International Journal of Research in Management, Science & Technology (E-ISSN: 2321-3264)
Vol. 5, No. 3, December 17
Available at www.ijrmst.org

Video Summarization using key frame extraction

Ms.Aarti Kumthekar[1], Ms. Sneha Patil[2]
[1,2]*Electronics and Telecommunication Department, AGTI's DACOE, Shivaji University, Maharashtra, India*

[1]aartikumthekar@gmail.com
[3]sneharsha@gmail.com

***Abstract*—Recent years have witnessed enormous increase in the video data on the internet. This rapid increase demands efficient techniques for management and storage of video data In order to reduce the transfer stress in network and invalid information transmission, the transmission, storage and management techniques of video information become more and more important. Video summarization is one of the commonly used mechanisms to build an efficient archiving system. The video summarization methods generate summaries of the videos which are the sequences of stationary or moving images. Key frame extraction is a widely used method for video summarization. The key frames are the characteristic frames of the video which render limited, but meaningful information about the contents of the video. Key frame is the frame which can represent the salient content and information of the shot. The key frames extracted must summarize the characteristics of the video, and the image characteristics of a video can be tracked by all the key frames in time sequence. Key frame extraction is performed on shots in order to save computation time by avoiding the effort of using all the frames in the video. The key frames selected should be those frames that represent the most important content of the shot.**

***Keywords*— key frame , histogram , video, edge detection**

I. INTRODUCTION.

General video is rich in content and consists of 25 frames per second [1]. Hence a one hour video would contain around 25x60x60 frames. Most of these frames contain redundant information and thus key frame extraction is essential. Thus, the use of key frames reduces the amount of data required in video indexing and provides the frame work for dealing with the video content. A basic rule of key frame extraction is to discard the frames with repetitive or redundant information. To extract valid information from video, process video data efficiently, and reduce the transfer stress of network, more and more attention is being paid to the video processing technology. The amount of data in video processing is significantly reduced by using video summarization and key-frame extraction. In recent years, many algorithms of key frame extraction focused on original video stream. It can introduce processing inefficiency and computational complexity when decompression is required before video processing. The researchers have attempted to exploit various features for the extraction of key frames in videos. Some of the low level features which are commonly used include histogram, frame correlation, motion information and edge detection etc. used the color histogram difference between the current frame and the last extracted key frame to draw out key frames from the video.

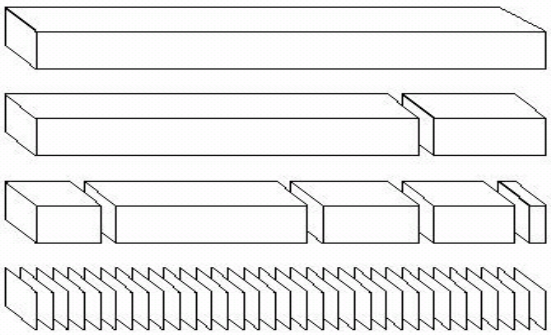

Fig 1: Structural hierarchy of a video

A video summary is a sequence of still or moving images, with or without audio. These images must preserve the overall contents of the video with a minimum of data. Still images chronologically arranged form a pictorial summary that can be assumed to be the equivalent of a video storyboard. Shot can be defined as a set of consecutive series of frames that constitutes a unit of action in the story. Practically, it is a part of the video that has been taken without interruption by a single camera. A video can be segmented into different units, such as frames, shots, or scenes. The structure of a video is shown in Figure. The complete moving picture in a video can be discretized to a finite image sequence, i.e., many still images. Each still image is called a "frame", which is the basic unit of the video. The image sequence is naturally indexed by the frame number. All the frames in one video have a same size and the time between each two frames is equal, typically 1/25 or 1/30 seconds. A video consists of a

sequence of images (often being called frames), which can be played consecutively at the speed of around 20 to 30 frames per second in order to view smooth motion. A video shot is defined as a series of interrelated consecutive frames taken contiguously by a single camera and representing a continuous action in time and space [11]. In general, shots are joined together in a process called editing to produce a video. The unbroken image sequence in a shot usually has consistent content. While scene is a more semantic notion, which is essentially a story unit.

For video summarization, a common first step is to segment the videos into temporal "shots," each representing an event or continuous sequence of actions. A shot represents a sequence of frames captured from a unique and continuous record from a camera. Then key frames are to be extracted. Video segmentation is the premise of key frame extraction, and key frames are the salient content of the video (key factors to describe the video contents). To index and retrieval a video, shot boundary detection is usually conducted to segments the video into shots by detecting boundaries between camera shots. A shot [10] is defined as the consecutive frames from the start to the end of recording in a camera. It shows a continuous action in an image sequence. There are two different types of transitions that can occur between shots, abrupt (discontinuous) also referred as cut, or gradual (continuous) such as fades, dissolves and wipes [1]. The cut boundaries show an abrupt change in image intensity or color, while those of fades or dissolves show gradual changes between frames.

II. LITERATURE SURVEY

Video summarization is one of the commonly used mechanisms to build an efficient video archiving system. The video summarization methods generate images (Money and Agius, 2008). Key frame extraction is a widely used method for video summarization. The key frames are the characteristic frames of the video which render limited, but meaningful information about the contents of the video (Li et al., 2001).The researchers have attempted to exploit various features for the extraction of key frames in videos. These features have been utilized in a variety of different ways. Some of the low level features which are commonly used include color histogram, frame correlation, motion information and edge histogram etc. (Jiang et al.,2009). Zhang et al. (1997) used the color histogram difference between the current frame and the last extracted key frame to draw out key frames from the video. Gunseland Tekalp (1998) compared the histogram of current frame with the average color histograms of the previous frames to compute the discontinuity value .A thorough survey of existing techniques reveals that the researchers have used many different visual features for the problem of key frame extraction.

Some of the techniques are as below:-

Pair wise pixel comparison [3] is a straightforward and simplest way, in which the number of pixels changed from a frame to the next is counted. When the total percentage of the pixels has changed, a shot is detected. In this algorithm individual pixels from frames are compared to find out frame difference. Pair-wise comparison evaluates the differences in intensity or color values of corresponding pixels in two successive frames. In this algorithm the pixel-wise difference algorithm gives quite acceptable results with adaptive thresholding. By considering difference between the difference signal values of adjacent frames is a worthwhile approach. In practice, it is observed that it is useful to reduce the difference signal with a threshold derived from the maximum and minimum difference signals over a small aperture. Even with the adaptive thresholding, the algorithm produces false alarms, if the shot before/after the shot boundary includes high motion activity. The reason can be explained as follows: The weakness of the pixel based features is the high sensitivity to the video content. It is difficult for this algorithm to understand whether the change in the continuity signal is due to shot boundary or due to disturbances/motion. In order to enhance the algorithm, adaptive thresholding can be used. However, the high level of activity in the images around shot boundary produces a larger difference signal than expected. As a result adaptively obtained threshold is larger. A threshold that is larger than expected results in missed shot boundary.

The main disadvantage of this method is its inability to distinguish between a large change in a small area and a small change in a large area. It is observed that cuts are falsely detected. The other disadvantage of this method is that it is quite sensitive to fast object movements and the camera motion - fast camera panning or zooming. Also it is sensitive to noise.

The motion activity is one of the motion features included in the visual part of the MPEG-7 standard [11]. It also used to describe the level or intensity of activity, action, or motion in that video sequence. The main idea underlying the methods of segmentation schemes is that images in the vicinity of a transition are highly dissimilar. It then seeks to identify discontinuities in the video stream. The general principle is to extract a comment on each image, and then define a distance (or similarity measure) between observations. The application of the distance between two successive images, the entire video stream, reduces a one-dimensional

signal, in which we seek then the peaks (resp. hollow if similarity measure), which correspond to moments of high dissimilarity. In this work, the extraction of key frames method based on detecting a significant change in the activity of motion is used. To jump 2 images which do not distort the calculations but we can minimize the execution time. First the motion vectors between image i and image i+2 is extracted then calculates the intensity of motion, we repeat this process until reaching the last frame of the video and comparing the difference between the intensities of successive motion to a specified threshold.

Likelihood ratio [12] is a region-based technique. It is a typical statistical difference method, which can be regarded as an extension to pixel difference. It can solve the problem of false detection due to small camera motions. Instead of comparing individual pixel, it compares the statistical characteristic, the so-called likelihood ratio, of the corresponding regions (i.e. blocks) in two successive frames. If the likelihood ratio is larger than a preset threshold, the region is regarded as being changed. A shot can be declared if a certain number of regions have changed. A shot boundary is found if more than a certain number of blocks have changed. It is less sensitive to camera and object motion and noise.

Eyuphan Bulut & Tolga Capin[13] proposes a new approach to find key frames in a motion captured sequence. Treat the input motion sequence as a curve, and find the most salient parts of this curve which are crucial in the representation of the motion behavior. We apply the idea of saliency to motion curves in the first part of our algorithm. Then in the second part, we apply key frame reduction techniques in order to obtain the most important key frames of the motion. This method is effective to a certain extent.

Zhuang et al. [14] proposed an unsupervised clustering method. A video sequence is segmented into video shots by clustering based on color histogram features in the HSV color space. For each video shot, the frame closest to the cluster centroid is chosen as the key frame for the video shot. Notice that only one frame per shot is selected into the video summary, regardless of the duration or activity of the video shot.

Zuzana Cernekova [15] proposed a new approach for shot boundary detection in the uncompressed image domain based on the MI and the joint entropy (JE) between consecutive video frames. Mutual information is a measure of information transported from one frame to another one. It is used for detecting abrupt cuts, where the image intensity or color is abruptly changed. A large video content difference between two frames, showing weak inter-frame dependency leads to a low MI. The entropy measure provides with better results, because it exploits the inter-frame information flow in a more compact way than a frame subtraction.

Ali Amiri [16] proposed a novel video summarization algorithm which is based on QR-decomposition. Some efficient measures to compute the dynamicity of video shots using QRdecomposition was utilize in detecting the number of key frames selected for each shot. Also, a corollary that illustrates a new property of QR-decomposition. This property was utilized in order to summarize video shots with low redundancy.

Hanjalic et al. [17] developed a similar approach by dividing the sequence into a number of clusters, and finding the optimal clustering by cluster-validity analysis. Each cluster is then represented in the video summary by a key frame. The main idea in this paper is to remove the visual redundancy among frames.

DeMenthon et al. [18] proposed an interesting alternative based on curve simplification. A video sequence is viewed as a curve in a high dimensional space, and a video summary is represented by the set of control points on that curve that meets certain constraints and best represent the curve.

Doulamis et al. [19] also developed a two-step approach according to which the sequence is first segmented into shots, or scenes, and within each shot, frames are selected to minimize the cross correlation among frames' features.

An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

III.PROPOSED METHODOLOGY

The proposed methodology is extracting efficient key frame for video summarization based on the block based Histogram difference and edge matching rate.

The method for key frame extraction consists of three steps:

1. Input a video stream, extraction of all the frames and calculate the block based histogram difference of each consecutive frame.
2. Choose the current frame as a candidate key frame whose histogram difference is above the threshold point.
3. Extract the edges of the candidate key frames and calculate the edge matching rate of adjacent frames.

If the edge matching rate is above average edge matching rate, the current frame is considered

as a redundant frame and should be eliminated from the candidate key frames.

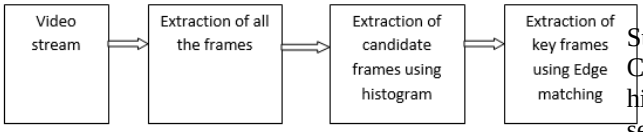

Fig 2: Block diagram of proposed work

The block diagram of this proposed work is shown in fig 2. The extraction of all the frames from video stream is the first step of the key frame extraction, which mainly refers to detecting the transition between successive shots. A shot represents a sequence of frames captured from a unique and continuous record from a camera. There are two different types of transitions that can occur between shots, abrupt (discontinuous) also referred as cut, or gradual (continuous) such as fades, dissolves and wipes.

After extraction of all the frames from video stream, key frames are to be extracted. Extraction of all the frames is nothing but shot segmentation is the premise of key frame extraction, and key frames are the salient content of the video. Thus summarization can be done using key frame extraction. The Histogram difference of every two consecutive frame is calculated so that key frames can be extracted and then the edges of the candidate key frames are extracted by Prewitt operator. Then, the edges of adjacent frames are matched. If the edge matching rate is above average edge matching rate, the current frame is deemed to the redundant key frame and should be discarded.

Algorithm for video summarization:

Frame extraction using histogram

Let F(k) be the kth frame in video sequence, k = 1,2,…., Fv(Fv denotes the total number of video). The algorithm of key frame extraction is described as follows.

Step 1: Partitioning a frame into blocks with m rows and n columns, and B(i, j, k) stands for the block at (i, j) in the kth frame;

Step 2: Computing x2 histogram [8] matching difference between the corresponding blocks between consecutive frames in video sequence. H(i, j,k) and H(i, j, k +1) stand for the histogram of blocks at (i, j) in the kth and (k +1) th frame respectively. Block's difference is measured by the following equation:

$$D_B(k,k+1,i,j) = \sum_{i=0}^{L-1}[H(i,j,k) - H(i,j,k+1)]^2/H(i,j,k)$$

….. (11)

Where wij stands for the weight of block at (i, j).

Step 3: Computing x2 histogram difference between two consecutive frames:

$$D(k,k+1) = \sum_{i=1}^{m} \sum_{j=1}^{n} w_{ij}\, D_B(k,k+1,i,j)$$

(12)

Step 4: Computing threshold aut… Computing the mean and … histogram difference over … sequence. Mean and standard va… as follows:

$$MD = \{\sum_{k=1}^{Fv-1} D(k,k+$$

…..

(13)

$$STD = \sqrt{\sum_{k=1}^{Fv-1} \frac{(D(k,k+1)-MD)^2}{Fv-1}}$$

…..

(14)

Step 5: Computing the difference between all the general frames and reference frame with the above algorithm

$$D_C(1,k) = \sum_{i=1}^{m} \sum_{j=1}^{n} w_{ij}\, D_{CB}(1,k,i,j), k = 2,3,4 \ldots \ldots$$

…… (15)

Step 6: Searching for the maximum difference within a shot:

Max(i)={

$$D_C(1,k)\}max, k = 2,3,4, \ldots \ldots F_{CN}$$

…… (16)

Step 7: Determining "Shot Type" according to the relationship between Max(i) and MD: Static Shot(0) or Dynamic Shot:

$$ShotType_C = \begin{cases} 1 \; if \max(i) \geq MD \\ 0 \; Others \end{cases}$$

……

(17)

Step 8: Determining the position of key frame: if ShotTypeC = 0, with respect to the odd number of a shot's frames, the frame in the middle of shot is chose as key frame; in the case of the even number, any one frame between the two frames in the middle of shot can be choose as key frame. If ShotTypeC = 1, the frame with the maximum difference is declared as key frame.

Extract Edges of the Candidate Key Frames

The candidate key frames obtained from the above treatment do well in reflecting the main content of the given video, but exist a small amount of redundancy, which need further processing to eliminate redundancy. As the candidate key frames are mainly based on the Histogram difference which depends on the distribution of the pixel gray value

in the image space, there may cause redundancy in the event that two images whose content are the same exist great difference from the distribution of the pixel gray value. For example, the substance content of images A and B don't change, but the two images are both identified as key frames as a result of the different gray value distribution, resulting in redundancy. So the use of the edge matching rate to match the edges of adjacent frames for eliminating redundant frames. The formula for calculating the edge matching rate is as follows

$$P(fi,fi+1) = s/n \quad \text{.....(18)}$$

Where, n=Max (fi, fi+1)

$$s=\sum_i^m \sum_j^n h(i,j)$$

$$h(i,j) = \begin{cases} 1, v_{fk}(i,j) = v_{fk+1}(i,j) \\ 0, Otherwise \end{cases}$$

Where vfk(i,j) and vfk+1(i,j) are the pixel values of the position (i, j) in the frame fk and the frame fk+1, respectively. M and n indicate the height and the width of the image, nfi and nfi+1 represent the number of the pixels on the edge of the frame fi and the frame f i+1 respectively. Assume the key frame sequence as {f1, f2, f3,…….fk} (the total number of the candidate key frames is k), we make use of the following steps to eliminate redundant frames:

Step 1: Use the edge operator to extract edges of the candidate key frames and obtain their corresponding edge images.
Step 2: Set j=2.
Step 3: Calculate the edge matching rate p(fj-1 , fj) between the current frame fj and the previous frame fj-1 w. If p(fj-1 , fj) is above average edge matching rate, the current frame fi will be marked as a redundant frame.
Step 4: j =j+1, if j>k, go to (e). Otherwise, return to (c) and continue processing the remaining frames.
Step 5: Remove the frames which have been marked as
Step 6: Redundant frames from the candidate key frames.
Step 7: The remaining candidate key frames are the ultimate key frames. With the edge detection and edge matching, we eliminate redundant key frames, improve the accuracy rate of the key frame extraction and reduce the redundancy.

IV. RESULTS

In proposed methodology the key frames are extracted from the video which can best reflect the video and give user as much information as possible. The way to achieve this is to first load the video then the video is divided into several frames and selecting one key frame which best describes the entire video. Proposed Key frame Extraction Algorithm is implemented with GUI (Graphical User Interface) based approach which provides easy user interaction which is shown in fig .The analysis of result is done using following methods:

1. Key frame extraction using Histogram
2. Key frame extraction using Edge matching rate

1. Key frame extraction using Histogram

First the video is loaded of 21 seconds containing the 317 total numbers of frames. Implemented project loads avi video as shown in fig 3.

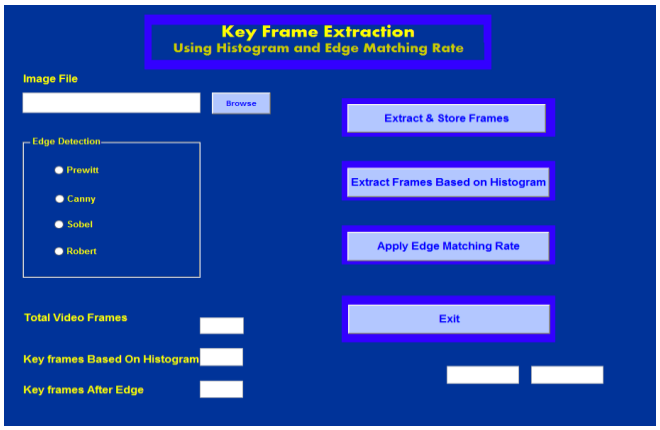

Fig 3: Selection of the video and extracting frames based on the histogram.

After selecting the video, based on histogram difference the frames of the two consecutive frames is extracted. These Frames are called as Candidate Key frames. The candidate key frames are 22 as shown in fig 4.

Fig 4: Frames extracted based on histogram

2. Key frame extraction using Edge matching rate

The edge matching rate is applied on these candidate key frames. Any of the four edge detector operator i.e. Prewitt, Canny, Sobel and Robert is used for extracting edges of the frames. By applying step key frame extraction using histogram, we are getting 22 frames then by edge matching rate, the key frames obtained are just 9.Thus video summarization is done with decreased in the required storage area.

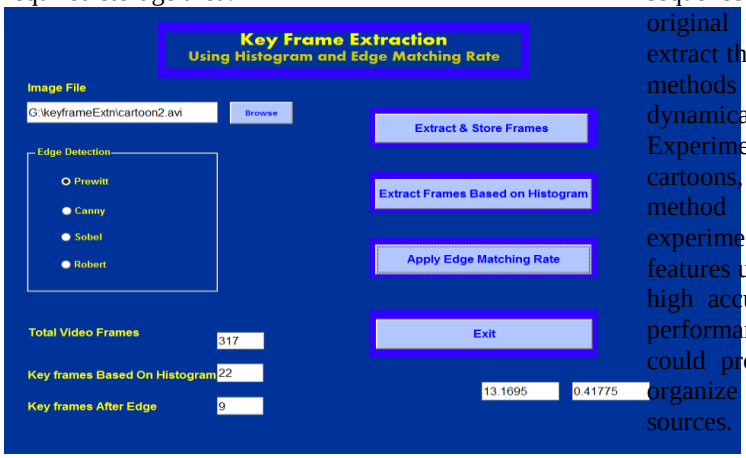

Fig 5: Selection of the operator and extraction of the frames based on edge matching rate

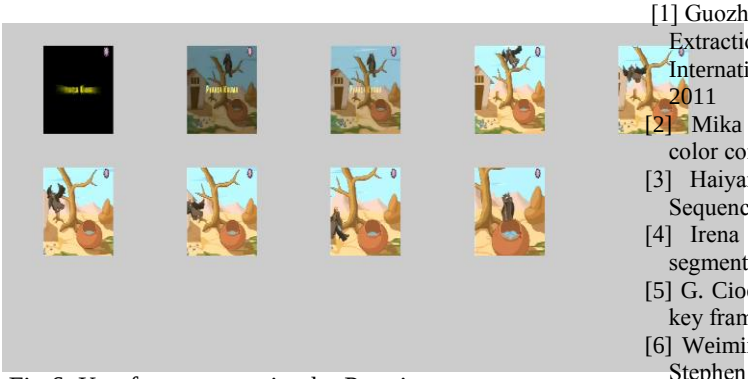

Fig 6: Key frame extraction by Prewitt operator

Table I shows the results of the proposed work for different edge detectors.

Video sequences	Total frames	Key frame extraction by histogram	Key frame extraction by edge detectors			
			Prewitt's operator	Canny operator	Robert operator	Sobel Operator
Video 1	262	5	4	3	3	3
Video 2	317	22	9	9	9	8
Video 3	344	7	7	5	5	5

It is observed that Prewitt is the best method for detecting the edges of the images and also one of the oldest methods for edge detection.

V. CONCLUSION

In this study, a novel algorithm of key frame extraction based histogram difference with edge matching rate is proposed. It compensates for the shortcomings of other algorithm and improves the techniques of key frame extraction. Automated video summarization was enabled by partitioning the video into meaningful segments and extracting key frames in the video. Key frame extraction aims to reduce the amount of video data, and the frame sequence must preserve the overall contents of the original video. Our proposed system is able to extract the key frames from most of the videos. The methods used are computationally simple and dynamically determines the number of key frames. Experiments on other type of videos such as cartoons, documentaries etc., have shown that the method is adaptive to the video content. The experimental results shows that the frame difference features using histogram and edge matching rate has high accuracy rate and low error rate. From the performance analysis, our video summarization could provide a fast, easy and effective way to organize and retrieve information from video sources.

REFERENCES

[1] Guozhu Liu and Junming Zhao "Key Frame Extraction from MPEG Video Stream" Third International Symposium on Information Processing, 2011
[2] Mika Rautiainenand David Doermann,"Temporal color correlograms for video retrival", 2002
[3] HaiyanXie," Key Frame Segmentation in Video Sequences", 2008
[4] Irena Koprinska Sergio Carrato,"Temporal video segmentation: A survey",feb 2000
[5] G. Ciocca1, R. Schettini"An innovative algorithm for key frame extraction in video summarization",
[6] Weiming Hu, NianhuaXie, Li Li, XianglinZeng, and Stephen Maybank ,"A Survey on Visual Content-Based Video Indexing and Retrieval". IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, Vol. 41, no. 6, november 2011
[7] Janko Calic and Ebroul Izquierdo,Multimedia and Vision Research Lab, Queen Mary, University of London,"Efficient key-frame extraction and video analysis",
[8] M. Ceccarelli , A. Hanjalic , R.L. Lagendijk'"A sequence analysis system for video databases",
[9] B. Furht and P. Saksobhavivat, *A Fast Content-Based Multimedia Retrieval Technique Using Compressed Data*, www.cse.fau.edu/~borko/paper_SPIE- 1.pdf.
[10] Keyframe Based Video Summarization Using Automatic Threshold & Edge Matching Rate. Mr. Sandip T. Dhagdi, Dr. P.R. Deshmukh, 2012.International Journal of Scientific and Research Publications, Volume 2, Issue 7, July 2012 ISSN 2250-3153

[11] Alan Hanjalic, "Shot-boundary detection: unraveled and resolved?" IEEE transactions on circuits and systems for video technology, vol. 12, no. 2, pp. 90-105, 2002

[12] T. Kikukawa, S. Kawafuchi, "Development of an automatic summary editing system for the audiovisual resources", IEEE Trans. on Electronics and Information, J75-A, pp. 204-212, 1992.

[13] LeGall D., Mitchell J.L., Pennbaker W. B., Fogg C.E., "MPEG video compression standard", Chapman & Hall, New York, USA, 1996

14] Y. Zhuang, Y. Rui, T. S. Huan, and S. Mehrotra, "Adaptive key frame extracting using unsupervised clustering," in Proc. Int. Conf. Image Processing, Chicago, IL, 1998, pp. 866–870.

[15] Zuzana Cernekova, Ioannis Pitas "Information Theory-Based Shot Cut/Fade Detection and Video Summarization" in IEEE proc. in circuits and systems for video technology, VOL. 16, NO. 1, JANUARY 2006.

[16] Ali Amiri and Mahmood Fathy "Hierarchical Keyframe-based Video Summarization Using QR-Decomposition and Modified k-Means Clustering" in Hindawi Publishing Corporation EURASIP Journal on Advances in Signal Processing, Volume 2010.

[17] A. Hanjalic, "Shot Boundary Detection: Unraveled and Resolved?," IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, no.2, pp. 90-105, February 2002.

[18] D. DeMenthon, V. Kobla, and D. Doermann, "Video summarization by curve simplification," in Proc. 6th Int. ACM Multimedia Conf., Bristol, U.K., 1998, pp. 211–218.

[19] N. Doulamis, A. Doulamis, Y. Avrithis, and S. Kollias, "Video content representation using optimal extraction of frames and scenes," in Proc. IEEE Int. Conf. Image Processing, Chicago, IL, 1998, pp. 875–878.

[20] J. Calic and E. Izquierdo, "Towards Real-Time Shot Detection in the MPEG Compressed Domain", Proceedings of the Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS'2001, Tampere, Finland, May 2001.

[21] Jianhua Lu, and Ming L. Liou, "A Simple and Efficent Search Algorithm for Block-Matching Motion Estimation", IEEE Trans. Circuits And Systems For Video Technology, vol 7, no. 2, pp. 429-433, April 1997.

[22] Tianming L, Zhang HJ, Qi FH (2003). "A novel video key-frame extraction algorithm based on perceived motion energy model"

[23] Raman Maini & Dr. Himanshu Aggarwal, "Study and Comparison of Various Image Edge Detection Techniques" ,International Journal of Image Processing (IJIP), Volume (3) : Issue (1) .