Spectral Clustering Based Technique for Key Frame Extraction for Video Summarization

[1]Johnwesley, [2]Darakhshinda Parween
[1,2]Assistant Professor
[1,2]Department of Computer Science and Engineering
[1,2]Sree Dattha Institute of Engineering and Science, Sheriguda, Hyderabad-501510, Telangana, India

Abstract— The popularity of digital video is increasing rapidly. To help users navigate in huge data base, the algorithms that automatically index video based on key frames are needed. One approach is to extract key frames based on clustering approach.
Key frame is a frame which can represent the salient content of the shot. Key frames provide a suitable abstraction and framework for video indexing, browsing and retrieval. They allow users to quickly browse over the video by viewing only a few high-lighted frames. The use of key-frames greatly reduces the amount of data required in video indexing and provides an organizational framework for dealing with video content. Because of its importance, much research effort has been given in key frame extraction.
The presented key frame extraction scheme relies on an improved spectral clustering algorithm for the clustering of frame sequences of video. Video frame are clustered into groups by their similarity differences. Then the clustering process is carried for shots of the video using spectral clustering method. Then the key frame identification scheme is established for key frame extraction process and key frame of the video shot are extracted. Finally proposed model presents a key frame extraction method based on clustering scheme using the spectral clustering technique and the key frames of the shot are extracted.

Keywords— **Key Frames Extraction, Video Summarization, Spectral Clustering**

I. INTRODUCTION

There has been a tremendous growth in the multimedia information availability on the Web and other archives in the recent years. The latest developments in multimedia technology, combined with a considerable growth in computer performance and the expansion of the Internet have provided people with access to a tremendous amount of video information. There also been a rise in the number of low bandwidth technologies such as wireless and mobile that is typically resource poor. Together these developments indicate the need for technologies that examine vast amounts of multimedia information to facilitate full content selection based on previews. A browsing facility that provides an information oriented summary for selection of actual content is a necessity. This has lead to an increasing demand of efficient techniques to summarize, index, retrieve and store the video content.

The volume of digital video data has been increasing significantly in recent years due to the wide use of multimedia applications in the areas of education, entertainment, business, and medicine. With developing of internet, more and more video information need to be and managed at low cost. Digital video contains so much information that even simple operations such as browsing, searching and retrieval are often very expensive in both time and computation complexity. The key frame is used to represent the key picture frame shot and to reflect main content of the shot. With granularity from small to large, the segmentation results can be frame, shot, scene, and video. Figure 1 below illustrates the relationship among them.

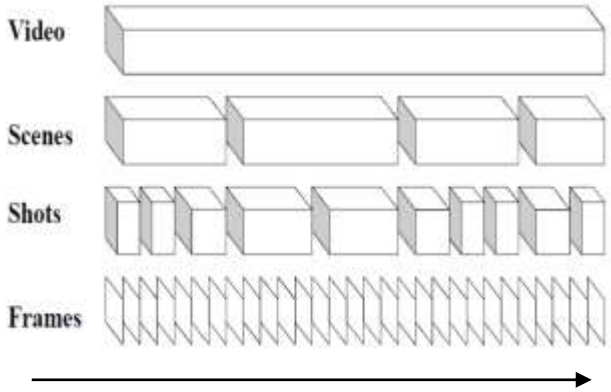


Fig 1. Structure of video

II. RELATED WORK

Key frames extraction is one of the active issue in visual information retrieval research. A review of the major approaches that have been proposed by different researchers in the field to tackle this problem is given below.

An improved spectral clustering method is proposed in [1], to cluster the shots into groups that both estimates the number of clusters , and employs the fast global k-means algorithm in the clustering stage after the eigenvector computation of the similarity matrix.

A method to perform a high-level segmentation of videos into scenes is proposed in [2]. A scene can be defined as a sub division of a play in which either the setting is fixed, or when it presents continuous action in one place. This fact is exploited and a novel approach for clustering shots into scenes by transforming this task into a graph partitioning problem is presented. A new algorithm for identifying key frames in shots from video programs is presented in [3]. The optical flow computations are used to identify local minima of motion in a shot-stillness emphasizes the image for the viewer. This technique allows identifying both gestures which are emphasized by momentary pauses and camera motion. The key frame extraction presented in [4] is from consideration of set-theoretic point of view, and systematic algorithms are derived to find a compact set of key frames that can represent a video segment for a given degree of fidelity. The proposed extraction algorithms can be hierarchically applied to obtain a tree-structured key frame hierarchy that is a multilevel abstract of the video. The key frame hierarchy enables an efficient content-based retrieval by using the depth-first search scheme with pruning. Then, systematic extraction algorithms based on the point set theory are presented.

With the features of MPEG compressed video stream, a new method for extracting key frames is is presented in [5]. Firstly, an improved histogram matching method is used for video segmentation. Secondly, the key frames are extracted utilizing the features of I-frame, P-frame and B-frame for each sub-lens. Fidelity and compression ratio are used to measure the validity of the method.

A simple clustering algorithm presented in [6], has given tool for clustering process in proposed methodology. A novel criterion called shot reconstruction degree (SRD) is defined in [7] which is the degree of retaining motion dynamics of a video shot. Compared with the widely used fidelity criterion, the key frame set produced by SRD can better capture the detailed dynamics of the shot. Using the new SRD criterion, a novel inflexion based key frame selection algorithm is developed.

A triangle model of perceived motion energy (PME) to model motion patterns in video and a scheme to extract key frames based on this model is proposed in [8]. The key frame is a simple yet effective form of summarizing a long video sequence. The number of key frames used to abstract a shot should be compliant to visual content complexity within the shot and the placement of key frames should represent most salient visual content. Motion is the more salient feature in presenting actions or events in video and, thus, should be the feature to determine key frames. The key-frame selection process is threshold free and fast and the extracted key frames are representative.

An efficient video content representation using automatic key-frames extraction is presented in [9]. The proposed video-content representation provides the capability of the more efficient browsing digital video sequences. Firstly, each video sequence is partitioned into shots by applying a shot-cut detection algorithm. Here a multidimensional shot-level feature vector by fusing audio and visual information to describe the average frame properties of the shot is presented.

A technique to automatically extract a single key frame from a video sequence is presented in [10]. The technique is designed for a system to search video on the World Wide Web. For each video returned by a query, a thumbnail image that illustrates its content is displayed to summarize the results. The proposed technique is composed of three steps: shot boundaries detection, shot selection, and key frame extraction within the selected shot. The shot and key frame are selected based on measures of motion and spatial activity.

The key-frame extraction technique in content-based video retrieval is proposed in [11]. Dealing with problems existed in the traditional clustering algorithms, an improved shots key frame extraction algorithm based on fuzzy C-means clustering is presented. Using the color feature information in the video frames, and then through the improvement of the clustering algorithm of video sequences to acquire the center value of various classes and the membership degree of every frame relative to the classes, finally the shots will be clustered into several sub-shots.

An innovative approach to the selection of representative frames of a video shot for video summarization is proposed in [12].By analyzing the differences between two consecutive frames of a video sequence the algorithm describes the complexity of sequences in terms of visual content changes. Three descriptors color histogram, wavelet statics, and an edge direction histogram are used here to express the visual content .similarity measure are computed for each descriptor.

III. PROPOSED WORK

A. The proposed Algorithm

The algorithm used for key frame extraction is by using a spectral clustering. A methodology for key frame extraction based on clustering is presented in following steps.

To perform key-frame extraction the video frames of a shot are clustered into groups using an improved spectral clustering algorithm. The main steps of the typical spectral clustering algorithm are described as below. Suppose there is a set of objects S = s1, s2, . . . , sN to be partitioned into K groups.

Step 1. Compute similarity matrix A ∈ RN×N for the pairs of objects of the data set S.

Step 2. Define D to be the diagonal matrix whose (i, i) element is the sum of the A's i-th row and construct the Laplacian matrix

$L = I - D^{-1/2}AD^{-1/2}$. (1)

Step 3. Compute the K principal eigenvectors x1, x2, . . . , xK of matrix L to build an N× K matrix X = [x1 x2 . . . xK].

Step 4. Renormalize each row of X to have unit length and form matrix Y so that:

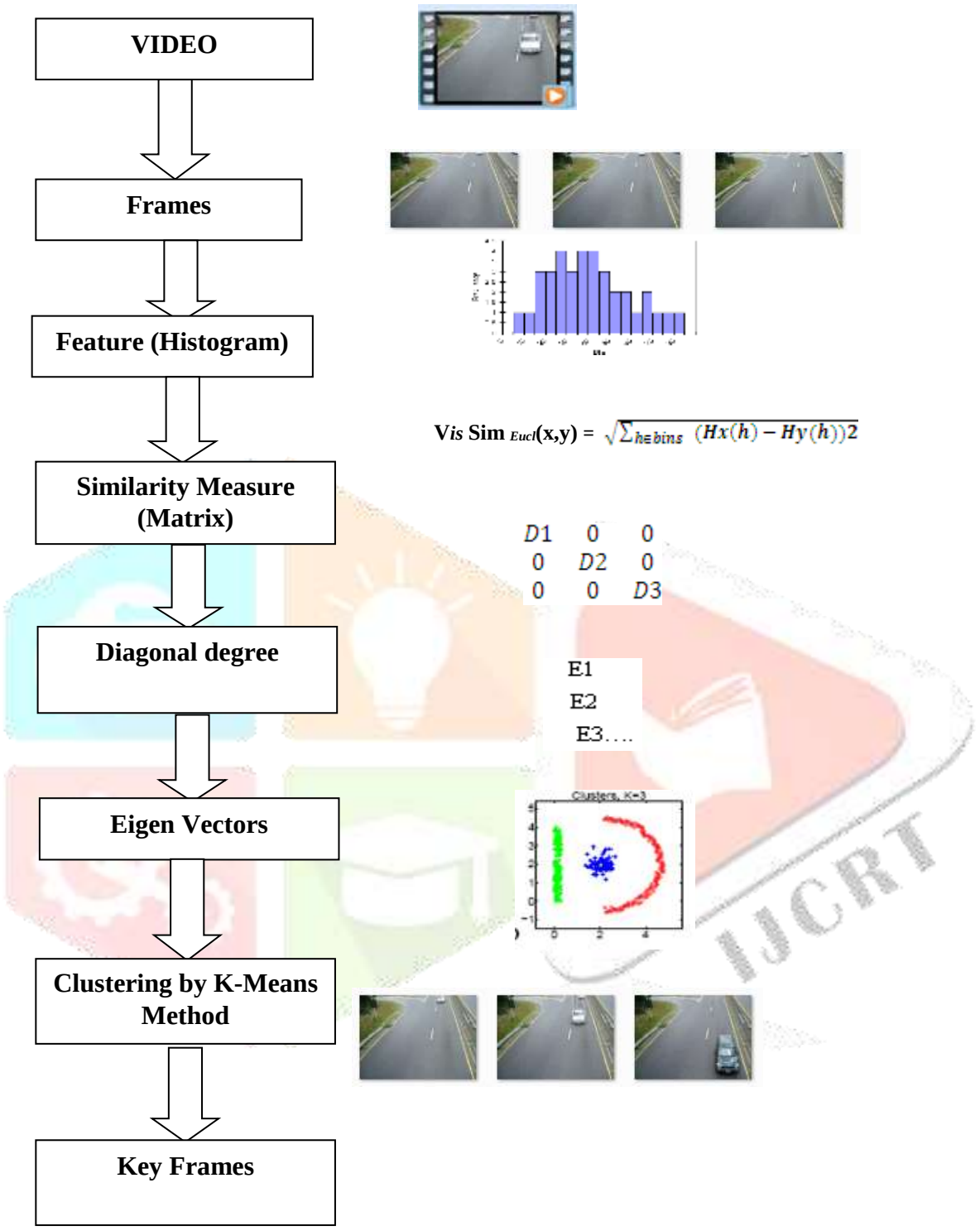
$Y_{ij} = x_{ij} / (\sum x_{2ij})^{1/2}$ (2)

Step 5. Cluster the rows of Y into K groups using k-means.

Step 6. Finally, assign object Si to cluster j if and only if row i of the matrix Y has been assigned to cluster j.

Spectral Clustering algorithms are based on the spectral graph theory. They treat the data clustering as a graph partitioning problem without make any assumption on the form of the data clusters, namely, the clustering of data sets mapped to the Laplacian matrix's row vector which composed of the first 'k' feature vectors. Spectral Clustering algorithm is a point-to-point cluster algorithm, and has a good application prospects. In recent years, Spectral Clustering algorithm is more studied and more increasingly widespread as a cluster analysis algorithm. It is a new branch in the cluster analysis, it was originally used for load balancing and parallel computing, VLSI design and other areas, and it is beginning to be used in machine learning recently, and quickly becomes an international hot spot in the field of machine learning. Currently, Spectral Clustering is attracted more attention in the field of text mining, information retrieval and image segmentation, and has achieved research results. In the first step a video is read from a given dataset. Next the video frames are extracted from the video. Then a feature, histogram is selected, and the feature values are obtained for the frames and a feature vector is retained. With the feature vector a similarity measure is performed distance measure of histograms. Similarity measure is the measure of distance between histograms of frames. Here for similarity measure an Euclidean distance relation is used as shown in the flow diagram. After the similarity measure a diagonal degree is calculated in second step. With the diagonal degree a laplacian matrix is calculated using the relation (1) as shown in above step. Compute the principal Eigen vectors of laplacian matrix as obtained above step. Renormalize the Eigen vectors and form matrix 'Y' using the relation (2) as shown in the proposed algorithm. Next in the fourth step clustering is done for the Eigen vectors by employing the k-means clustering algorithm. The k-means is one of the methods of clustering which clusters and extract the key frames based on the centroid of the cluster.

The flow chart of the proposed algorithm is as shown in below.



Below terms briefly describes the terminologies used in proposed algorithm for key frame extraction.

B. Image histogram

An image histogram is a type of histogram that acts as a graphical representation of the tonal distribution in a digital image. It plots the number of pixels for each tonal value. By looking at the histogram for a specific image a viewer will be able to judge the entire tonal distribution at a glance. Image histograms are present on many modern digital cameras. Photographers can use them as an aid to show the distribution of tones captured, and whether image detail has been lost to blown-out highlights or blacked-out shadows.

Fig.2. Sunflower Image

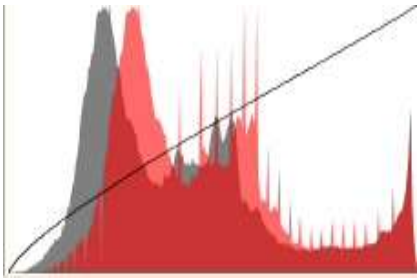


Fig 3. Histogram of sunflower Image

The horizontal axis of the graph in fig.3 represents the tonal variations, while the vertical axis represents the number of pixels in that particular tone. The left side of the horizontal axis represents the black and dark areas, the middle represents medium grey and the right hand side represents light and pure white areas. The vertical axis represents the size of the area that is captured in each one of these zones. Thus, the histogram for a very dark image will have the majority of its data points on the left side and center of the graph. Conversely, the histogram for a very bright image with few dark areas and/or shadows will have most of its data points on the right side and center of the graph. In our key-frame extraction problem, suppose we are given a data set H = H1, . . . HN where Hn is the feature vector (normalized color histogram) of the n-th frame. Similarity measure is the measure of distance between two feature values for their similarity.

C. Frame Similarity Measures

One of the most commonly used techniques to determine similarity between shots is to perform a pair-wise comparison between color histograms of key frames extracted from the shots. This can be viewed as a frame similarity measure. In most methods, the first step of this process is to perform a color reduction, also referred to as color quantization, on the original frames. This step can greatly reduce the computational time for the steps that follow it. Next, color histograms are obtained for each of the key frames and a visual similarity between key frames can be calculated using several methods. The two most commonly used methods to obtain visual similarity are Euclidean distance V isSimEucl(x; y) and histogram intersection V is Simint(x; y).

$$\text{Vis Sim Eucl}(x,y) = \sqrt{\sum_{h \in bins} (Hx(h) - Hy(h))2}$$

Using the above relation the similarity measure is carried out for a pair of objects. In the proposed work, the same distance relation is used and the similarity matrix is computed for the histogram feature value of frames. For example the similarity matrix for a set of feature values is as shown below in fig 4

D. Selecting Key Frames

In most cases, a key cluster consists of only one consecutive sequence of frames which are long enough and the frame that is closest to the centroid of the key cluster in the feature space is chosen as the key frame.

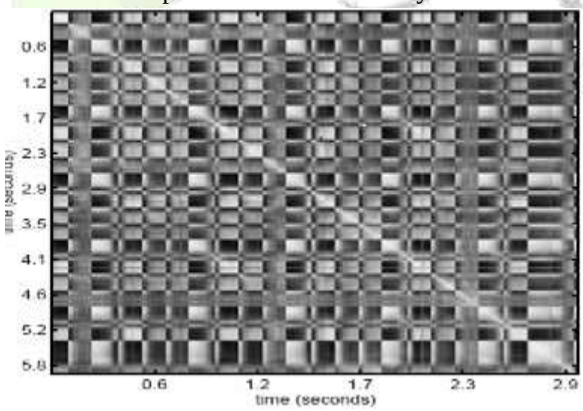


Fig 4. Affinity matrix for a set of feature values

E. The K-means Clustering Algorithm

This part briefly describes the standard k-means algorithm. It is a partitioning clustering algorithm, this method is to classify the given date objects into k different clusters through the iterative, converging to a local minimum. So the results of generated clusters are compact and independent. The algorithm consists of two separate phases. The first phase selects 'k' centers randomly, where the value 'k' is fixed in advance. The next phase is to take each data object to the nearest center. Euclidean distance is generally considered to determine the distance between each data object and the cluster centers. When all the data objects are included in some

clusters, the first step is completed and an early grouping is done. Recalculating the average of the early formed clusters. This iterative process continues repeatedly until the criterion function becomes the minimum. The process of k-means algorithm as follow:

Input the number of desired clusters, k, and a database D={d1, d2,…dn} containing n data objects. Randomly select 'k' data objects from dataset D as initial cluster centers. Repeat until all the objects are selected. Calculate the distance between each data object and all 'k' cluster centers cj(1<=j<=k) and assign data object di to the nearest cluster. For each cluster j (1<=j<=k), recalculate the cluster center until no changing in the center of clusters. The k-means clustering algorithm always converges to local minimum. Before the k-means algorithm converges, calculations of distance and cluster centers are done while loops are executed a number of times, where the positive integer 't' is known as the number of k-means iterations.

IV. EXPERIMENTAL RESULT AND DISCUSSIONS

The theoretical aspects of frame work for key frame extraction are discussed in previous chapter. This chapter is over the implementation of the same using Mat lab. Mat lab tool is used for development of modules since it is found to be one of the most efficient tool for Image and Video Processing.

The proposed algorithm is implemented in MATLAB 7.11.0(R2010b) version on windows 7 operating system installed on a machine with 4 Gb of primary memory and Intel I5 second generation processor. The videos used for the experimentation are of .avi and .mpg format. The experimental dataset used in this project are from the you tube action dataset.

A. Datasets

We have used different kind of videos for testing and analysis of the algorithm. For example: Rhinos Video, Traffic movement, accident …etc videos.

The descriptions about sample video clips used for the testing are given below.

Fig 5. Ice hockey video

This Video is taken from the Youtube data set. The video describes the ice hockey sports video.

For the input video ice hockey.avi file there are totally 1675 frames extracted. The size of the input video file is 2.02Mb. Corresponding to the input file 4 key frames were extracted.

Below steps describes the steps taken for extraction of key frames from the video.

Read and display the video. Firstly a video is taken from the data set and it is read using Mat lab. It reads in all the video frames from the file associated with the object. Select the feature histogram feature vector and extract features values from frames by first converting RGB image into HSV image.

Compute the similarity measure for the frame histogram feature values which calculates the distance values and obtain a similarity measure (matrix). Here the distance measure of features is calculated using the Euclidean distance relation. Find diagonal degree from the similarity measure obtained.

Calculate Laplacian matrix from the obtained diagonal degree measure values of feature values.

Obtain the principal Eigen vectors and renormalize each row of laplacian matrix.

Cluster the Eigen vector values using the K-means clustering method as decribed below.The version of the K-means algorithm is defined below:

Select k random centroid points on our multi-dimensional space.

Compute each frame against all the cluster centroids.

Each frame is assigned to the cluster that minimizes the error function.

Re-compute cluster centroids.

On every iteration, check to see if the centroids converged. If not, we go to step 2.

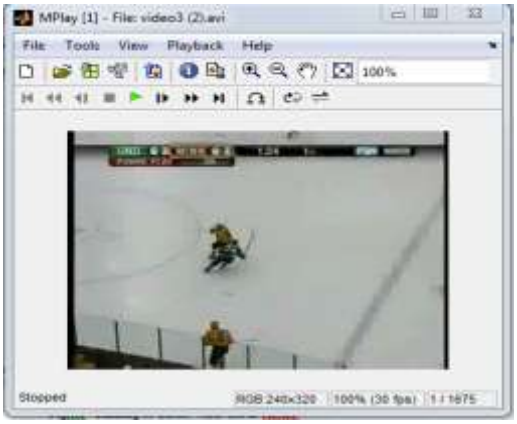


Fig. 6 Reading of ice hockey video file in Matlab

After proper design and implementation of proposed methodology, next is to evaluate results of both methodology and implementation. First the video read step is presented. Here the video read of ice hockey which is shown above.

This video ice hockey is represented in the single frame represents the ice hockey sports video in Mat lab.

B. Key Frame Generation for Ice Hockey video

This process involves generation of frames, converting color images into HSV images and calculating the similarity measure, finding diagonal degree values and calculating Laplacian matrix , principal Eigen vectors and clustering using k-means method of clustering. There are many methods to cluster the frames. Here the global k- means algorithm is used which is efficient and clusters based on users need.

The figure below shows the key frames obtained after the implementation of our proposed methodology.

Fig 7. Key frames of ice hockey Video

C. Evaluation criteria

Based on literature survey, two metrics are selected in order to evaluate the results of a video summary: Compression rate and computation time.

Compression rate is used as a measure of video conciseness. It is computed as:

$$CR = m / N \quad (1)$$

Where 'm' is the number of key frames and N is the total number of frames in the original video. This metric gives an indication of the size of the summary with respect to the size of the original video. Computational time which is the time required to extract the key frames

D. Experimental Analysis

The proposed algorithm is implemented in MATLAB 7.11.0(R2010b) version on *windows 7 operating* system installed on a machine with 4 Gb of primary memory and Intel I5 generation processor. For the evaluation purpose six video clips are downloaded from different datasets. *Table 1* shows the overview of the tests performed. It demonstrates the total frames, number of key frames extracted. In the next columns it gives a comparison between the sample videos with reference to the evaluation metrics namely, computational time which is the time required to extract the key frames; compression ratio which is described as above.

Table 1 Performance results of proposed algorithm for different video genres

Name of Sample Video	Total Frames	KeyFrames Extracted	Computational Time (sec)	Compression Ratio(%)
Rhinos	114	4	5.57	96
Traffic	120	3	4.57	97
Accident	587	6	27.70	99
House Tour	556	5	37.47	99
Sports1	701	8	24.50	98
Comedy	539	5	30.14	99

The same results can be demonstrated in a better way by the charts. Fig.8 shows the plot of computational time or time taken by the algorithm to extract the key frames per sample video. Similarly fig.9 describes the compression ratio (CR) per sample video.

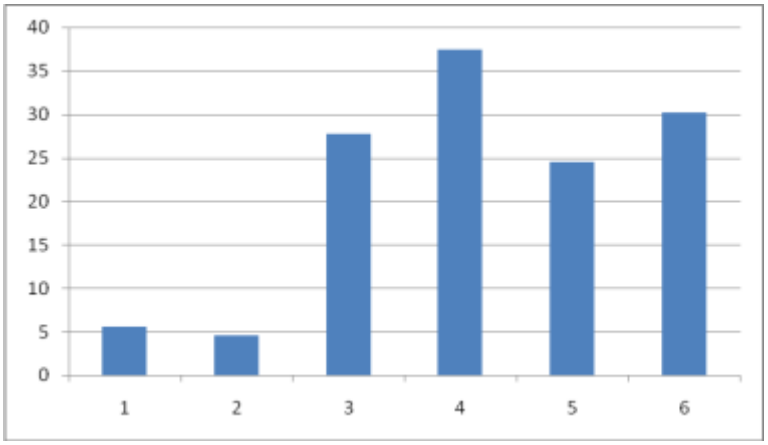


Figure 8 Computation Time per sample Video

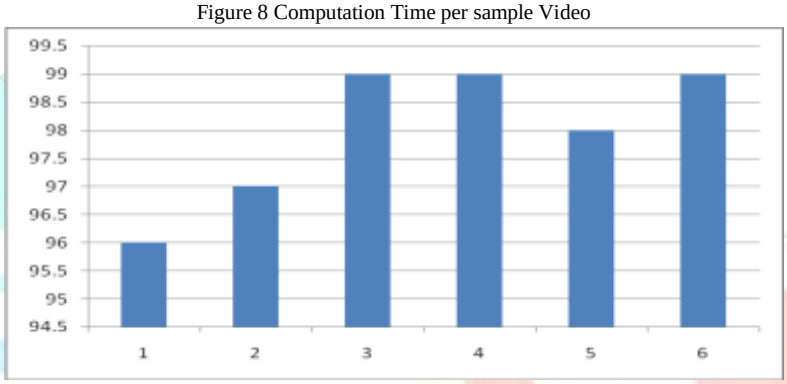


Fig.9 Compression Ratio per sample video

V. CONCLUSIONS

Key frame is a frame which can represent the salient content of the shot. Key frames provide a suitable abstraction and framework for video indexing, browsing and retrieval. They allow users to quickly browse over the video by viewing only a few high-lighted frames.

In this work a new method for key-frame extraction using spectral clustering is proposed. Key-frames are extracted using a spectral clustering method employing the global k-means algorithm in the clustering procedure. Key frames are extracted using a shot clustering method by an algorithm in clustering procedure. The key frames are undergone for their similarity measure and clustered for the shots using k-means method. Furthermore, the number of key frames is estimated by examining the Eigen values of similarity matrix corresponding to pair of shot frames.

Many intensive researches has been done on key frame extraction and also large number of articles have been published on this subject during last few decades. The video shot summarization has been subject for many researchers. There are many new methods invented for shot summarization. The above method includes some of the technique to find the key frame, cluster the shots.

REFERENCES

[1] Vasileios T. Chasanis, Aristidis C. Likas, and Nikolaos P. Galatsanos , “Scene Detection in Videos Using Shot Clustering and Sequence Alignment”,2009.

[2] Zeeshan Rasheed and Mubarak Shah, “Detection and Representation of Scenes in Videos”,2005.

[3] Wayne Wolf , “Key Frame Selection by motion Analysis”, 1996.

[4] Hyun Sung Chang, Sanghoon Sull, and Sang Uk Lee, “Efficient Video Indexing Scheme for Content-Based Retrieval”, 1999.

[5] Guozhu Liu, and Junming Zhao, “Key Frame Extraction from MPEG Video Stream”,2009

[6] Andrew Y.Ng , Michael I.Jordan , Yair Weiss. , “On Spectral Clustering : Analysis and Algorithm”.

[7] Tieyan Liu , Xudong Zhang , Jian Feng b, Kwok-Tung Lo, “Shot reconstruction degree: a novel criterion for key frame selection”, 2004.

[8] Tianming Liu, Hong-Jiang Zhang, and Feihu Qi, “A Novel Video Key-Frame-Extraction Algorithm Based on Perceived Motion Energ Model”, 2003.

[9] LANG Congyan XU De CHENG Wengang FENG Songhe, “Automatic key-frames Extraction To Represent A Video”,2004.
[10] Frkdkric Dufaux, “Key frame selection to represent a video”,2000.
[11] Rong Pan,Yumin Tian, Zhong Wang, “Key-frame Extraction Based on Clustering”,2010.
[12] Gianluigi Ciocca, Raimondo Schettini, “Dyanamic Key-frame Extraction for Video Summarization”.