 ISSN: 2321-8169
726 – 728

Video Retrieval Using Embedded Audio Content

Akash Tawde[1], Vipul Tare[2], *Mrs.* Geetanjali Khanvilkar[3]
Department Of Computer Engineering[1,2,3]
RMCET(Mumbai University)
Ambav, Devrukh, India[1,2,3]
tawdeakash4@gmail.com[1], *vipul.tare1993@gnail.com*[2], *geetanjalikhanvilkar@gmail.com*[3]

Abstract- With the advancement of multimedia, digital video creation has become very common. There is enormous information present in the video, necessitating search techniques for specific content. We propose a system that performs unsupervised, automated classification and annotation of videos based on its embedded audio content. The classified and annotated videos can be later searched and retrieved, when a user inputs a keyword. This paper proposes video retrieval through audio feature mapping, for same we used the API(Application program interface) like FFMPEG for separating the audio content from the video. FFMPEG records, converts and forms streamed digital audio and video in numerous formats. After separating audio it convert that audio to text. To do so Microsoft SAPI API is used which separates each word from text. Video is annotated with this extracted keywords. Such keyword based searching is obviously results in more effective retrieval of relevant videos shortly.

*****

1. INTRODUCTION

Today tagging, though widely used, has its own drawbacks. There is need for videos to be searched based on its content. A system is needed which would search the videos for the occurrence of a given keyword and gives the audio-visual content, even specifying the correct position where the keyword was found in the videos. This can be achieved if we are aware of the content of the videos. The knowledge about the content of the videos comes from the metadata of the content. This has given rise to the need of automatic and unsupervised classification and annotation of videos.

2. RELATED WORK

Basic Architecture of the system is shown in Fig. 1 This diagram serves the general purpose to explain the broad structure and working of the proposed system.

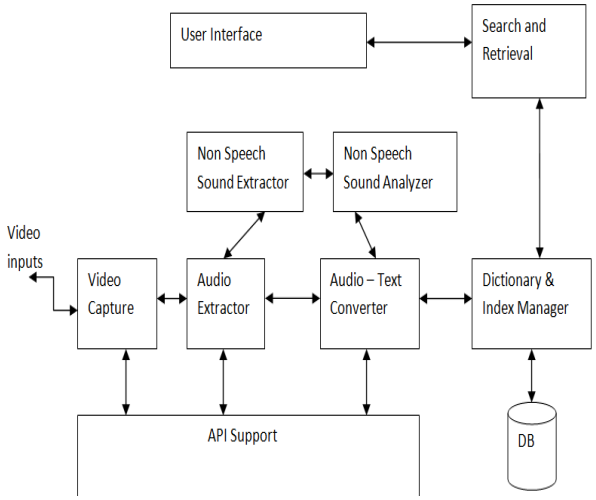

Fig. 1 Basic Architecture of the system

The system consists of the following modules integrated within the framework.

Video Capture: This module captures a series of video data frames or streams video content which acts as the basic input. A link to the source is maintained in a database to enable future retrievals.

The Audio Extractor: This stage is responsible for cleaning up any unwanted noise from the input and extracts the audio stream out of the composite audio-video stream.

Non-Speech Sound Extractor: This stage is responsible for extracting the non speech sounds out of the composite audio-video stream.

Non-Speech Sound Analyzer: This stage is responsible for analyzing the non speech sounds which has been extracted in the previous stage.

Audio to Text Conversion: This phase is primarily responsible to convert the audio stream into textual form. All common words mentioned in an exclusion list are removed from the text to realize the final set of interesting keywords.

Dictionary and Index Manager: The dictionary and index manager maintains a dictionary of encountered keywords. Indexing these keywords ensures efficient retrieval of the information sought.

Search Engine: The search engine is mainly responsible to search the keyword indexes for match as per user requirements.

3. PRAPOSE SYSTEM

We have to give video as the input to the system then system will capture that video and extracts the audio of that video, after extracting audio it convert audio to the text

 ISSN: 2321-8169
726 – 728

format. This is done by using the API's(Application Program Interface). Then indexing will perform for keywords of video and get stored in database. When the user gives the keywords in search box then searching and retrieval is perform from the database and accurate result will be given as the output.

3.1 Audio Extraction

It generates an audio file from a video the embedded audio stream is extracted from video content and is then converted to an audio file. The audio file can be of any formats like mp3, wav etc. We can use audio extraction APIs for this purpose which provides more flexibility, convenience and efficiency. After this that audio file is converted into the text file, i.e. to separate lyrics of that audio, which is in the form of text file. Now we perform the following operations on that text file.

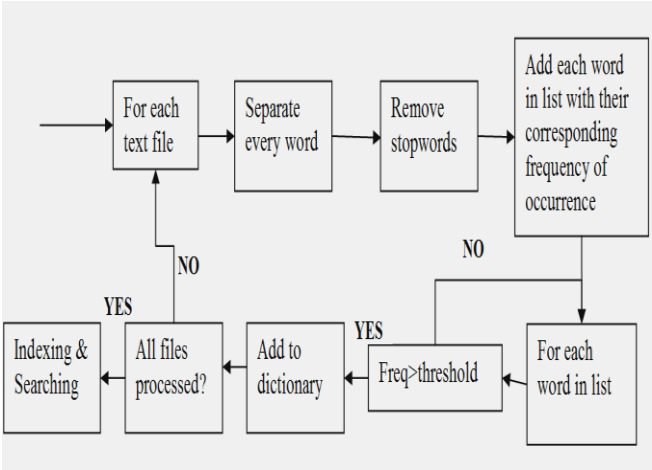

Fig. 2 Flow for Keyword Extraction

Steps :-

1. Separate each word on basis of white spaces.
2. Remove the stop words.
3. If word is in the LIST increment its frequency else insert the word and set frequency to 1.
4. After inserting the word in the LIST if its frequency is greater than threshold promote it to DICTIONARY.

3.2 Video Indexing

Video indexing is done to facilitate a proper search of the videos based on the keyword given by the user. After the creation of keyword dictionary, whenever a new video is added in the database, speech to text conversion is performed on the extracted audio and for all the keywords encountered indexing is done.

Steps :-

1. For each keyword encountered in the text file following steps are performed
2. Check the existing dictionary for the encountered keyword.
3. If the keyword is not found go to step 3. Else go to step 4.
4. Add the keyword in the dictionary.
5. Add a link for the video and the frame(s) in the video at which keyword is encountered to the index field of corresponding keyword in the dictionary.
6. Increment its rank/relevance.

3.3 Video Retrieval

When the user wants to retrieve video, he has to give some keywords in the search box. On the basis of that keyword the video will get retrieved. Only those videos will get retrieved which are most related to that keywords.

Steps :-

1. This keyword is searched in the dictionary for all the keywords with relevance greater than a particular threshold level.
2. If the keyword is not found go to step 3. Else go to step 6.
3. Check the remaining dictionary.
4. If the keyword is still not found go to step 5. Else go to step 6.
5. Display a message to user "KEYWORD NOT FOUND".
6. Read all the links to the videos and corresponding frame(s) from index of the keyword.
7. Display all the videos in order of there rank/relevance.
8. Increment the rank/relevance of video.

4. FUTURE SCOPE

In the future, the project can be extended for classification of other modalities including text as well as visual in the video. The text modality deals with the detection of the presence of an onscreen text i.e. indexing and searching is done on the basis of words found onscreen. The visual modality deals with pattern matching and image mining performed on the frames extracted from the video. A combination of two or more modalities can also be used. This project is a Desktop application, in future it can be extended as a server application.

5. CONCLUSION

In the further course of the project, we intend to study the mentioned APIs, which would be useful for the implementation of the project. For development purposes the project will be divided into modules as depicted in the basic diagram. Individual modules will be separately developed and individually tested. Hence our system searches the video's more accurately and efficiently.

ISSN: 2321-8169
726 – 728

REFERENCES

[1] C. V. Jawahar, balakrishna Chennupati, balamanohar Paluri "video retrieval Based on textual Queries".

[2] Tahir Amin, Mehmat Zeytinoglu, Ling Guan "Intractive Video Retrieval Using Embedded Audio Content".

[3] Geetanjali Khanvilkar, Dr.B.B.Meshram "Video Data Mining: Event Detection from the Association Perspective using FP-growth Tree" December 2011

[4] Zi Huang; Yijun Li; Jie Shao; Heng Tao Shen; Liping Wang; Danqing Zhang; Xiangmin Zhou; Xiaofang Zhou; "Content-based video search: Is there a Need, and Isit Possible?", Information-Explosion and Next Generation Search,2008.

[5] Geetanjali Khanvilkar Asmeena Mhate "Domain Specific Video Data Mining Using Multicore Fp-Growth Tree".