Video Retrieval System-An Approach based on Image Comparison

Tarun Sharma [#1], Chinta Rahul [*2], Shaik Shahul [#3], M. Venkatesan
[#] School of Computing Science and Engineering
VIT University, Vellore, Tamilnadu, India
[1] tarunsha009@gmail.com
[2] chinta.rahul2@gmail.com
[3] shaik.shahul42@gmail.com
[4] mvenkatesan@vit.ac.in

***Abstract*— Currently people in the world are using many multimedia devices. The best application of the multimedia devices is to watch the videos. The videos are shared on the multimedia device by using various social networks or through the internet. The video, which is shared, is normally a short part of the main video. Most of the time it happens that the video name or any other description is not available. So, how to find the complete video? Our paper will be very helpful for this. Our paper will compare that video in the database and return the complete video along with its description. The videos are the sequence of frames or say images. In this paper, we are using the image comparison technique to match and find the videos. In future we can extend this paper to the Video Surveillance field, as we know these days' security is an important aspect of our life. So adding some artificial intelligent in this we can extend our work in security area also. Using this, we can detect the unwanted activity, unauthorized entry and many more.**

Keyword - Complexity, Efficiency, Techniques, Video Search, Frames, Image Comparison

I. INTRODUCTION

It is beyond doubt if we say that today's generation is totally dependent on the technology. Most of the persons are any how connected with the technology & internet. When we use internet to get some information we have to use the tools like search engines like Google, Yahoo, Bing etc. These search engines try to display the result in the shortest time. If a person wants to search some text or textual content he can easily find that. Even though we have techniques and tools to search the image and audio file too. But till now we are unable to create any efficient technique which provides us the video content search.

From the previous couple of years lot of research work regarding this is going on. Still we are not able to compare the motion images. We have lots of tools which support only retrieval of still images but not a single for the motion pictures. Now we are going to discuss a proposed [16] system as follows:

- Automatic Video Division
- Video Component Library including resolution, size
- Query with different articles.
- Interactive scanning over the Internet.
- Compressed-area video control.

Fundamentally we need to build up an online video recovery framework. Our goal is excessively taking into account the image comparison and sound coordinating we are going to explore the entire video of related video clip. In the next sections we have examined about the related work, techniques which we are using our system expected result and about the future enhancement.

II. RELATED WORK

A developing collection of exploration is analyzing users' general Web looking qualities, with less study inspecting inquiries by clients looking for sight and sound data. Jansen and Pooch [1999] give an in-depth survey of Web client looking studies when all is said in done (i.e., without respect to text or interactive media). Spink et al. [1999] directed exploration into Web client looking aims. Multimedia exploration has regularly centered around image recovery using recorded image accumulations [Enser 1995; Goodrum and Kim 1998; Hastings 1995; O'Connor et al. 1999; Turner 1990]. Some image exploration has concentrated on the configuration of mixed media IR frameworks [Aslandogan et al. 1997].

Different analysts have explored sound and video recovery [Brown et al. 1996]. Smith et al. [1998] give investigation on the interest of looking for video when designing

a media classroom. Goodrum and Spink [1999] particularly investigated clients' image questions, terms and sessions utilizing the same information utilized as a part of our study. Twenty-eight (28) terms were utilized

to recognize inquiries for both still and moving pictures, bringing about a subset of 33,149 image questions by 9,855 clients. They give information on: (1) frame inquiries – the quantity of pursuit terms, and the utilization of visual modifiers, (2) frame search sessions – the quantity of questions per client, alterations made to consequent inquiries in a session, and (3) frame terms – their rank/recurrence circulation and the most exceptionally utilized inquiry terms.

Goodrum and Spink [1999] discovered a mean of 2.64 frame inquiries for each client containing a mean of 3.74 terms for every question. Frame inquiries contained an extensive number of remarkable terms. The most as often as possible happening frame related terms seemed under 10% of the time, with most terms happening just once. This can be differentiated to before work by Enser [1995] who analyzed composed inquiries for pictorial data in a non-advanced environment.

Over the accompanying 7 years TRECVid worked on an assortment of video classes and a scope of substance based undertakings including programmed identification of video shot limits, recognition of semantic ideas inside of shots, completely programmed, self-loader and intelligent quest for video shots or for known recordings, close copy video location, video synopsis, semantic occasion discovery in CCTV and television news story division. These assignments are done in a colossally collective and steady environment with sharing and gift of information and different assets among members being the default, all for the sake of advancing the field of video recovery.

III. PROPOSED METHODOLOGY

1. Development Process:

In this project we are trying to implement the system which can give the results based on the video search. Our planning is to do it in different phases. As we all know that video is the set of continuous frames, which cannot be compared directly. So first we have to convert video in the image format. In the first phase we are going to use the techniques to convert the Video in frames.

1.1 Video to Image Conversion:

As we know People are used to watch TV.TV has the frame rate of about 30 frames per second, frequently called 'full' frame rate. In the video surveillance world, IP cameras are turning out to be more popular which has the full frame rate. Without any doubt, today, most expert IP cameras are intended to stream up to 30 frames for each second. Today we have lots of video to frame conversion software's in the market. So before extracting frames from the video first we have to decide what frame rate per second we need to use.

Normally movies have the average of 24 frames per second & normal DSLR have 30 frames per second. So everything should be considered for that project. Let's consider the following figure:

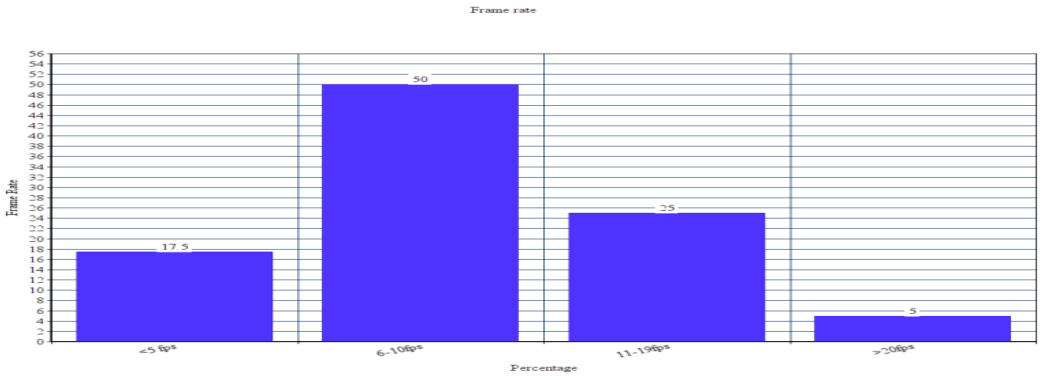

Figure 1 Ratio Between numbers of Frames per seconds

6 - 10 fps is the normal rate utilized for observation recording. It bodes well on the grounds that it gives a genuinely smooth playback while minimizing capacity costs which are still a genuine element for the staggering number of arrangements.

1.2 Image Comparison:

The next phase will be to compare the different video; means now we have to use the different techniques which can give the result that video is same or not. So for that image comparison will be used.

Here we have done the image comparison based on Pixel by Pixel values. Here we are using instance of colour to make things simpler. We have to rescale the image because of the resolution issue. Many times we face the issues like in our database we have the High definition (HD) Video and frames and the sample video is MP4 resolution. So to comparing to different scaled frames is different and there are lots of chances of error.

Next is image comparison for that we used the naive similarity algorithm to compare images. Here our task is to compare the set of images continuously and get the desired result. Here we have set the ranges from 0. If there is no difference between images then the value will be 0. Otherwise we get the different values for different images.

The process of image comparison can be understood from the below given algorithm:

Table 1 Image Comparison Algorithms

Input: Video consisting of multiple frames
Output: Difference between the frames

Algorithm Image Comparison

1. Calculate the signature of the reference image which has to compare.
2. Get the all other files in the component list with whom we have to compare the reference image.
3. For each image calculate the signature.
4. Calculate the difference between the each store image signature with the reference image signature.
5. Return difference.

Algorithm Calculate Signature

Input: signature size N

1. Assign the proportion (x, y) two coordinates values with central pixel.
2. Calculate the average around of the pixel of the all signature values using the pixel proportion value (x, y).
3. Return Signature.

Algorithm To Calculate Average Around

1. Assign the sample size for the image pixels.
2. Get the pixel value and store in the accumulator.
3. Calculate the average of the accumulator values.
4. Return Average Around.

Algorithm To Calculate Distances of two signature values

1. Calculate the signature for the image.
2. Get the RGB values of each pixel
3. Compare the two signature images RGB values by following steps,
    a. Difference between each colour i.e. Red, Green, Blue colours of two signature image.
    b. Multiply each colour difference with itself.
    c. Add the all difference.
    d. Take square root of it.
4. Return the distance.

Figure 2 shows the generalized algorithm of Image comparison. It briefly concentrates on discovering few areas of a given picture that matches with the pictures in the image store as opposed to scanning for equity of the items in given images. It is an algorithm to find similar regions in a set of images.

Figure 3 explain the whole process. First we have video clip which is extracted in the Frame format. Our database (Video) is also converted in the frame format. Now the image comparison algorithm processes until the matched content is found. After matching it gives the name of that particular content in the result.

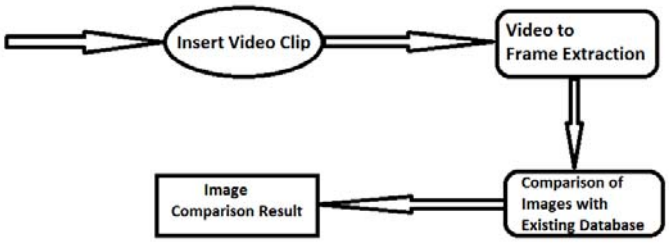

Figure 2 Process of Video Comparison

1.3 Audio Comparison:

Third phase will comprise sound/audio recognition techniques and then we have to synchronize the audio and video techniques to give the desired result. So here we have to extract the audio from the given video clip. There are lots of techniques to extract the audio from the video.

So here firstly we are converting the audio sound in the byte code. When our image comparison is done so from that result we come to know that a particular database is matched with the given Video clip. So now when we get a particular content matched then on this only we apply the audio comparison technique. We are matching the both the audio using the byte codes only.

2. Principle of Operation:

In this whole process we are planning to use the association Rule (It is planned to recognize solid principles found in databases utilizing distinctive measures of interests) means we are planning to take the different measures of images in percentile and for audio too. Now suppose we are assuming some criteria like if images comparison percentage is 70(%) and audio matching percentage is 80(%) then only we can say that the given video is matched with the particular file.

In that we have to be very careful about some issues. First if users have the different resolution video and database has different. Second if we need to maintain the efficiency as size of video database will be large. May be two different video has the same type of graphics or two different video may have same audio. So everything should be considered for that project.

Next thing is that whenever we are going to covert video in image/frames format, then size of total frames is very large than video. So how to compare too much frames will be the next thing which we need to keep in mind. Storage is also an issue as even video file is too large and after converting it into frames it requires lots of storage space. So we have to take care of all type of complexities.

IV. EXPECTED RESULT/OUTCOME

Our dream is whenever you are getting any short clip in your mobile or in a laptop, instead of wasting time by searching the name of the clip, you just go to our system/software, upload your video clip and that's all. If that clip has the full length video or Film you will get that result.

To develop this system we are going to use Net Beans, as programming language Java will be used. As to come up with this system we need to do lots of development which requires different tools. To convert video in the image/frame format and extracting audio from the video file we are using FF-MPEG tool. Different Image processing algorithms will be used to compare the images and give the result. As we know that we can do image comparison using pixel by pixel or based on color. But the important thing is which algorithm takes less time to compare more images. Some text and audio conversion or searching can be done.

Below are the screenshots of our system which we tried to develop:

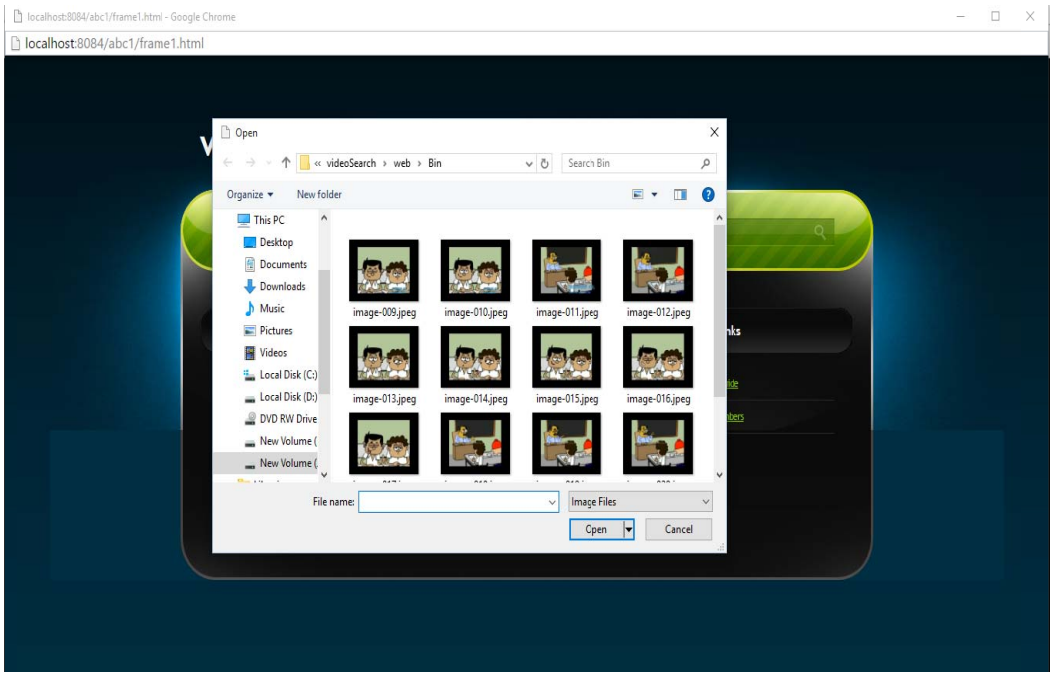

Figure 3 Screenshot of Video Retrieval System

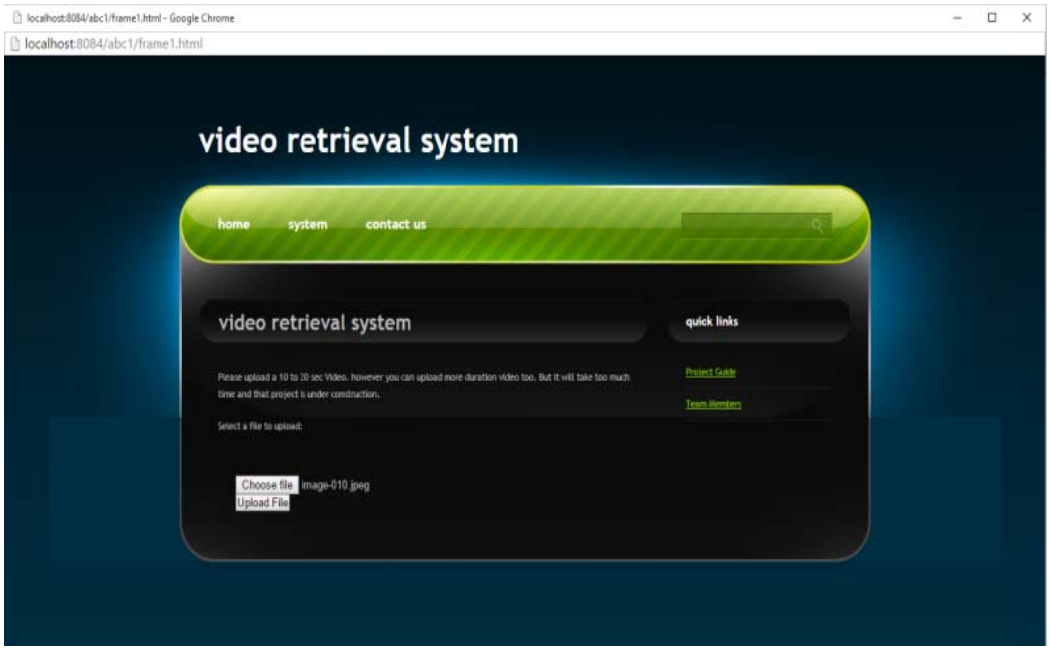

Figure 4 Screenshot of Video Retrieval System

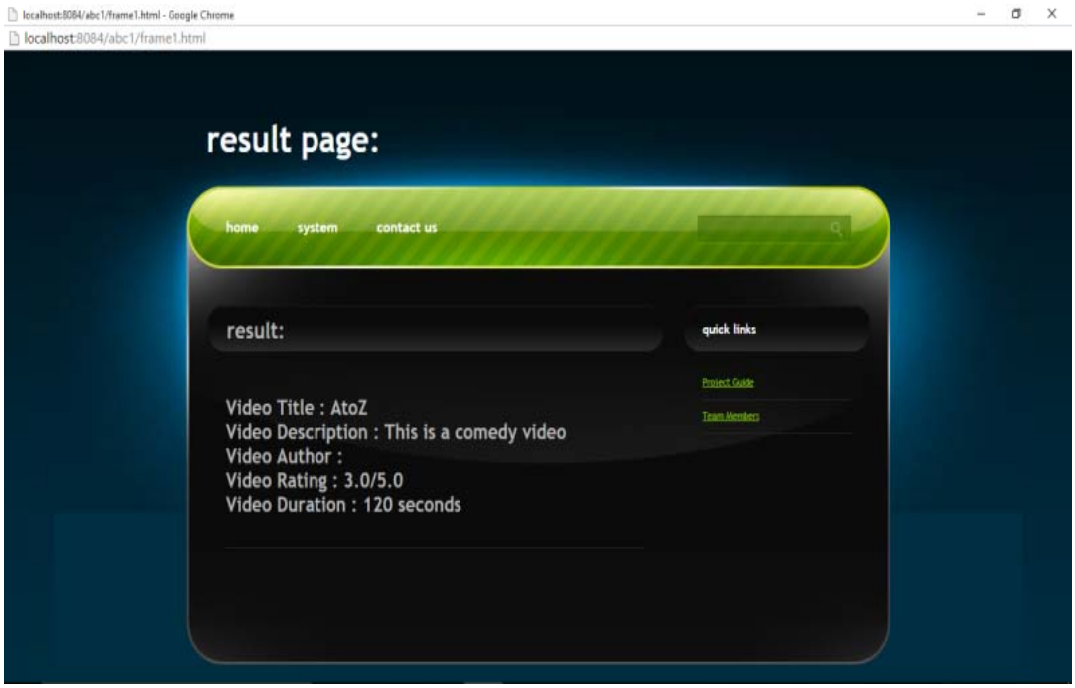

Figure 5 Screenshot of Result Page

The screenshots of figure 3, figure 4 and figure 5 shows the different pages of the web based video retrieval system. In the figure 3 we have shown how to upload the video clip for which the search result is desired.

Here the main thing is that when you upload the video clip then we are converting it into the images and storing it in the database. After completing this task we are comparing the images with our database which already contains the set of images.

Figure 4 shows the result page. After comparing the images from the database it will give the result from the best matches' i.e. name of the file.

V. FUTURE ENHANCEMENT

In our above work we have used the pixel by pixel comparison of the images. We can't say it's the very fast algorithm as it needs time to compare the whole database. So we need some much more efficient algorithm to compare the images.

Second thing which we can do is integrating the audio search. Right now we are doing this work with the byte code comparison. But by doing some more work on audio comparison we can make it better.

For the future enhancement of this work we are continuing our research broadly focusing on the audio to text conversion of the sound, and then comparing it to the subtitles of the videos. By doing this we don't need to convert or extract the whole video in the image format. If we are able to compare the text of subtitle then from that we can get the specific location also like our audio matched at 00:30:45 time (Time in HH:MM:SS format) . So from this we need to extract only that specified images from the video and we know that audio comparison is more fast and efficient than the image comparison. So I think this will be the more efficient and more reliable solution to the video search field.

REFERENCES

[1] Efficient and Effective State-based Framework for News Video Retrival', International Journal of Advancements in Computing Technology, vol. 2, no. 4, pp. 151-161, 2010.
[2] Honghai Liu, Shengyong Chen and N. Kubota, 'Intelligent Video Systems and Analytics: A Survey', IEEE Transactions on Industrial Informatics, vol. 9, no. 3, pp. 1222-1233, 2013.
[3] Shih-Fu Chang, W. Chen, H. Meng, H. Sundaram and Di Zhong, 'A fully automated content-based video search engine supporting spatiotemporal queries', IEEE Trans. Circuits Syst. Video Technol., vol. 8, no. 5, pp. 602-615, 1998.
[4] Corrigendum: A bibliometric study of Video Retrieval Evaluation Benchmarking (TRECVid): A methodological analysis', Journal of Information Science, vol. 38, no. 2, pp. 201-201, 2012.
[5] C. Thornley, S. McLoughlin, A. Johnson and A. Smeaton, 'A bibliometric study of Video Retrieval Evaluation Benchmarking (TRECVid): A methodological analysis', Journal of Information Science, vol. 37, no. 6, pp. 577-593, 2011.
[6] A. Smeaton, P. Wilkins, M. Worring, O. de Rooij, T. Chua and H. Luan, 'Content-based video retrieval: Three example systems from TRECVid', Int. J. Imaging Syst. Technol., vol. 18, no. 2-3, pp. 195-201, 2008.
[7] Yovisto: Academic Video Search', Choice Reviews Online, vol. 51, no. 10, pp. 51-5333-51-5333, 2014.
[8] X. Wei, Y. Jiang and C. Ngo, 'Concept-Driven Multi-Modality Fusion for Video Search', IEEE Trans. Circuits Syst. Video Technol., vol. 21, no. 1, pp. 62-73, 2011.

[9] A. Dore, M. Soto and C. Regazzoni, 'Bayesian Tracking for Video Analytics', IEEE Signal Process. Mag., vol. 27, no. 5, pp. 46-55, 2010.
[10] C. Thornley, A. Johnson, A. Smeaton and H. Lee, 'The scholarly impact of TRECVid (2003-2009)', J. Am. Soc. Inf. Sci., vol. 62, no. 4, pp. 613-627, 2011.
[11] S. Chang and H. Zhang, 'Content-Processing for Video Browsing, Retrieval, and Editing', Multimedia Systems, vol. 7, no. 4, pp. 255-255, 1999.
[12] M. Thida, H. Eng, D. Monekosso and P. Remagnino, 'Contextual Analysis of Videos', Synthesis Lectures on Image, Video, and Multimedia Processing, vol. 6, no. 2, pp. 1-102, 2013.
[13] P. Rajendran and T. Shanmugam, 'A content-based video retrieval system: video retrieval with extensive features', IJMIS, vol. 2, no. 2, p. 146, 2011.
[14] G. Awad, P. Over and W. Kraaij, 'Content-Based Video Copy Detection Benchmarking at TRECVID', ACM Transactions on Information Systems, vol. 32, no. 3, pp. 1-40, 2014.
[15] M. Lee, 'Distributed video coding with video analytics information for video surveillance application', Electronics Letters, vol. 49, no. 20, pp. 1265-1266, 2013.
[16] Shih-Fu Chang, W. Chen, H. Meng, H. Sundaram and Di Zhong, 'A fully automated content-based video search engine supporting spatiotemporal queries', IEEE Trans. Circuits Syst. Video Technol., vol. 8, no. 5, pp. 602-615, 1998.

AUTHOR PROFILE

Tarun Sharma is doing his Masters of Technology in Computer Science and Engineering from VIT University, Vellore, Tamilnadu. He has one year of experience as an Associate Quality Analyst at Girnar Software (SEZ) Pvt. Ltd. Jaipur (Rajasthan). His research area includes Database, Data Mining, and Big Data. He has good knowledge of development tools like Net Beans, Visual Studio, SQL and many more. He has done training from CSIR-CEERI, Pilani. mailto:tarunsha009@gmail.com.

Chinta Rahul is doing his Masters of Technology in Computer Science and Engineering from VIT University, Vellore, Tamilnadu. His area of interest is Networking and Programming. He has good knowledge in development tools like Net Beans, Visual Studio, SQL and many more. mailto:chinta.rahul2@gmail.com.

Shaik Shahul is doing his Masters of Technology in Computer Science and Engineering from VIT University, Vellore, Tamilnadu. His research area includes Database, Data Mining. He has good knowledge in development tools like Net Beans, Visual Studio, SQL and many more.

M. Venkatesan is working as an Associate Professor in School of Computing Science and Engineering, Vellore Institute of Technology, Vellore, Tamilnadu. He is working in the area of spatial data mining. His research area includes databases, data warehouse, big data analytics, data mining and applications of data mining in disaster management. He has good knowledge in data mining tools like Clementine, rapid miner. He is the principal investigator in ISRO funded project on data mining in a landslide. He completed his BE, M. Tech and PhD in the area of spatial data mining. He published 25 papers in international journal and 30 papers in the reputed conferences.