



Content Based Video Retrieval with Frequency Domain Analysis Using 2-D Correlation Algorithm

Navdeep Kaur

CGC Group of College, Gharuan (Pb.)
India

Mandeep Singh

Associate Prof, Chandigarh University (Pb.)
India

Abstract— Content Based Video Retrieval is very interesting area in the field of multimedia information retrieval. Content based video retrieval involves retrieving user's choice related video clips from video database using their visual attributes such as color, texture, shape etc. These days, video is the most common way to exchange information with advancements in technology. When a user has to detect some useful objects from video, then he will have to watch complete long video, which leads to wastage of time and is very complex procedure. So, content based video retrieval is open area of research nowadays. To improve the performance of content based video retrieval system, it is very necessary to develop efficient video parsing and feature extraction algorithms. In the presented work, first the video is parsed into shots using edge detection algorithm and then shots are stored in a separate file. After detecting shots, key frames are extracted to represent video shots. Then this system is made to receive the user query in the form of image. Matching between user query and key frames is done using frequency domain analysis with 2-D correlation algorithm. This algorithm is compared with the existing technique of surf feature based point matching algorithm. It is observed that the proposed research work is providing more efficient results and also saves time.

Keywords— CBVR, frequency domain, 2-D correlation, video parsing, DCT, feature vector.

I. INTRODUCTION

In the context of multimedia, information retrieval such as images, text, audio or videos is an open research area. Various methods have been discovered to retrieve the effective and accurate information. Basically, there are two retrieval methods, which are: text based information retrieval and content based video retrieval. The First method is the text based retrieval- in which the images and videos are manually annotated with keywords or descriptors. Most of the search engines on the internet like Google tend to find the multimedia information by searching the textual labels attached with the multimedia data. The keywords are attached to the multimedia data in such a way that they tend to best describe the image or video itself based on their properties. Second method is content based information retrieval, which deals with retrieval of images or videos based on their visual contents from large database of videos. With the help of content based video retrieval, the user is able to retrieve important clips of video based on his demands rather than watching the whole video. So, content based video indexing and retrieval is promising area of research. There are three levels to describe visual features of image or video described below:

- Low level description
- Middle level description
- High level description

Low level description of the visual contents is based on color, texture and shape features. The low level visual features of an image are directly related to the image content. Next level is the middle level description which is concerned with the background and spatial attributes of the concerned object. The high level feature representation is dealing with human brain and perception. Examples are events, scenes and human thinking such as emotions [17]. This type of description is very difficult to map to some mathematical model. A lot of work has been done in the field of content based image and video retrieval systems. Some of the important commercial systems are QBIC, Excalibur, Virage. Experimental CBVR systems have also been developed by academic institutes such as Photobook, Netra, Visualseek and Chabot to check new technologies.

II. ARCHITECTURE OF CBVR SYSTEM

The main functionalities of content based retrieval system are best described in terms of its architecture. Content based video retrieval is also known as by the name of content based visual information retrieval i.e. CBVIR. The main considerations in a content based video retrieval system is on storage of video data, extracting its features properly, providing suitable user-interface, choosing a proper similarity matching algorithm. Video database is maintained by storing the video with any type of extension such as .mpeg, .mp4, .avi etc. Video parsing, also known as video structure analysis- deals with structured hierarchy of video, which includes detecting video clips, scenes, shots and key-frames from video. The main components of CBVR system are following:

1. Video database
2. Video parsing
3. Feature extraction
4. User query
5. Similarity Matching
6. Video browsing and feedback

Video indexing and retrieval is a promising concept to save time in the conditions when one is required to retrieve some important information from a long video and is not interested to watch the whole video which is very time consuming. The content based video retrieval offers a solution to such type of problem.

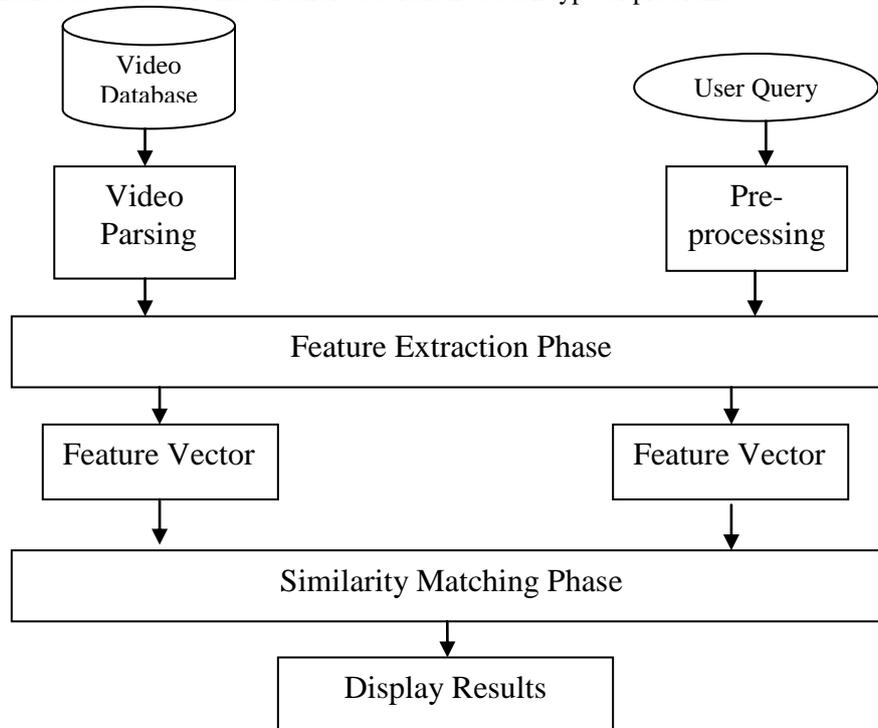


Fig 1: Architecture of CBVR

III. RELATED WORK

Content based video retrieval has been one of the challenging research areas in the field of image and video processing. Abundant research has already been done in this field and it is still being done as it is a vast field. A brief overview of the previous research work done in the field of content based video retrieval is presented in the following paragraphs:

Color, texture and edge features were used for video retrieval purpose in this paper [4]. In the first phase, the shots are detected based on measure of entropy based texture characteristics. Then key frames were extracted. The feature vector of key frame features in the form of RGB color value and entropy value was created. The entropy features are retrieved from GCL matrix. In the second phase of system, the user enters a video frame as a query. Then the entropy feature of this frame is calculated by the system. The authors had proposed the combination of various features such as entropy-edge based feature method, entropy-color based feature method. The authors used the entropy –edge feature based method in the construction domain, news domain and entropy-color feature based method in nature domain and e-learning domain and provided the accurate results.

In [5], the authors have proposed a new framework to extract video shots and key-frames from video database. They combined temporal multi-resolution analysis (TMRA) work using SVM (Supported Vector Machines) to classify the video frames into normal frames, then clustered the similar frames into different shot categories. Blocked color histogram in the YUV color space was retrieved as a feature from temporal multi resolution analysis. Wavelet transform was applied on video first and then it was fed to support vector machine classifier. Then CUT and GT were detected and flash detection was performed in the end. The key frames were extracted using minimal and maximal points of wavelet coefficient with blocked color histogram.

In [6], the authors have proposed an algorithm to overcome the gap between low level features and high level features. The threshold adopted by the authors is adaptive i.e. its value can vary depending upon the requirements. After storing videos into database, the video segmentation is done to retrieve video shots using some threshold value. Then frame features were retrieved after extracting key-frames. The feature (color, shape, texture) and high level features (object annotations). Then the videos in database were represented using their feature vectors. Here, graphical annotations were used to show the regions of interest. The features of query image and the video are compared using dynamic programming approach.

In [8], authors have proposed an object-based method for the extraction of key frames from the visual features of video. The key frames are extracted by combined effect of segmentation of object approach with the color, shape and texture features. First frame of each shot is presumed as key frames. Then further key frames are extracted based upon the similarity distance between consecutive frames. The fuzzy segmentation procedure is used with the watershed algorithm. The shape based features are used to classify objects such as thin and elongated objects will fall in different categories. This method is based on spatial segmentation of each frame in order to detect the important events. After performing the experiment, the authors have shown that compression ratio and peak signal to noise ratio were reduced to a greater extent.

This system retrieves similar videos based on a local feature descriptor called SURF (Speeded Up Robust Feature) [9]. SURF is a robust local feature descriptor that detects the points of interest from an image, and represents these points by a feature vector. The authors of this paper discussed two main points. First, they implemented the system using SURF descriptor. Second, the authors have tried to reduce the dimension of feature vector generated by SURF feature. Otherwise surf feature vector occupies a large space in database because of its high dimension. Stochastic dimensionality reduction method is used to reduce the size of this feature vector. This reduced vector set is trained into its model vector. The size of surf feature vector is $64 \times M$. Shot detection is determined from histogram difference. Fuzzy K nearest neighbour is used as minimum distance classifier. This system also analyses the performance efficiency of the low dimensional SURF descriptor.

IV. ALGORITHM

We are developing a content based video retrieval system, in which the user can extract the objects of his interest from the video by entering a query image containing the object of interest. The algorithm comprise of following steps:

A. Video Acquisition and enter query image:

At first, the video database is maintained with the video having .mp4, which can be of varying duration.

B. Algorithm for shot detection:

Following steps are used for the proposed algorithm:

- I. *Read video frames:*
The database is created with a video having format .mpeg, .mp4 etc., which will be of varying durations or a long video. Then video frames are read one by one to check their dimensions.
- II. *Define block size and create ROI rectangle:*
After processing the dimensions of video frame, the frames are divided into number of blocks, so as to maintain accuracy and increase speed. The block size is variable. The ROI rectangle indices are created for each block of frame.
- III. *Edge detection from each frame:*
Edge detector system object is created using sobel operator to detect edges from each image.
- IV. *Calculate mean from edge detected frame and calculate absolute difference between consecutive frames:*
Mean value is extracted from each block of every frame after edge detection. Then absolute difference of means of consecutive frames is compared until last frame of video.
- V. *Detect shots from compared absolute difference by setting a threshold value:*
The threshold value is adaptive and the shots are defined if there are significant changes in more than one block of consecutive frames. Number of retrieved shots varies by varying the value of threshold.
- VI. *Display the identified shots and the information about number of shots and frames detected:*
The identified shots are displayed with their edge information. Only 3 frames are used to represent a video shot. Further the number of shots and frames that have been detected are also displayed to the user.
- VII. *Store video frames from detected shots in a video file:*
A new video file is created by the system and it will store the frames retrieved from the shots retrieved for further processing.

C. Key frame extraction:

Key frames are obtained after detection of shots. Key frame extracted should be sufficiently able to represent semantics and characteristics of a complete shot. There can be more than one key frame to represent a single shot. In this system, we have chosen the central frame of a shot to represent as a key frame.

D. Conversion into frequency domain using discrete cosine transform:

The key frames and the query image are operated on by discrete cosine transform to use them into frequency domain. Discrete cosine transform is a mathematical operator, used to convert data from spatial domain into frequency domain. The general equation for a 2D (N by M image) DCT is defined by the following equation [15]:

$$F(u, v) = \left(\frac{2}{N}\right)^{\frac{1}{2}} \left(\frac{2}{M}\right)^{\frac{1}{2}} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} \Lambda(i) \cdot \Lambda(j) \cdot \cos \left[\frac{\pi \cdot u}{2 \cdot N} (2i + 1) \right] \cos \left[\frac{\pi \cdot v}{2 \cdot M} (2j + 1) \right] \cdot f(i, j)$$

Here, we can see that discrete cosine transform uses only cosine functions and real coefficients.

E. Frequency domain analysis done with 2-D correlation algorithm:

Matching between query image and the corresponding video key frames is using 2-D correlation algorithm. Here, 2-D Correlation between two matrices A and B is computed as follows [16]:

$$r = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{\left(\sum_m \sum_n (A_{mn} - \bar{A})^2 \right) \left(\sum_m \sum_n (B_{mn} - \bar{B})^2 \right)}}$$

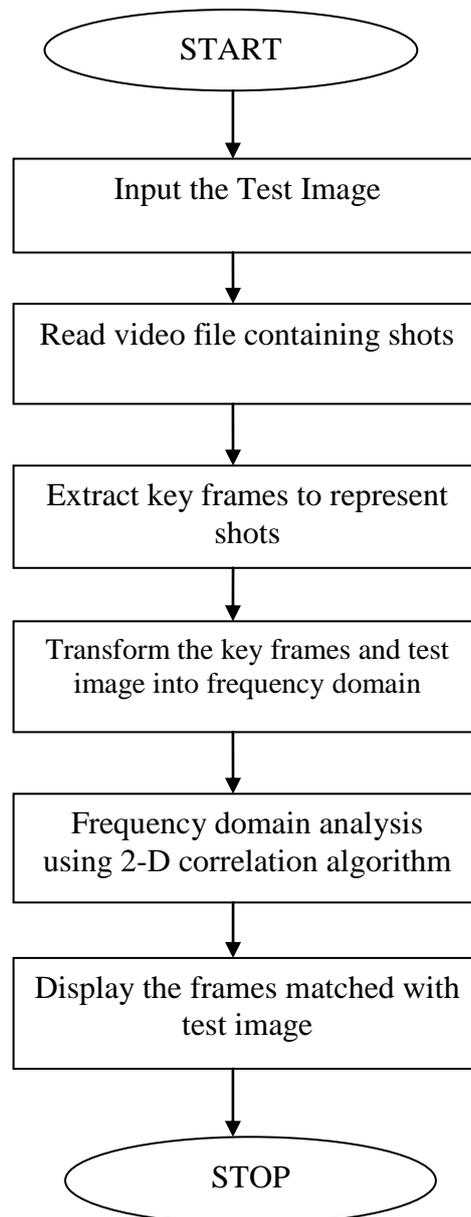


Fig 2: Flow Chart of proposed system

F. Display the objects from video that matched with the query image

After frequency domain analysis using 2-D correlation algorithm, those frames are displayed to user as output whose properties matched to query image with maximum extent using 2-D correlation algorithm.

V. RESULTS

We have tested this algorithm on different types of videos and efficient results were produced. Here, the user enters a query image containing an object, and our system searches that object throughout the whole video. The matched frames are displayed to the user as a result.

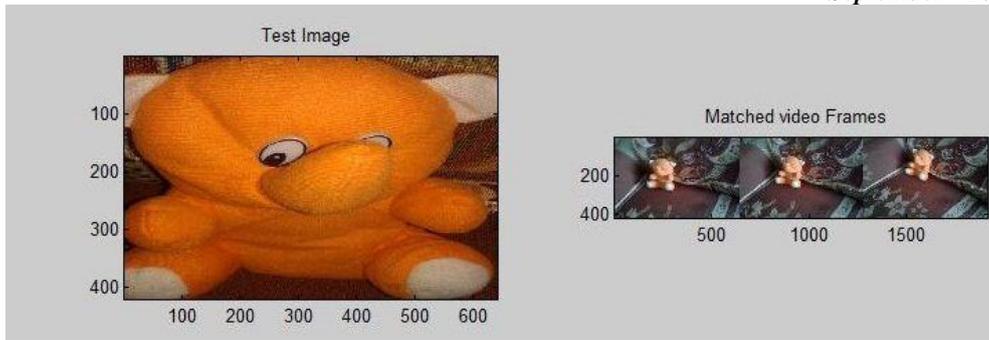


Fig 3: matched video frames with query image

Here, the system can also detect the frame number from which the desired object has detected by matching algorithm. We have also compared the performance of our proposed research work (frequency domain based correlation algorithm for feature extraction) with the existing technology of feature extraction using surf feature based point matching algorithm. It has been observed that our proposed system is producing better results than the later.

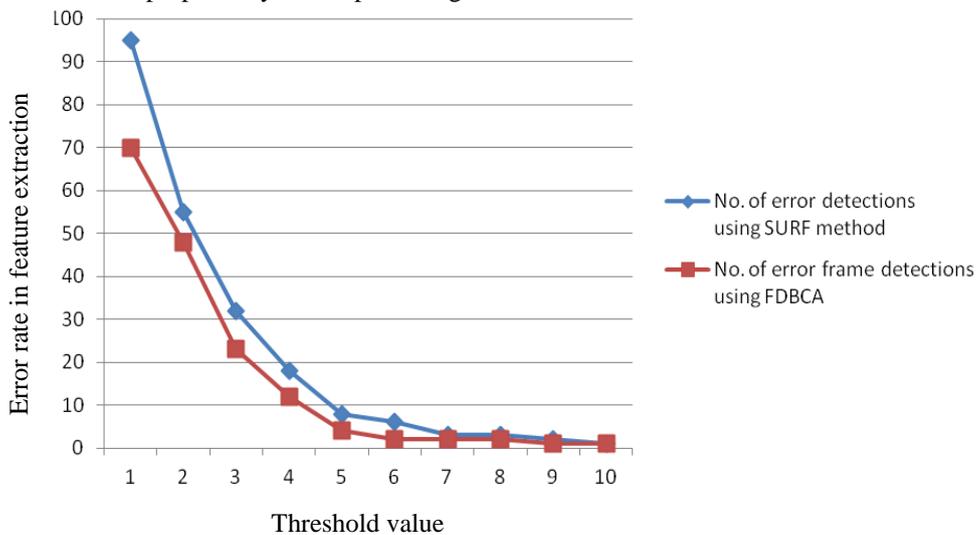


Fig 4: graph showing performance comparison between different techniques

VI. CONCLUSION

Content based information retrieval is wide and active area of research. Although content based video retrieval is very complex task because videos consist of very rich data and have very long duration. The proposed work has done feature extraction from video with frequency domain analysis using 2-D correlation matching algorithm. Shot detection has been carried out using edge detection algorithm. The main purpose of this system is to detect all the objects from video that matched with the user's query image. The comparison of feature extraction using surf feature and using frequency domain analysis has been carried out. Our system is retrieving results accurately. The accuracy of system changes with the varying value of threshold and size of video.

REFERENCES

- [1] Bay et al., SURF: "Speeded Up Robust Features", *Computer Vision and Image Understanding (CVIU)*, Vol. 110, No. 3, pp. 346--359, 2008.
- [2] David L., "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, Vol.60, No. 2, 2004.
- [3] Flickner, M et al , "Query by image and video content: the QBIC system" , *IEEE Computer* Vol.28, No.9, 23-32, 1995.
- [4] Kamde P. et al., "Entropy supported video indexing for content based video retrieval", *International Journal of Computer Applications*, Volume 62, No.17, January 2013.
- [5] Feng H. et al., "A new general framework for shot boundary detection and key frame extraction", *ACM*, 1-59593-244-5, 2005.
- [6] Elimnir H. et al., "Multi feature content based video retrieval using high level semantic concept", *International Journal of Computer Science Issues*, Vol. 9, Issue 4, No 2, July 2012.
- [7] Kaur N., Singh M., "Content (color) based image retrieval using RGB component Analysis", *1st National Conference on Information Technology and Cyber Security/Vol.1/PP 171-174/ITCS13/33*, 2013.
- [8] Barhoumi W. et al., "On-the-fly extraction of key frames for efficient video summarization", *Science Direct*, Vol. 4, pp. 78-84, 2013.

- [9] Asha S. et al., "Content based video retrieval using surf descriptor", International conference on advances in computing and communications, DOI 10.1109/ICACC.2013.49, 2013.
- [10] R.Gonzalez and R.E. Woods, Digital Image Processing, Prentice Hall, 2011.
- [11] Kaur N., Singh M., "Content based video retrieval using time based activity", IJARCSSE, Vol. 4, No. 6, 2014.
- [12] Liu G, Zhao J. Key frame extraction from MPEG video stream. Int Symp on Information Processing; 423-427, 2010.
- [13] T. Gevers, and A.W.M.Smeulders, "Pictoseek: Combining color and shape invariant features for image retrieval," IEEE Trans. on image processing, Vol.9, No.1, pp102-119, 2000.
- [14] Patel B., Meshram B., "Content based video retrieval systems", International journal of ubi computing, Vol. 3, No. 2, 2012.
- [15] Particle, "A very basic introduction to time/frequency domain", 2004.
- [16] www.mathworks.in
- [17] Heba A. et al. , "A new approach in content-based image retrieval using fuzzy", © Springer Science + Business Media,2008, Vol. 40, pp.55-66, DOI 10.1007/s11235-008-9142-9.
- [18] Weiming hu, Nianhua Xie, Li Li, Xianglin zeng, "A Survey on Content Based video indexing and retrieval", IEEE transactions on System, man and cybernetics, Vol. 41, No.6, 2011.