

# Video Retrieval by Extracting Key Frames in CBVR System

Miss. Poonam V. Patil<sup>1</sup>, Prof. S. V. Bodake<sup>2</sup>

P.G. Student, Department of Computer Engineering, PVPIT College of Engineering, Bavdhan, Pune, India<sup>1</sup>

Assistant Professor, Department of Computer Engineering, PVPIT College of Engineering, Bavdhan, Pune, India<sup>2</sup>

**ABSTRACT:** Due to the availability of increased internet bandwidth and transfer media, the multimedia objects usage is widely used. daily around millions of videos are being uploaded through various sources. it needs some kind of retrieval technique which will relieve users from tracking videos relatively similar manually. so users are shifted from text based video retrieval to content based video retrieval (cbvr). in videos large number of frames forms a scene. this scene includes repetition of nearly same frames with slight differences, which increases the storage space and decreases the performance in video processing. here instead of searching the whole set of frames in videos, only selected key frame are used for further processing. we first use the dual-threshold method to segment the videos into segments with homogeneous content and extract key frames from each segment. sift features are extracted from the key frames of the segments. then, we propose an svd-based method to match two video frames with sift point set descriptors .to obtain the video sequence matching result, we propose a graph-based method. it can convert the video sequence matching into finding the longest path in the frame matching-result graph with time constraint.

**KEYWORDS:** Content based video retrieval (CBVR), key frames, near-duplicates, dual-threshold method, feature extraction SVD, graph-based matching

## I. INTRODUCTION

Most of the multimedia search systems rely on available metadata or contextual information in text form. These all challenges motivate us to present an optimized key frame extraction for detection of near duplicates in Content Based Video Retrieval. Now-a-days, server space is becoming a major issue for high level organizations to maintain enormous amount of data. Due to the availability of higher bandwidth in Internet and multimedia processing technologies, maximum amount of content is being spread among large number of users. this concept of multimedia item processing have been demonstrated .As a part of enhancement in this process users are now shifted from text to content, which includes the semantic or content description of the query being processed.

In CBVR there is requirement of feature extraction and many methods, such as bag of features (BoFs) or visual word for video copy detection, locality sensitive hashing (LSH), hierarchical indexing structure for efficient video retrieval, and so on, have been proposed for efficient video search. However, by using these indexing methods, the temporal information of the SIFT feature points in different frames will be lost.

The concept of near-duplicate elimination for web video search with content and context was first introduced by Xiao Wu, et al. [1]. With the exponential development of online networking, there exist tremendous quantities of near-duplicate web videos, extending from straightforward arranging to complex blend of distinctive altering impacts. Notwithstanding the bottomless video content, the social web gives rich arrangements of setting information connected with web videos, for example, thumbnail picture, time length of time duration. video copy detection method proposed Arun Hampapuret al. [2] Content-based copy detection (CBCD) is a reciprocal way to deal with watermarking for distinguishing duplicates of media. Watermarking depends on the capacity to distinguish from a duplicate an unmistakable example that was brought into the first media. CBCD strategies recognize duplicates by measuring separations between sub-distances based signatures removed from the first and the duplicate. The discriminating test in content-based copy detection is the outline of signatures, which are invariant to the distinctions in quality crosswise

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

over duplicates of the same media. In [3], the authors propose in this paper a segmentation and graph-based video sequence matching method for video copy detection. Specifically, due to the good stability and discriminative ability of local features, use SIFT descriptor for video content description. In [4] a video retrieval system is proposed, which demonstrates the non semantic approach used with the help of query by example, query by sketch. When content are introduced it has been suggested to replace this query by keywords and query by natural language. . In [5], it is demonstrated that video retrieval techniques has wide scope which is utilized in the work of video summarization, video segmentation, video colorization and video analysis.

The paper is organized as follows. Section II includes the general structure of video as well as general background techniques of video retrieval such as Content-Based Video Retrieval and Video Copy Detection System with framework structure. Section III include problem definition . Section IV consists of Proposed system architecture with detail explanation of each method which used in proposed system to improve performance of video retrieval .Section V includes a conclusion of proposed system , direction to continue this work in video applications. In the recent years, researchers are continuously focusing on solving the real problems .Section VI includes future scope of proposed system it is demonstrated that video retrieval techniques has wide scope which is utilized in the work of video summarization, video segmentation, video colorization and video analysis.

## II. RELATED WORK

Lots of work has been done in video The videos are being processed in terms of video indexing, shot boundary detection, video classification, concept based video retrieval etc. And now this is in the progress of switching from concept to content based video retrieval. The generic framework of content based video retrieval system is as

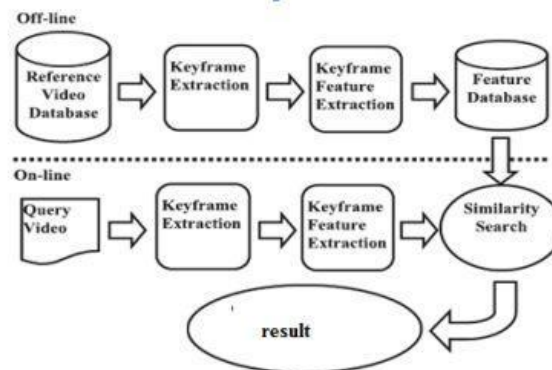


Fig 2.1. A framework on Content Based Video retrieval System

The framework of content-based video detection. It is composed of two parts: 1) An offline step. 2) An Online step.

1) An offline step: Key frames are extracted from the reference video database and features are extracted from these key frames. The extracted features should be robust and effective to transformations by which the video may undergo. Also, the features can be stored in an indexing structure to make similarity comparison efficient.

2) An online step: Query videos are analysed. Features are extracted from these videos and compared to those stored in the reference database. The matching results are then analysed and the detection results are returned.

The framework accepts input as a video sequence with the standard structure consisting of shots, key frame and scene. A video is the sequence of shots played in terms of certain frame rate and then scenes are played in fixed interval of time. The existing work in CBVR can be summarized with three major points contained in the video data: 1) video structure analysis; 2) feature extraction in terms of low level features and high level features; 3) content based query formulation; 4) similarity measurement; 5) accuracy and precision, recall calculation of retrieved videos.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

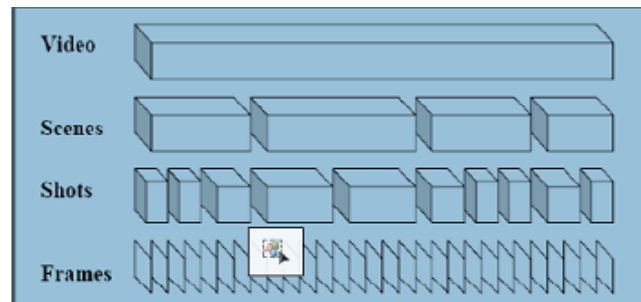


Fig 2.2. Video Segmentation

A video may have visual information as well as audio information, both representing the more semantic to the scenes. The application range of videos is increasing with ease of use. Videos are structured in the hierarchical order of video clips, shots, scenes and frames. A shot is sequence of consecutive frames. A frame is visual input data which is captured by the camera action with some time interval. In the different effects of camera action have been described. Videos are the dense structure of frames displayed at a certain rate of frames per second. This frame rate decides the bit rate of intensity values with fixed amount of height and width in the form of pixels values. The CBVR makes the video processing easy with the consideration of shots, scene, frames and finally key frames.

Shot contain thousand of frames but only no. of frames contain more information. These selected frame contain maximum amount of content, very special imp for content analysis on video signal .These type of frame is called Key frame. Key frame extraction is more essential in the video processing because of less space required for key frame, the reason that actual work done is only on important content of information, not on the supplementary frames(near duplicate frames). There by it reduces the redundancy and gives improved performance.

Near duplicate is nothing but temporal redundancy of visual information of continuous video frames. Copy Video is also the duplicate video, copy video is: A video V1, by means of various transformations such as addition, deletion, modification(of aspect, color, contrast, encoding, and so on), cam cording, and so on, is transformed into another video V2,then video V2 is called a copy of video V1.

### III. PROBLEM DEFINITION

To improve performance and speed of CBVR by an optimized Key frame extraction by eliminating redundant video frames in CBVR .For these purpose we used efficient algorithm and techniques on and retrieve video by detecting redundant video mean it is near duplicate or fully duplicate to input video in the process of CBVR .

### IV. PROPOSED APPROACH

With the rapid development and wide application of multimedia hardware and software technologies, each day tens of thousands of video data are generated and published. Among these huge volumes of videos, there exist large numbers of copies or near-duplicate videos. According to 27 percent redundant videos that are duplicate or nearly duplicate to the most popular version of a video search engines. It needs some kind of retrieval technique which will relieve users from tracking videos relatively similar manually. So users are shifted from text based video retrieval to content based video retrieval (CBVR).In videos large number of frames forms a scene. This scene includes repetition of nearly same frames with slight differences, which increases the storage space and decreases the performance in video processing.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

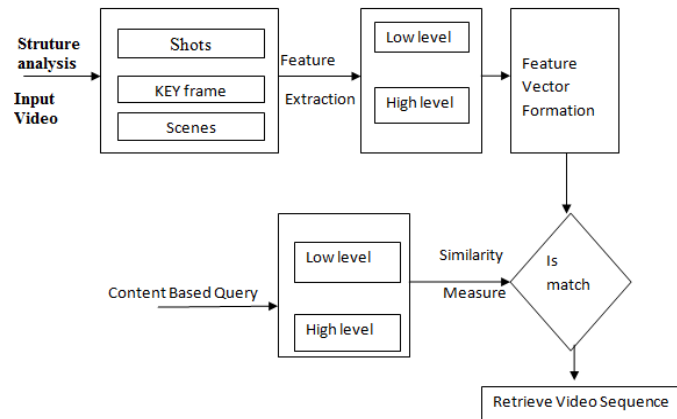


Figure 4.1: System architecture

## Step 1: Video Segmentation:-

In video segmentation divide the video into frames, for video segmented is done commonly by using the shot boundary detection with the different distance measurement algorithm . The video shots can be obtained by the shot boundary detection method. But to detect near duplicates frames in video data set , used Auto-dual threshold method. A video retrieval by removing near duplicates in CBVR These method aims to eliminate the near-duplicate frames along the video time direction; it does not take into account the concept of the shot, also does not require post processing to obtain the actual shots.

## Step 2: Key frame Selection by eliminate Redundant Video Frames:-

Normally, visual information of video frames is temporally redundant. So, video sequence matching is not necessarily to be carried out using all the video frames .By using the auto dual-threshold segmenting method, continuous video frames can be segmented into temporally continuous and visually similar video segments. Three frames are extracted from each video segment, which are the first frame, the key frame and the last frame of this segment. The key frame is determined by the frame that is the most similar to the average frame(i.e., the average feature value of all the frames within the segment). The key frame is used for video sequence matching, while the first and the last frames for accurately determining the segment location for copy detection and assisting matching. Each segment is assigned a continuous ID number along the time direction.

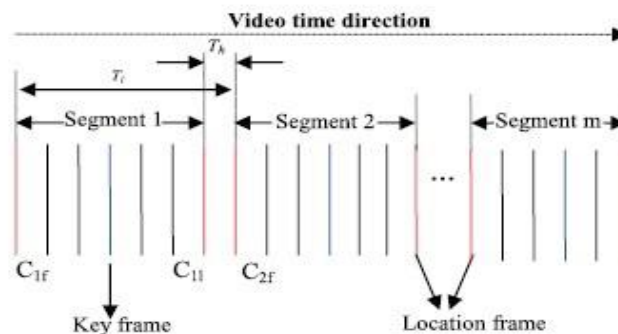


Figure 4.2: Eliminate redundant video frames, select key frames.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

## Step 3: Feature Extraction:-

In video analysis the features applicable to images are considered same for video signal . Features are extracted for the key frame and stored into feature vector. Features are of two types that are spatial ( low level ) and temporal( high level ). Spatial features are further classified as color, shape and edge. Temporal features are also further classified as motion and audio. So, video features can be classified as low level and high level features. Low level features describe the color, shape and high level features describes texture, pattern including the major part of information. It is very tedious to bridge the gap between low level and high level features .These features need to be extracted to work on with the content. Unlike the pixel values from each image frame of a video the feature must be computed and extracted for the feature comparison in retrieval process. Feature extraction occurs in two forms based on color ,shape

1) Features with Color: Color-based feature are generally extracted with color histograms and correlograms. Use of different types of color space models such as RGB, YCbCr makes the difference in feature extraction related with color. HSV provides different kinds of feature extracted. For Color feature are using Local Color Histogram (LCH) and Global Color Histogram (GCH), Average RGB Algorithms.

2) Feature with Shape: In this feature extraction the edges are the main elements of interest wherein the shape is measured on the basis of edge differences. In this feature extraction procedure using shape is demonstrated in different color spaces. For Shape and Edge are using Eccentricity Algorithm

## Step 4: Similarity matching :-

In singular value decomposition method, match the SIFT feature point sets based on SVD. The SVD method has been widely used in pattern recognition, data compression, signal processing, and other fields. Specially, the goal of this method is to use the SVD method to reduce and correct the wrong match between the two points in two SIFT feature sets. The method focused on the one point-to-one point correspondence. However, SVD method to measure the similarity between two SIFT feature point sets, and emphasize the similarity of frame-to-frame

## Step 5: Video Sequence Matching:-

In a new graph-based video sequence matching method in matching result graph, as per query video .

## V. RESULTS

By considering of the approaches referred in the previous sections, the motion vector, edge histogram, color histogram and texture features are extracted for all the video shots in the database and stored in the feature library. Based on the input query clip, in video retrieval system, retrieve the related videos from the video database. The above mentioned features are evaluated for the query clip and compared against the features in the feature library. With the help of a similarity measure, the comparison of the features is achieved.. Finally, based on the distance measures the video sequences are sorted in ascending order and the similar videos are retrieved.. Clearly, the similarity measure plays a major role as the original feature space in deciding “close” or “far”.



(i)

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015



(ii)

Figure 5.1 Results of the proposed CBVR system: (i) input query clip and the (ii) retrieved video clips

## VI. CONCLUSION

In CBVR key frame extraction and elimination of near duplicates will improve the video storage, search and retrieval speed. Using the Auto-dual threshold method for elimination of near duplicate frame extraction the number of frames can be reduced to 30-40 percent. A dual-threshold method to eliminate redundant video frames and use the SVD based method to compute the similarity of two SIFT feature point sets. These will definitely help to achieve better accuracy up to 35-36 percent.

## VII. FUTURE SCOPE

In this paper this method of key frame extraction is applied only in one pass. It can be implemented in multiple passes and improved the performance of video retrieval process. As key frames are the collections of only important content data, once extracted from the videos can be utilized anywhere for video retrieval such as video summarization, video colorization, video annotation etc.



## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

### REFERENCES

- [1] Xiao Wu, Chong-Wah Ngo, Alexander G. Hauptmann, "Real-Time Near-Duplicate Elimination for Web Video Search With Content and Context", IEEE Transactions on multimedia, vol. 11, no. 2, february 2009
- [2] Arun Hampapur and Ruud M. Bolle "Comparision Of Distance Measure For Video Copy Detection" IEEE International Conference on Multimedia and EXPO. 2001
- [3] Hong Liu, Hong Lu, Member, IEEE, and Xiangyang Xue, Member, IEEE "A Segmentation and Graph-Based Video Sequence Matching Method for Video Copy Detection", IEEE transactions on knowledge and data engineering, vol. 25, no. 8, august 2013.
- [4] YANG Shuping, LIN Xinggang, "Key Frame Extraction Using Unsupervised Clustering Based on a Statistical Model", Tsinghua Science and Technology\ April 2005, 10(2): 169 - 173 Applications And Reviews, Vol. 4i, No. 6, November 20i 1, pp no 797-8i9.
- [5] Kalpana Thakre, Archana Rajurkar and Ramchandra Manthalkar, "An Effective CBVR System Based On Motion, Quantized Color" Science & information Technology (IJCSiT), Vol 3, No 2, April 2011.
- [6] Sudeep D. Thepade , Ashvini A. Tonge "An Optimized Key Frame Extraction for Detection of Near Duplicates in Content Based Video Retrieval" International Conference on Communication and Signal Processing, April 3-5,2014, India
- [7] Sanjoy Ghatak, Debotosh Bhattacharjee "Extraction of Key Frames from News Video Using EDF, MDF AND HI Method for News Video Summarization" In- ternational Journal of Engineering and Innovative Technology (JEIT) Volume 2, Issue 12, June 2013
- [8] Weiming Hu, Senior Member, IEEE, Nianhua Xie, Xianglin Zeng, and Stephen Maybank , "A Survey on Visual Content-Based Video Indexing and Retrieval", IEEE Transactions On Systems, Man, And CyberneticsPart C: Applications And Reviews, Vol. 41, No. 6, November 20i i, pp no 797-819.
- [9] Xiao Wu, Wan-Lei Zhao, Chong-Wah Ngo, "Near-Duplicate key frame Retrieval with Visual Keywords and Semantic Context", CiVR'07, July 9-11,2007, Amsterdam, The Netherlands.
- [10] R.T. Minu, K.K. Thyagarajan, "Scrutinizing the Video And Video Retrieval Concept", international Journal oj Soft Computing and Engineering (JJSCE), ISSN: 2231-2307, Volume-I, issue-5, November 2011