoneme which has the information for converting the audio waves into specific order that is required to form text and the extracted words are validated by using the word dictionary attached to it. Automatic Speech Recognition ASR is technology which allows a computer to identify the spoken words and then convert it to textual form. ASR is aimed to allow a computer to recognize the speech in real time, with full accuracy, all words that are spoken by an individual, irrespective of the size of sentences, noise or characteristics and accent of the speaker. The speech may also contain a sequence of words with some pauses. As speech signal is given as an input to the ASR , it generates a speech wave form and then decodes speech signal into sentences. ASR converts the audio signal into a sequence of vectors. After that it generates a valid sequence of representations by using a syntactic decoder.

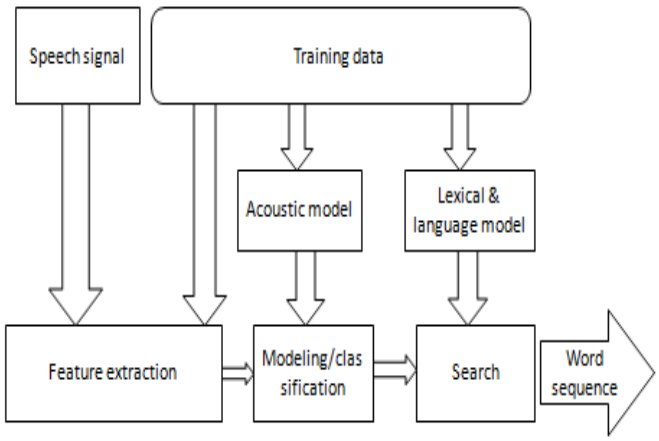


Figure 6. working of ASR

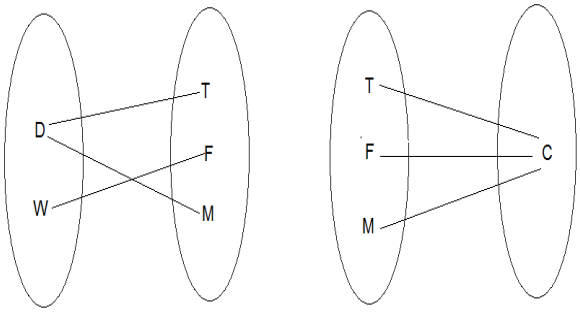
D. Retrieval of Video

Applying the entire procedure of video fragmentation, OCR algorithm, ASR algorithm results output as information is stored as OCR and ASR results into database. When user will fire a search query in form of text that will be directly compared with the stored results in database, by matching with the threshold value. After searching from database, clusters are formed of whole search results and then results are displayed.

IV. MATHEMATICAL MODEL

Problem Description
Let S be a technique for video retrieval system.
Such That S= {I,F,O} Where, I represents the set of inputs:
I= {D, W}
D= Input query (Text Format)
W= Input query (Image Format)
F is the set of functions:
F= {T,F,M}
T= Frame Extraction
F= OCR (Image To Text Conversion)
M= ASR (Wav to Text Conversion

O is the set of outputs:
O= {C}
C= Resulting Video



V. CONCLUSION

As we have observed that the video retrieval has become a crucial factor for geographical information systems, medical research, archiving information etc with the help of this project implementation we will be able to develop a search engine which retrieves and results all the videos according to their contents and not only on the basis of their title and metadata description. The problem of existing system is tried to overcome with the proposed system which in addition reduces the time complexity as user will be provided with the videos that have most appropriate contents related to the search query which will help to enhance the overall experience of the user.

VI. REFERENCES

[1] Haojin Yang and Christoph Meinel, _Content Based Lecture Video Re-trieval Using Speech and Video Text Information_ IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES, VOL. 7, NO. 2, APRIL-JUNE 2014.
[2] Weiming Hu, Senior Member, IEEE, NianhuaXie, Li Li, XianglinZeng, and Stephen Maybank ,_A Survey on Visual Content-Based Video Index-ing and Retrieval_IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS_PART C: APPLICATIONS AND REVIEWS, VOL. 41, NO. 6, NOVEMBER 2011 797.
[3] BassemBouaziz,Walid Mahdi. TarekZlitni, Abdelmajid ben Hamadou,_Content-Based Video Browsing by Text Region Localization and Classi_cation_,International Journal of Video & Image Processing and Network Security IJVIPNS-IJENS Vol:10 No: 01 40
[4] Aditi P. Sangale , Santosh R. Durugkar, International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 4 Issue 5 May 2015, Content Dependent Video Retrieval System.