Techniques for Key Frame Extraction: Shot segmentation and feature trajectory computation

Ijya Chugh[1], Ridhima Gupta[2,] , Rishi Kumar[3], Prashast Sahay[4]
Computer Science and Engineering Department
Amity University, Uttar Pradesh, Noida, India
[1]ijyac.06@gmail.com, [2]rdhmgupta@gmail.com, [3]rishikumar182000@gmail.com, [4]pras1211@gmail.com

***Abstract*— Video is a structured media and segmentation of it into basic temporal units is normally the beginning in the processing channel of content based retrieval. This paper looks into the various possible techniques for extraction of key frames from the video stream. As there have been many challenges faced by other researches and projects to properly execute or create a system which can perform video frame extraction and face recognition, this paper will try to resolve them in a more efficient manner. The paper focuses on streamlining all the requirements of a user to view only those scenes which he desires i.e. scenes of a particular character and also the scenes of a particular emotion. The techniques considered herein are Shot Boundary Detection and NFL-technique.**

Index Terms- Key frame extraction, Shot Boundary detection, Nearest Feature Line (NFL) - Break point method, Simplified Break Point Method.

I. INTRODUCTION

One the most effective and efficient way for capturing and storing digital media in today's world is video. Searching for videos on the web often involves the use of existing search interfaces delivered by online video portals. What if we could devise a method of video processing and searches that not only makes the work easier but also make it more interactive? The answer is here to stay, Content-based video retrieval techniques. Through this study we propose an innovative approach to support these searches in an effective manner. One key component of these techniques is the key frame extraction which can be analysed with image processing algorithms.

Video is by its nature a temporally structured media and segmenting it into its basic temporal units (shots) is usually the first step in the processing of content-based retrieval systems. Based on the detected shots there exist several key frame extraction approaches which then are used as representatives of the shot.

II. LITERATURE REVIEW

A. Shot Boundary Detection

This technique can be performed based on various methods such as pixel per Inch or better known as "PPI-pixel intensity", histogram based approach, edge and motion vector based, which are first implemented and then analysed. Taking into account each and every methodology and approach that exit, the most popular one is the "Histogram Difference". Method based on histogram for shot boundary detection is constituted of the global percentages of various colours that a picture can possibly contain. This technique based on histogram neglects space distribution and pixel.

Histogram technique is based on separating each frame into smaller blocks and taking a "Histogram difference" of frames in succession, and then specific weights are assigned to the blocks in consideration.

Threshold is thereafter formulated by the calculation of mean and standard deviation. If the difference between the frames is more than the threshold of a frame then that frame is the key frame.

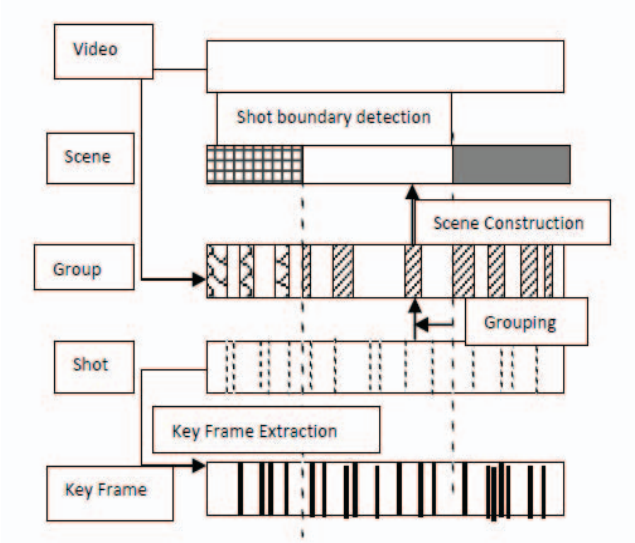

Fig. 1 Extraction Process

B. Algorithm:

Step 1: A video sent as input is read, and then processed by the ShotBoundaryDetection method.
Step 2: Using this method, the method divideIntoSubFrames is called to convert frames into sub frames. These sub frames are a part of respective blocks.

Step 3: The block difference for each sub frame is calculated by the divideIntoSubFrames method.
Step 4: Add up the block differences by the given Formula,

a. Block difference=BD (1)
b. Histogram of 1st block= H1 (2)
c. Histogram of 2nd block=H2 (3)
d. No of Gray Levels=GL (4)

FORMULA: *BD= (H1-H2)*GL.* [6] (5)

Step 5: Once the block difference is obtained, calculation for the mean deviation (MD) and standard deviation (SD) is done. The threshold is calculated in following manner:

Threshold, **T** = MD + (a* SD) (6)

Step 6: When block difference of a frame is greater than the calculated threshold then that frame becomes the key frame.
Step 7: Step1 to Step 6 are repeated till the complete video has been processed and every key frame has been determined and stored. [1]

The following three steps depict the implementation of the Shot boundary detection method:

1. *Segmentation of Image:* Here the frames attained from the video are broken down into x-rows and y-columns. The calculation of difference between two successive frames is calculated. The final calculated difference is considered as the sum of all the differences.
2. *Attention model:* It concerns itself with the visual point of view of an object by carefully and closely observing and listening, which is the ability to concentrate. The position of the pixels holds great significance. Pixels that are on the edges have much more importance than the other pixels. So, pixels of various positions are given different weights.
3. *Difference Matching*: Colour histograms are used for matching difference. By comparing several kinds of histogram methods it was found that x^2 histogram was the best.[7]

Fig. 2 Video clip

C. Key Frame Extraction Algorithm:

The algorithm discussed below focuses only on techniques that are considered to be different and from variable viewpoints, the fundamental dynamics of the video sequence. There exists an algorithm for key frame extraction which is defined as follows:

Step 1: Computation of dissimilarity between all reference and general frames is done using the above algorithm. (Max (i))

Step 2: Search the greatest difference within a Shot.

Step 3: Determine the type of shot depending on the relationship which exists between max (i) and mean deviation (MD): Static or Dynamic Shot. [9]

Step 4: Determine the position of the key frame as follows:

If ShotType=0, choose the frame in the middle as key frame. If there is even number of shit frames, then choose any one of the two middle frames as the key frame.

If ShotTypeC=1, choose the frame with the greatest difference as key frame.The shot boundary detection technique accepts the input video in "avi" or "mp4" format and video size can vary from around 3MB up to 4.5MB. [2]

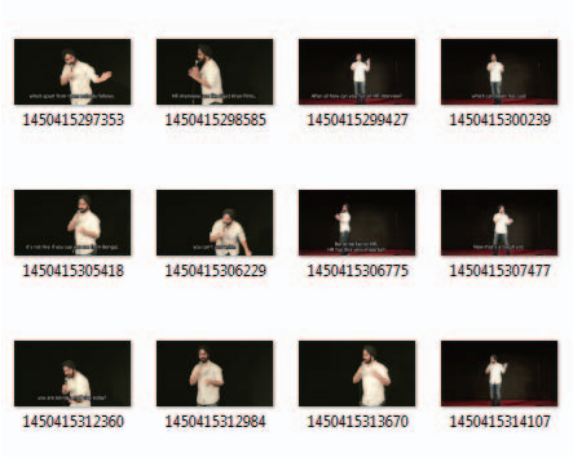

Fig. 3 Extracted key frames for the input clip Key frames

D. Nearest Feature Line Method

The difference in color histogram space between two successive frames is considered to be mainly cued by motion of the subject or by manipulation of the camera.

So, considering the straight line joining the feature points of a key frame with successive key-frame, we can compute the approximate trajectory of continuous frames between the start key frame and end key frame.

Let these variables *f1, f2*..... f_N be taken as the feature points which can be considered equivalent to the key frame sequence of shot ***C***, where ***N*** *stands f*or the number of key frames in a particular shot. The calculated trajectory starting from *f1, f2*.... *t*o f_N results in formation of a curve in the feature space. Thus the feature line $f_k f_{(k+1)}$ is used to estimate the curved segment between the feature points. [5]

The actual manifold which means that the manifold confined by successive feature points which lie close to the FL or are can be as linear as possible, and then we can achieve a

much better performance from the algorithm. A better performance can be observed by breaking the complete curve at Sharp corners and using the comers as break points (BP) key frames.

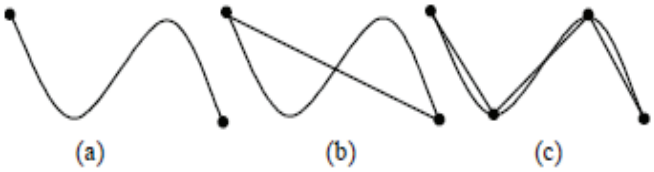

Fig. 4 Various trajectory curves

An efficient way to extract key frames approximately is SBP (Simplified Breakpoint). This technique is described by the algorithm below:

Algorithm:

1. Fix the number of frames that are in shot ***C***, and then initialize all these number of key frames to 0.
2. Continue till all frames are not processed:
   a) If distance between two consecutive frames is greater than δ or distance between alternate frames is greater than C1. δ, then generate a new key frame and increment the number of key frames by 1.
   b) If distance between two consecutive frames is less than δ or distance between alternate frames is less than C1. δ, then no key frames are generated.
3. Exit

User predefines δ as tolerance and C1 is a constant for steps 1~2.

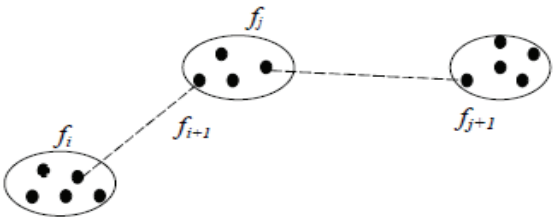

Fig. 5 Calculation of difference between frames

Once the key frames are extracted, the number of key frames obtained N, reflects the character of the shot. If the value of N is minuscule or $N = 0$, it can be concluded that there is very little change in the shot, i.e. the shot can be regarded as a single feature point projected in its own feature space. If the value of N is very huge, it can be concluded that there is a great change in the feature of the shot or that the user defined tolerance δ, selected in step 1-2, is very small. Therefore, δ can be adjusted calculated again. [3]

Hence it can be said that the two techniques in NFL:

1. NFL based "SBP (Simplified Breakpoint) key frame extraction"
2. BP (Breakpoint) key frame extraction with NFL

Fig. 6 Frames extracted

The dissimilarity between the above two sets of images is that in the first set, the method used to extract key frames is classic curve splitting while in the second set, the method for key frame extraction is simplified breakpoint search technique.[4]

III. APPLICATIONS

Presented techniques can be used to simplify numerous manual efforts. Every second over 60 hours of videos are uploaded on social networking sites, which if processed effectively can help extract meaningful information easily. The above explored techniques can be used to extract shots to create summaries of videos and identify subjects. Following are the areas where these techniques can be implemented.

A. *Sports*: Sports matches can run for hours at a time and it not always possible to watch every second of the game. Highlights are a solution to this issue but creating the highlights video is not an easy task to do manually. Some fans are only interested in the performance of their favorite players. For these avid fans, customized videos can be created.

Fig. 7 Sachin from a particular game

B. *Social Gathering Videos:* Weddings and other cultural functions are attended by a large number of people wherein everyone is interested in seeing themselves on the screen. This can be done by extracting those key subjects from the videos for the guests to see. In wedding videos, special moments between the bride and groom can be extracted and compiled for them to remember their special day.

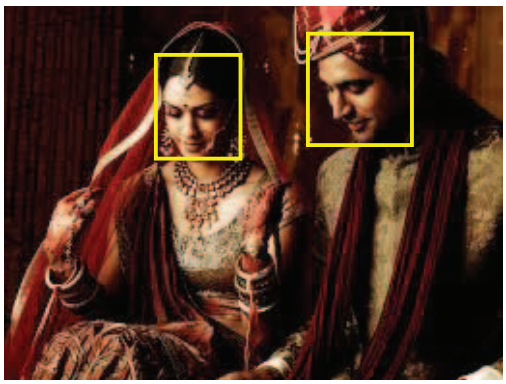

Fig. 8 Detecting bride and groom

IV. CONCLUSION

Shot boundary detection is one of the methods to identify the considerable changes in the content of the video. The extraction of the key frame is done by using a reference frame approach per shot. This method will not just be able to detect all shot boundaries but also store the frames which are suitable and which can designated as key frames to represent the video summary. This algorithm's observed efficiency is ranging from 95% to 98%. One of the significant points to be noted is that there is an exponential increase in the number of key frames whenever special effects such as wipe, dissolve and fading are observed in the video. [8]

The Nearest Feature Line method combines key frame extraction technique based on breakpoints and the NFL classification technique. This way, NFL's best performance can be achieved. Experimental results confirm that the suggested combination of methods perform better as compared to the conventional classification methods like NN and NC and it also supersedes in performance when compared to the traditional NFL method.

REFERENCES

[1] Deepika Bajaj and Shanu Sharma-"Video Depiction of Key Frames- A Review", Sixth International Conference on Computer and Communication Technology 2015 *(ICCCT '15)*. ACM, New York, NY, USA, 183-187, 2015.

[2] Anastasios D. Doulamis, Nikolaos D. Doulamis and Stefanos D. Kollias , "Relevance feedback for content based retrieval in video databases: a neural network approach", National Technical University of Athens, Heroon Polytechniou, 157 73 Zografou, Greece, 1999.

[3] D.Besiris, F. Fotopoulou, N. Laskaris, G. Economou, "Key frame extraction in video sequences: a vantage points approach", Department of Physics, Electronics Laboratory, University of Patras , IEEE, 2007.

[4] H.J.Zhang, D.Zhong and S.W.Smoliar, "An Integrated System for Content-Based Video Retrieval and Browsing," Pattern Recognition, Vol.30, No.4, pp.643-658, 1997.

[5] D.Zhong, S.F.Chang, "Spatio-Temporal Video Search using the Object-Based Video Representation", IEEE International Conference on Image Processing, Vol 1, pp.21-24, 1997.

[6] M.-K. Shan, S.-Y. Lee, "Content-based Video Retrieval based on Similarity of Frame Sequence", Proc. IEEE Conf. on Multimedia Computing and Systems, pp.90-97, 1998

[7] Evaggelos Spyrou and Yannis Avrithis, "Keyframe Extraction using Local Visual Semantics in the form of a Region Thesaurus", National Technical University of Athens Image, Video and Multimedia Laboratory Zographou, 15773 Athens, Greece, IEEE 2007.

[8] Frkdkric Dufaux, "Key frame selection to represent a video", Compaq Computer Corp., Cambridge Research Lab.,", 0-7803-6297- 7/00/ IEEE 2000.

[9] Li Zhao, Wei Qi, S.Z. Li, "Content-based retrieval of video shots using the nearest feature line method", submitted to IEEE WACV 2000.

[10] D.H. Ballard, C.M. Brown, "Computer vision", Prentice-Hall, Inc., Englewood Cliffs, New Jersey 07632,1982.

[11] Ali Amiri, Mahmood Fathy, Atusa Naseri , "Keyframe extraction and video summarization using QRDecomposition", Iran University of Science and Technology, Tehran, Iran, IEEE 2007.