Volume 5, Issue 10, October-2015 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Improved Efficient Technology for Video Retrieval Using Automated Extraction Process of Embedded Audio

Anil Kale[*]
IT Dept. KGCE, Karjat
Mumbai University, India

Dr. D. G. Wakade
Director, P. R. Patil College of Engineering,
Amravati, SGBAU, India

Abstract— The growth of online video material over the internet is generally combined with description or user-defined tags, which is the mechanism for accessing such videos. However user-assigned tags have limitations for retrieval Feature extraction plays an important role than attributes which specifies with videos. This allows to reduce time, space and cost. In the first step the video information with the tags and description from users and applying the compression technique for reducing the video size and it depends on the quality. Audio content of video are extracted and cleaned for further processing the next step converts audio into textual format .The captions are captured from textual format. The text mining can be applied for based on metadata, tags, video description and caption; so that we efficiently retrieve the videos.

Keywords— Text mining, CBR, Feature Extraction.

I. INTRODUCTION

As the amount of available multimedia data has steadily increased lately, users need to be able to access and manage such enormous multimodal corpora efficiently and effectively. Thus, content-based retrieval (CBR), which can analyse the actual contents of the multimedia and facilitate users to access large-scale video data, has been an increasingly active research area since the 1990s.

Efficient and effective video classification and annotation demands automated unsupervised classification and annotation of videos based on its embedded video content as manual indexing is unfeasible.

The advances in the digital and network technology have produced a flood of multimedia information. The people can easily access digital videos which is one of the major constituent of multimedia information.

The growing amount of digital video is driving the need for more effective methods for indexing, searching, and retrieving of videos based on its content. While recent advances in content analysis, feature extraction, and classification are improving capabilities for effectively searching and filtering digital video content, the process to reliably and efficiently index multimedia data is still a challenging issue.

II. LITERATURE SURVEY

“There is much video available today. To help viewers find video of interest, work has begun on methods of automatic video classification. “[1]

Today people have way in to a very great amount of viewing part, both on television and the internet. The amount of viewing part that a viewer has to select from is now so greatly sized that it is infeasible for a man-like to go through it all to discover viewing part of interest. one careful way that viewers use to narrow their selections is to look for viewing part within special groups or like group of books. Because of the very great amount of viewing part to categorize, make observations has started on automatically putting in order viewing part.

A greatly sized number of moves near have been attempted for giving effect to automatic order of viewing part. After going over again the literature of ways of doing, we found that these moves near could be separated into four groups: text-based moves near, audio-based approaches, visual-based approaches, and those that used some mix of wording, sound, and seeing points. Most writers made into one a range of features into their move near, in some cases from more than one modality. wording features are especially useful in putting in order some like group of books. Sports and news both have a tendency of to have more graphic wording than other like group of books. wording features formed (from) from transcripts are better than sound or seeing features at noting between different types of news parts.

Audio features require fewer computational resources to obtain and process than visual features. Audio clips are also typically shorter in length and smaller in file size than video clips. Many of the audio-based approaches use low-level features such as ZCR and pitch to segment and describe the audio signal with higher-level concepts such as speech, music, noise, or silence, which are then used for classification. Some of these approaches assume that the audio signal will only represent one of these high-level concepts [2], which is unrealistic for many real-life situations.

Most of the visual-based features rely in some manner on detecting shot changes and are therefore dependent on doing so correctly. This is the case whether the feature is the shot itself, such as average shot length, or applied at the shot level, such as average motion within a shot. Detecting shot changes automatically is still a difficult problem, primarily due to the variety of forms transitions between shots can take.

Frame-based features are costly to produce if each frame is to be considered. This also results in a tremendous amount of data to process for full-length movies. This can be made easier by only processing some frames, such as the keyframes of video shots. This assumes that the key frame chosen is representative of the entire shot. This assumption will be violated if there is much motion in the shot.

Color-based features are simple to instrument and cheap to process. They are useful in moves near desiring to use through motion pictures sense of right. For example, amount and distribution of light and color group condition of mind. Some unhelped sides are that color histograms lose spatial information and color-based comparisons have pain of when images are under different illumination conditions. The crudeness of the color histogram also means that frames with similar color distributions will come into view as similar without thought or attention of the current What is in. For example, the color histogram of a viewing part frame having in it a red Apple on a blue tablecloth may come into view as similar to the histogram of a red gas bag in the sky.

Object-based features can be costly and difficult to derive. Wei et al. Report that the detecting text objects are efficient enough to be applied to all video frames but that detecting faces is so expensive that they limited it to the first few frames of each shot. Most methods require that the objects be somewhat homogeneous in color or texture in order to segment them correctly, which may also require confirmation from humans [3]. Objects that changed shape, such as clouds, would also prove difficult to handle.

III. SYSTEM ARCHITECTURE

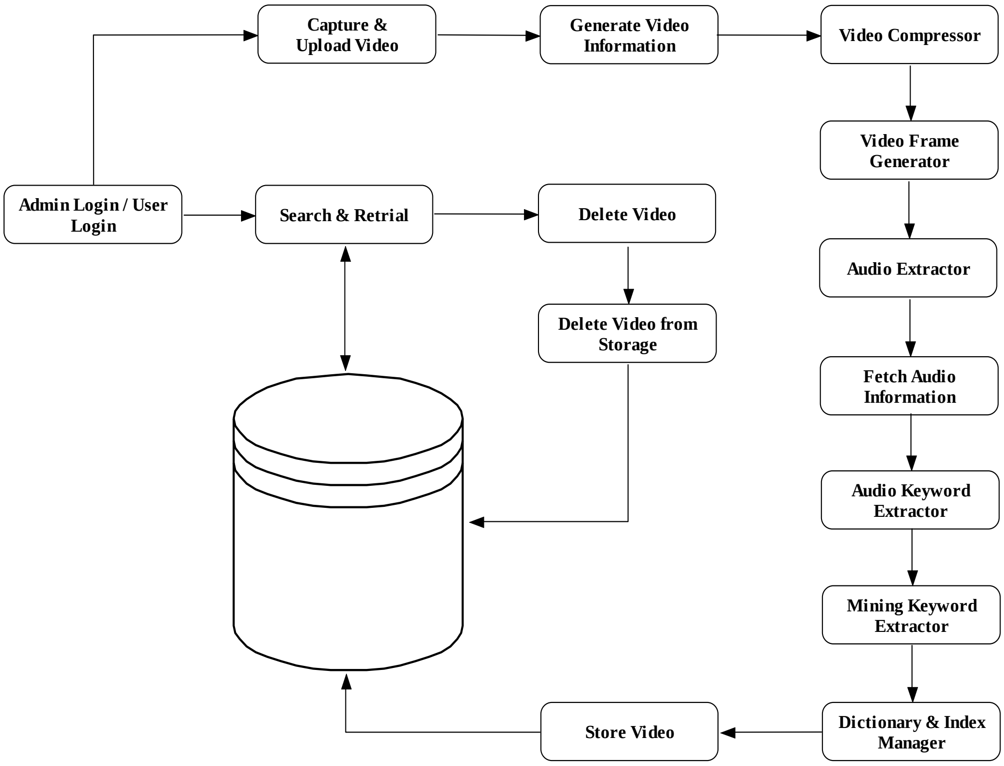


Fig. 1 System Architecture

i) Video capture

This Module will capture a series of video data frames or streams video content which acts as the basic input. This will also collect basic video information (tags, video description) from user.

ii) Generate video information

This module will extract the uploaded video information like video name, video size, video updated by name etc.

iii) Video Compression

This module will compress the video so as to reduce the video size without losing any information in order to improve the response time and page performance.

iv) Video Frame Generator

This module will extract and generate video frames from the video so that initially these frames will be displayed on the website in order to increase the loading time

Instead of displaying videos on the web page, particular frames of the videos will be displayed & user clicks on the corresponding frames the video will be played, hence instead of storing multiple videos in page only video frames will be stored & this will increase the page loading & response time. if there are the situations where multiple videos to be shown on a web page , then only video frames will be displayed & video will be loaded on corresponding frame click.

v) Audio Extraction

This module will responsible for cleaning up any unwanted noise from the input and extracts the audio stream out of the composite audio-video stream.

vi) Fetch Audio Information

This module would be responsible to extract the audio metadata into textual form.

vii) Audio Keyword Extractor

This module phase will extract the embedded audio content. Audio is a rich source of information in the digital videos that can provide useful descriptor for indexing the video databases. Audio archives contrast with image or video archives in a number of important dimensions. First, they capture information from all directions and are largely robust to sensor position and orientation, allowing data collection without encumbering the user. Second, the nature of audio is distinct from video, making certain kinds of information (for example, what is said) more accessible, and other information (for example, the presence of nonspeaking individuals) unavailable. In general, processing the content of an audio archive could provide a wide range of useful information.

viii) Dictionary and Indexing

This will maintain a dictionary of encountered keywords. Search Engine would be responsible to search the keyword indexes for match as per user requirements.

ix) Retrieval

When user enters a keyword for search in the user interface provided to him following steps are performed.

1) This keyword is searched in the dictionary for all the keywords with relevance greater than a
2) Particular threshold level.
3) If the keyword is not found go to step 3. Else go to step 6.
4) Check the remaining dictionary.
5) If the keyword is still not found go to step 5. Else go to step 6.
6) Display a message to user "NO VIDEOS FOR SUCH KEYWORD".
7) Displays the random videos stored on the server with the message
8) Read all the links to the videos and corresponding frame(s) from index of the Keyword
9) Display all the videos in order of their rank.

x) Server Side

Server will be responsible for accepting and validation the uploaded videos, after validation it will perform compression and extract the embedded content of the video

Once all the processing is done, server will save video indexes to the database. It will also displays the graph of the most viewed / liked and disliked videos and Settings module will displays the list of uploaded videos along with their details and admin can delete or configure the videos from this section

Once the video will be uploaded, folder structure will be created on the server for a given video name.

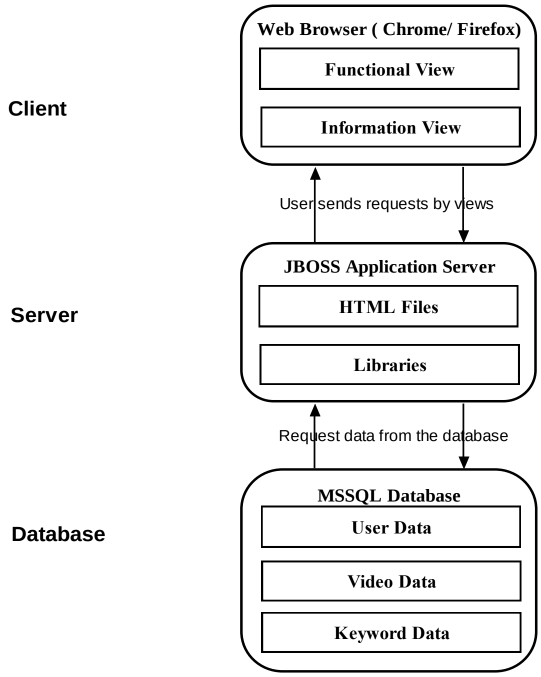


Fig. 1 Client Server Architecture

xi) Client Side

Here a methodology used at the client side is presented i.e. how user's will be presented to the list of uploaded videos , how videos will be searched by entering the prime keywords , how videos are categorized along with the video like / dislike and video download options and the interface with the central database server of the web portal.

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.

1) Video searching based on suggestion
2) Video searching based on keyword
3) Video searching based on category
4) Video like / dislikes and views
5) Video downloads
6) Display of similar or matched videos

IV. RESULTS

i) Video Upload and delete

Here User can upload as well as delete the video.

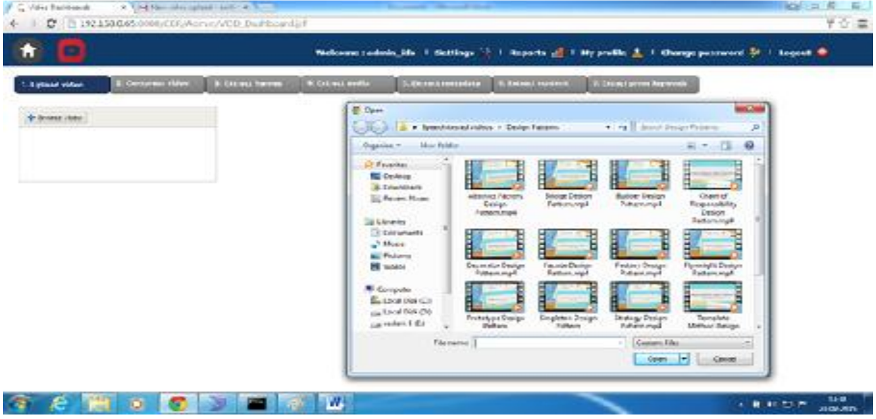
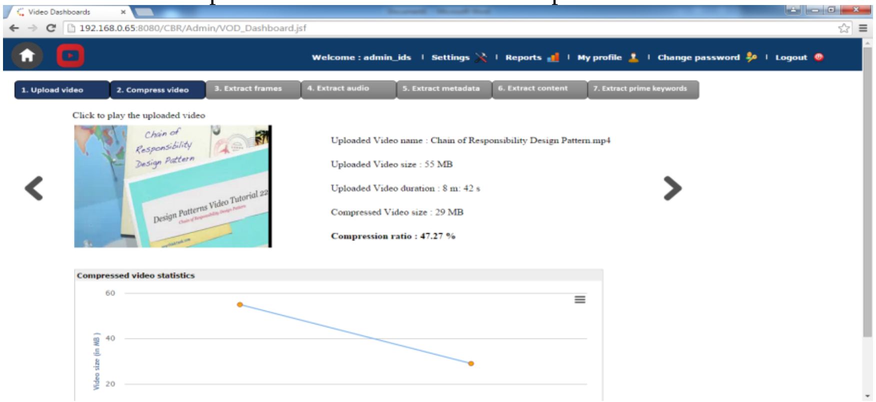


Fig.3 Video Uploading Screen

ii) Video Compression

Here uploaded video has been compressed so that it allocates less space.



Uploaded Video name : Chain of Responsibility Design Pattern.mp4

Uploaded Video size : 55 MB

Uploaded Video duration : 8 m: 42 s

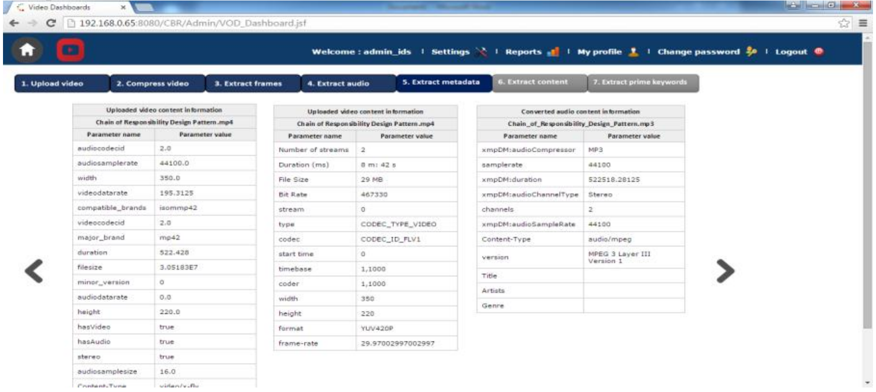
Compressed Video size : 29 MB

Compression ratio : 47.27 %

Fig. 4 Video Compression Screen

iii) Extraction of Audio Information

Here metadata is captured from audio.



Uploaded video content information	
Chain of Responsibility Design Pattern.mp4	
Parameter name	Parameter value
audiocodecid	2.0
audiosamplerate	44100.0
width	350.0
videodatarate	195.3125
compatible_brands	isommp42
videocodecid	2.0
major_brand	mp42
duration	522.428
filesize	3.05183E7
minor_version	0
audiodatarate	0.0
height	220.0
hasVideo	true
hasAudio	true
stereo	true
audiosamplesize	16.0
Content-Type	video/x-flv

Uploaded video content information	
Chain of Responsibility Design Pattern.mp4	
Parameter name	Parameter value
Number of streams	2
Duration (ms)	8 m: 42 s
File Size	29 MB
Bit Rate	467330
stream	0
type	CODEC_TYPE_VIDEO
codec	CODEC_ID_FLV1
start time	0
timebase	1,1000
coder	1,1000
width	350
height	220
format	YUV420P
frame-rate	29.97002997002997

Converted audio content information	
Chain_of_Responsibility_Design_Pattern.mp3	
Parameter name	Parameter value
xmpDM:audioCompressor	MP3
samplerate	44100
xmpDM:duration	522518.28125
xmpDM:audioChannelType	Stereo
channels	2
xmpDM:audioSampleRate	44100
Content-Type	audio/mpeg
version	MPEG 3 Layer III Version 1
Title	
Artists	
Genre	

Fig. 5. Metadata Extraction from Audio Information.

iv) *The text keyword database*

Here the keywords has been extracted from the uploaded audio.

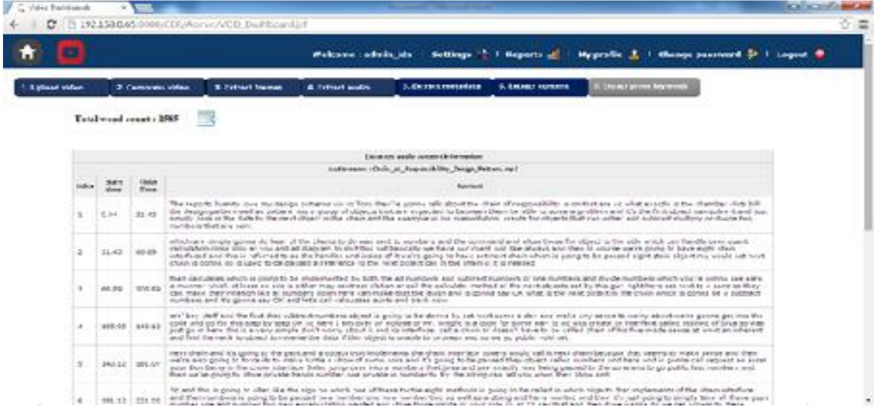


Fig. 6 Text Extraction from Audio Screen

v) *Video Retrieval*

Here user can search the video by giving query in search option.

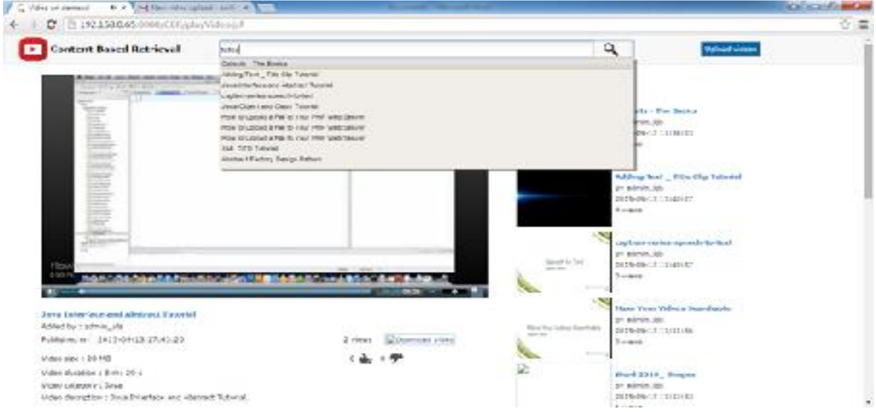


Fig. 7 Video Retrieval Screen

V. CONCLUSIONS

The system would do video classification and annotation based on the embedded audio content. The meta-database would be created for the available database of the videos, thus facilitating easy search on a very large database. The automation will provide ease an also would increase efficiency of the system.

Video classification is usually accompanied with video annotation which helps in retrieving the video archives. Video annotation is about video metadata creation which can be manual or automatic. This will help to improve the performance of the application and accuracy in searching

REFERENCES

[1] Darin Brezeale and Diane J. Cook, Senior Member, IEEE “Automatic Video Classification: A Survey of the Literature”, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, VOL. UNKNOWN, NO UNKNOWN, UNKNOWN 2007.

[2] ACM Multimedia 2001, MM'OI, Sept. 30-Oct. 5, 2001, Ottawa, Canada. Copyright 2001 ACM I-581 13-394-4/01/0009...$5.00 “A Robust Audio Classification and Segmentation Method”, Lie Lu, Hao Jiang and HongJiang Zhang.

[3] G.Y. Hong, B. Fong, A.C.M. Fong “An intelligent video categorization engine” , Kybernetes Vol. 34 No. 6, 2005 pp. 784-802 q Emerald Group Publishing Limited 0368-492X DOI 10.1108/03684920510595490.

[4] Journal of Industrial and Intelligent Information Vol. 1, No. 4, December 2013 , ©2013 Engineering and Technology Publishing doi: 10.12720/jiii.1.4.235-238 “An Automated Video Classification and Annotation Using Embedded Audio for Content Based Retrieval”, Anil Kale, D. G. Wakde.

[5] 2013 International Conference on Cloud & Ubiquitous Computing & Emerging Technologies, 978-0-4799-2235-2/13 $26.00 © 2013 IEEE DOI 10.1109/CUBE.2013.32 “Video Retrieval Using Automatically Extracted Audio”, Anil Kale, D. G. Wakde.

[6] Tahir Amin, Mehmet Zeytinoglu and Ling Guan “INTERACTIVE VIDEO RETRIEVAL USING EMBEDDED AUDIO CONTENT” Ryerson University, Toronto, Ontario, M5B 2K3, Canada tamin, mzeytin.

[7] Alan F. Smeaton, Paul Over, Wessel Kraaij “Evaluation Campaigns and TRECVid” MIR’06, October 26–27, 2006, Santa Barbara, California, USA.

[8] A. Hauptmann, R. Yan, Y. Qi, R. Jin, M. Christel, M. Derthick, “Video Classification and Retrieval with the Informedia Digital Video Library System” TREC’02 Gaithersburg, MD, November 2002.

[9] Arnon Amir3,Winston Hsu4, Milind Naphade1, Janne O Argillander2, Giridharan Iyengar2, Apostol (Paul) Natsev1, Marco Berg3, John R Kender1, John R. Smith1, Donqing Zhang1 Yan1, Shih-Fu Chang4 , Lyndon Kennedy4, Jelena Tesic1, Martin Franz2, Ching-Yung Lin1, Gang Wu1 , Rong. "Video Retrieval System" IBM Research TRECVID-2004.

[10] CEES G.M. SNOEK , MARCEL WORRING. "Multimodal Video Indexing: A Review of the State-of-the-art Multimedia Tools and Applications" 25, 5–35, 2005 c 2005 Springer Science + Business Media, Inc. Manufactured in The Netherlands.

[11] Ying Li, Member, IEEE, Shrikanth Narayanan, Senior Member, IEEE, and C.-C. Jay Kuo, Fellow, IEEE "Content-Based Movie Analysis and Indexing Based on AudioVisual Cues" IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 14, NO. 8, AUGUST 2004 1073.