

# Video Retrieval: An Adaptive Novel Feature Based Approach for Movies

Viral B. Thakar  
CHARUSAT

E-mail: viralbthakar@gmail.com

Chintan B. Desai  
CHARUSAT

E-mail: chitundesai10@gmail.com

Prof. S.K. Hadia  
CHARUSAT

E-mail: skhadia.ec@ecchanga.ac.in

**Abstract**— Video Retrieval is a field, where many techniques and methods have been proposed and have claimed to perform reliably on the videos like broadcasting of news & sports events. As a movie contains a large amount of visual information varying in random manner, it requires a highly robust algorithm for automatic shot boundary detection as well as retrieval. In this paper, we described a new adaptive approach for shot boundary detection which is able to detect not only abrupt transitions like hard cuts but also special effects like wipes, fades, and dissolves as well in different movies. To partition a movie video into shots and retrieve many metrics were constructed to measure the similarity among video frames based on all the available video features. However, too many features will reduce the efficiency of the shot boundary detection. Therefore, it is necessary to perform feature reduction for every decision. For this purpose we are following a minimum features based algorithm.

**Index Terms**— Abrupt transitions, Features reduction, Minimum features based algorithm, Video Retrieval

## I. INTRODUCTION

The increasing use of multimedia streams nowadays compels the development of efficient and effective methodologies for manipulation and access of the databases which are having that information. Today, a typical end-user of a multimedia system is usually overwhelmed with video collections, facing the problem of organizing them so that they are easily accessible. As per the analysis given in [1], there were more than 1 trillion numbers of videos on the YouTube during 2011-12 and 201.4 billion number of videos viewed online in the month of October 2011. With such gigantic video data resources, sophisticated video database systems are highly demanded to enable efficient browsing, searching and retrieval. However, the old school video indexing methods, which are totally human dependent, are time-consuming, lack the speed of automation and are diverge by too much human

subjectivity. Therefore, more advanced approaches are needed to support automatic indexing and retrieval directly based on videos content, which provide efficient search with satisfactory responses to the scenes and objects that the user seeks.

Video is just a sequential collection of frames, shots and scenes. In this paper we will explain some of the ways to have efficient retrieval of the videos by using minimum features based methods so that it can be less time consuming [1, 2, 3, and 4].

The effective and efficient video retrieval system totally relies on the effectiveness of the algorithm used for the video segmentation and indexing. A video is nothing but the hierarchical combination of frames, shots and scenes. The shot boundary detection is the process of dividing a video into its core component called SHOT. A shot 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.

A ‘representational frame’ can be chosen from the scene to act as a scene’s representative. Than the comparison between this representative frame and other query frame will be done in different manner. There are so many methods are available like gray scale or color histogram, pixel comparison, edge detection, some of the predefined models, etc.

The Paper flow will be in this manner: section II will give brief idea about literature review & recent trends in the field of video retrieval. Section III will explain the proposed algorithm for the retrieval method. Section IV contains some of the retrieved results by practical. And section V is the conclusion.

## II. BACKGROUND

From the analysis of the different material regarding the automatic video segmentation & retrieval we can say that the basic flow for any video retrieval algorithm will be equivalent to the following block diagram.

From this basic representation we can say that the effectiveness of the video retrieval algorithm is totally dependent on the execution of the approach used for video indexing & segmentation.

There are basically two types of transition in any video sequence: Abrupt & Gradual. The abrupt transition includes CUT transition while gradual transition includes fades, wipe & dissolve kind of special effects.

A cut or abrupt transition defines a drastic rise or instantaneous change in the current frame compare to the past frame. Most of the research work has considered the cut transition as one single task but we would like to define two different types of cuts: Hard Cuts & Soft Cuts. Hard cuts are the cuts which are similar to the conventional type of cuts used for shot boundary detection. Soft cuts are the cuts which are not only used for shot boundary detection but also used to identify that whether consecutive shots are part of same scene or not. So Soft cuts are the cuts which are having more importance & they play a more significant role in the video retrieval [2, 3]. Figure 1, 3, 5 & 7 shows the example of these types of effects.

The other type of transition called gradual transition is also known as Special Effects because they are generated through the different Linear & Nonlinear combination of the consecutive frames of two different shots. Most of the movie videos include fades, dissolves and wipe type of special effects [2, 3, and 4]. Fig. 11 & 14 shows the example of these types of effects.

A Fade can be identified by a gradual transition between a constant image and shot or vice versa. In most of the cases the pure black image is used as a constant image. To identify the shot boundary we have to detect that constant or monotone image.

A Dissolve is the special effect to merge two different scenes or shots without generating eye catchy sequence. The most of the dissolve generating algorithms use linear combination of two scenes where we want to generate dissolve effect. So a dissolve image is generated by the linear combination of pixel intensities of two different images [4].

When we try to analyze the literature for the recent trends for video retrieval & shot boundary detection few commonly used methods shows up like color histogram, edge detection, standard deviation based approach, motion vector difference etc.

In [4] & [5], they have used color histogram and  $\chi^2$  histogram based approach for CUT detection which shows decent results. But the approach for gradual transition is not that much effective for dissolves.

In [6], they have used similarity analyses which again perform efficiently for cut detection but the results for gradual transitions are not that much impressive.

In [8], they have used a single novel feature for the detection of shot boundary. The single feature reduces the total computation time & also reduces the threshold

parameters. We have tried to develop such type of algorithm by using most commonly used features.

In [11] & [12], they have mentioned different features for the video shot boundary detection. We have used these documents to understand the merits & demerits of different features.

In [13] the authors have used invariant descriptor for the frame as the comparison feature. So basically they have analyzed the region of frame which remains almost constant during the whole shot. It is the method for the shot segmentation & indexing part.

In [14] the authors have used the Scale Invariant Transformation as the main feature for the retrieval. The retrieval is very much efficient but the time consumption for the process is very much high because of the calculation of the SIFT for every frame and their comparison.

So by reviewing the work done in [3-5,7,10,11,12], we came up with a conclusion that all the current established methods are pretty much efficient for cut detection but their response towards gradual transition is not that much impressive. Also the other methods which shows higher performance index are found bungling for the movie videos which contains large visual information and their random variations. Apart from this there is a need of the adaptive approach for the retrieval part of the video also.

### III. PROPOSED ALGORITHM

As described earlier a video retrieval task is basically divided into two sub-sections like Video segmentation & indexing and Retrieval.

#### A. VIDEO SEGMENTATION & INDEXING

To examine and understand the aptness of different classes of Video Indexing algorithms with respect to the detection of hard cuts as well as special effects like wipes, fades, and dissolves, at least one representative of each main approach was implemented and tested on a large variety of sequences. Based on these experimental results several shot boundary detection approaches were selected to be combined.

Different approaches have considered different features like histogram, color histogram,  $\chi^2$  histogram to analyze the color based feature of image. The main drawback of this is that it is not effective for the fades & dissolves. DCT has also been considering as an individual feature to consider the change in frequency domain but it also not that much efficient. So we have considered all this individual features to generate a unique feature which contains all the information about image like its texture, color and frequency domain transformation.

#### a) CUT DETECTION

As mentioned in earlier section a CUT is characterized as an abrupt change in the frame sequence. We have considered the following individual features for each frame to identify the Strong Cuts and Weak Cuts.

### 1) DISCRETE COSINE TRANSFORM:

The discrete cosine transform (DCT) represents an image as a sum of sinusoids of varying magnitudes and frequencies. In our system the normalize frame is converted into gray color space, it's equally divided into non-overlapped 8X8 blocks. The DCT operates on an X block of N x N image samples and creates Y, which is an N x N block of coefficients. The action of the DCT can be described in terms of a transform matrix A. The forward DCT is given by  $Y = AXAT$ , where X is a matrix of samples, Y is a matrix of coefficients, and A is an N x N transform matrix. The elements of A are

$$A_{i,j} = C_i \cos \frac{(2j+1)i\pi}{2N} \quad (1)$$

$$\text{where } C_i = \begin{cases} 1/\sqrt{N} & i = 0 \\ 2/\sqrt{N} & i > 0 \end{cases}$$

Therefore the DCT can be written as

$$Y_{xy} = C_x C_y \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} X_{ij} \cos \frac{(2j+1)y\pi}{2N} \cos \frac{(2i+1)x\pi}{2N} \quad (2)$$

### 2) HSV HISTOGRAM:

A histogram is a function that calculates the number of observations that fall into each of the disjoint categories known as bins. Thus, if we let n be the total number of observations and k be the total number of bins, the histogram  $m_i$  meets the following conditions

$$n = \sum_{i=1}^k m_i \quad (3)$$

Using above equation for individual H, S & V components of image we have come to the resultant HSV Histogram.

We have calculated the cumulative HSV histogram also. A cumulative histogram is a mapping that counts the cumulative number of observations in all of the bins up to the specified bin. That is, the cumulative histogram  $M_i$  of a histogram  $m_j$  is defined as:

$$M_i = \sum_{j=1}^i m_j \quad (4)$$

### 3) EDGE CHANGE RATIO:

The Edge Change Ratio is defined as follow. Let  $\sigma_n$  be the number of edge pixels in frame n,  $X_n^{in}$  and variable  $X_n^{out}$  are the number of entering and exiting pixels in the frame then calculate below equation it will give the value of Edge Change Ratio in between 0 to 1.

$$ECR_n = \max(X_n^{in}/\sigma_n, X_n^{out}/\sigma_n - 1) \quad (5)$$

Apart from these main features we have also considered grey level histogram and grey image correlation as two supporting features for our algorithm. This individual features is not that much effective so we have combined them to define our novel difference index which has outperformed all other methods.

The difference index between two consecutive frames is calculated as

$$\text{DiffIndex} = \frac{\text{DCTcorr}}{\text{HSVhistocorr} + \text{Graycorr} + \text{Grayhistocorr} + \text{ECR}} \quad (6)$$

Where DCTcorr, Histogramcorr and Greycorr are the correlations for two consecutive frames. Fig. 2, 4, 6, & 8 shows the few examples of strong cuts & weak cuts detected through above illustrated approach.

### b) FADE DETECTION

To detect the fade in/out effect we have used the Entropy as a feature. As explained in the earlier section fade is generated or ends in a monotone image. Entropy is a statistical measure of randomness of the image. Entropy is defined as

$$E = -\sum(p * \log_2 p) \quad (7)$$

Where p represents the histogram count.

For a monotone image the value of entropy is very near to zero. So by using entropy we can easily detect the shot boundary for the fade in/out very easily and efficiently. Figure 12 & 15 show the Entropy graph for the fade in and fade out effects for the same video sequence.

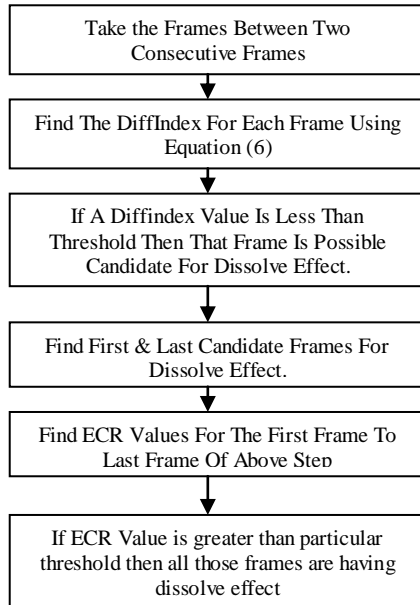
Apart from this we have also calculated the standard deviation of every frame. By combining the results of standard deviation and entropy we can accurately find the monotone frame which act as a key parameter for fade in or fade out detection. Figure 13 & 16 shows the graph of standard deviation for the same frames which we have considered for entropy.

### c) DISSOLVE DETECTION

As described earlier dissolves are generated by linear combination of two different shots. It's observed that for the effectiveness of shot boundary detection algorithm the detection of dissolve shots is crucial. We have observed that dissolve is combination of fade in & fade out effects, in which one shot is fade in while other is fade in at the same time. So we have used the same Entropy of image as one feature to detect the gradual transition of frames.

Apart from this another main quality of dissolve is that during dissolve type of transition the quality of frames is gradually decreasing compare to the last frame until the next shot is totally overruled the previous one. So the frames which are affected by the dissolve effects contain two different images overlapping on each other. To identify this overlapped region we have deployed following algorithm.

The Figure 17 shows the example frames which are having dissolve effect. Figure 18 shows the correlation graph for all the frames between two consecutive cuts for above example. It also shows the start and end point of the dissolve section. Figure 19 shows the ECR graph for those frames, which clearly shows the dissolve effect. The main advantage of this method is it can very effectively find the number of consecutive dissolves also. Figure 20 shows the result for one special case of three dissolves effect in Rajneeti.avi.



Now to extract the key frame or representative frame we have studied few algorithms. In most of them the authors have chosen the first frame of every detected shot as the key frame. The main disadvantage of this method is that it is less accurate to continuous dissolves. So to make it more effective we have selected three frames, i.e. first, last & middle, as the representative.

#### B. ADAPTIVE APPROACH

The adaptive threshold for the above difference index is calculated using the following method.

1. Divide all the frames in fix group size (Clustering Or Windowing).
2. For each group calculate mean & standard deviation of the DiffIndex.
3. Using equation (8) calculate the threshold value for CUT detection.

$$\text{Threshold} = \text{mean} + a * \text{standard deviation} \quad (8)$$

Where 'a' is an integer constant.

4. If for a particular frame difference index is greater than threshold announce it as possible result.

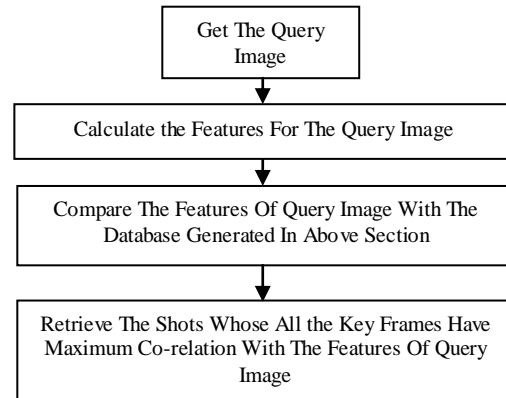


Figure.1 .Strong Cuts – Example - 1

By dividing the total frames into different groups of frames we are reducing the amount of false positives compare to calculating a global threshold for whole video. This kind of clustering based adaptive approach shows very promising results for the video retrieval.

#### C. VIDEO RETRIEVAL

The following flow chart shows the algorithm for the retrieval part. The choice of three key frames makes the retrieval more effective & efficient and this proposed algorithm outperforms the other methods.



#### IV. EXPERIMENT SIMULATION AND RESULT ANALYSIS

The performances of the implemented algorithms are evaluated based on the recall and precision criteria. Recall is defined as the percentage of desired items that are retrieved. Precision is defined as the percentage of retrieved items that are desired items.

$$\text{Recall} = \frac{\text{Correct}}{\text{Correct} + \text{Missed}} \quad (9)$$

$$\text{Precision} = \frac{\text{Correct}}{\text{Correct} + \text{False Positives}} \quad (10)$$

$$F_1(\text{Recall}, \text{Precision}) = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} \quad (11)$$

The proposed algorithm for video shot boundary detection is evaluated on 3 different movie videos which are different in size and nature, one video of a talk show and on an animated video. The values of recall, precision and F1 for every video input is mentioned in the table 1.

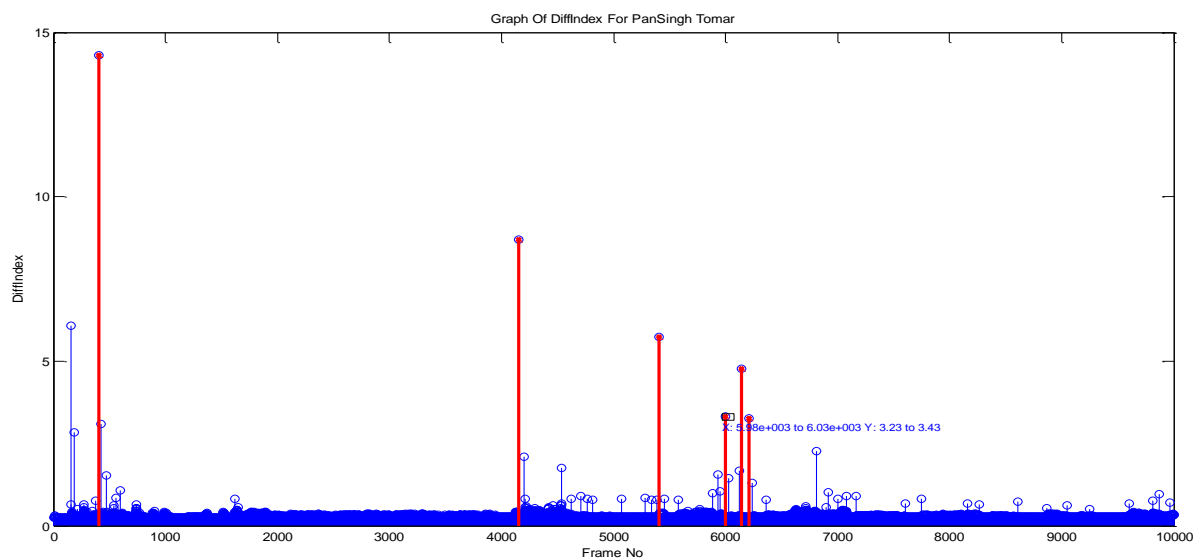


Figure.2. Difference Index for Strong Cuts – Example -1



Figure.3.Strong Cuts – Example – 2

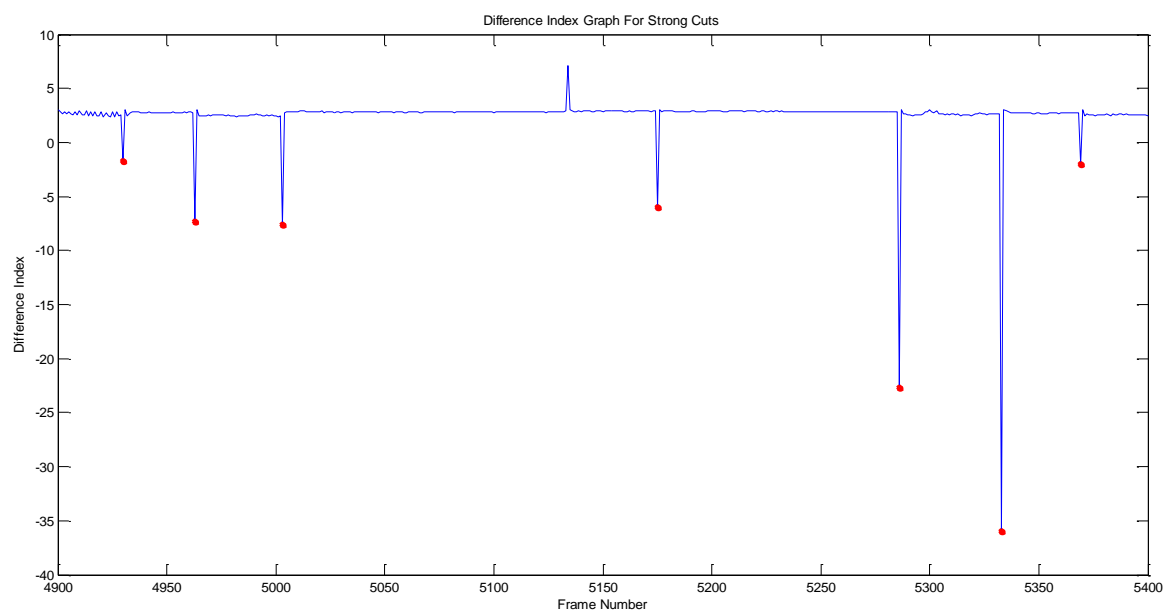


Figure.4. Difference Index for Strong Cuts – Example -2

### A. STRONG CUTS:

Here in Figure 1 & 3 we have shown few examples of Strong Cuts from the database. The graph of Difference Index calculated in equation (6) is shown in Figure 2 & 4. This result shows the effective detection of CUTS for the movie.

The Concept of Weak Cut is newly introduced by us. During our study phase we have found that for video retrieval scene detection is also important. The Weak Cuts are the shot which belongs to the same scene as shown in Figure 5 & 7. The graph shows that we have successfully identified the Weak Cuts & differentiated them from the Strong Cuts.

### B. WEAK CUTS:



Figure.5. Weak Cuts – Example - 1

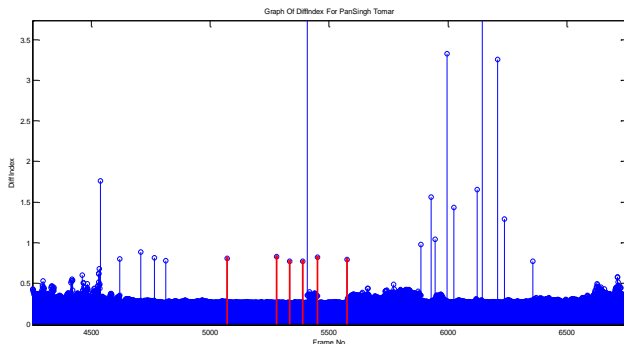


Figure.6. Difference Index for Weak Cuts – Example - 1



Figure.7. Weak Cuts – Example - 2

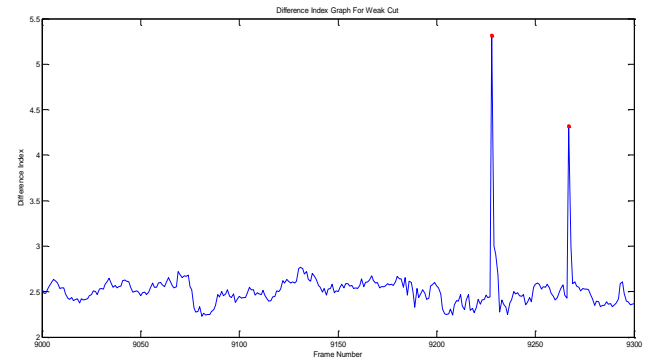


Figure.8. Difference Index for Weak Cuts – Example - 2

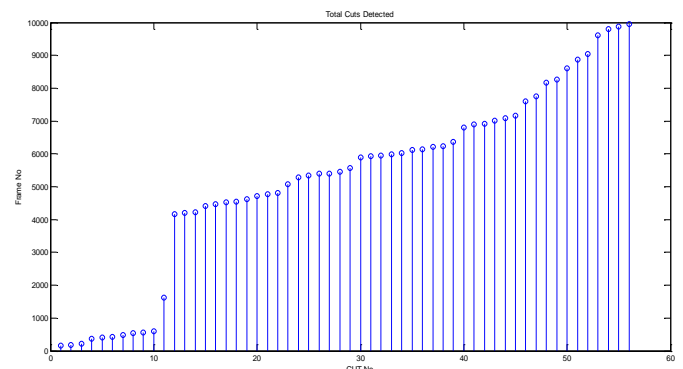


Figure.9. Total Cuts Detected – Example - 1

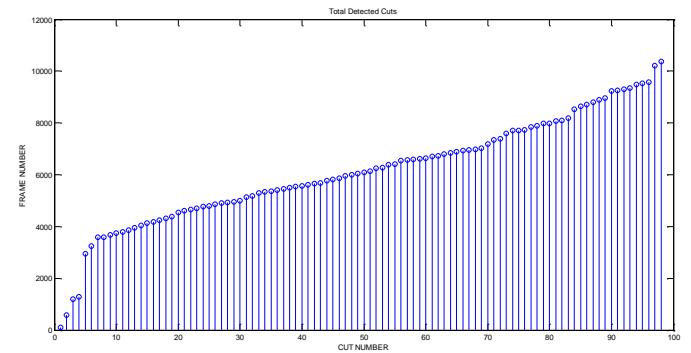


Figure.10. Total Cuts Detected – Example - 2

### C. FADE IN & OUT:



Figure.11. Fade in & Fade out – Example - 1



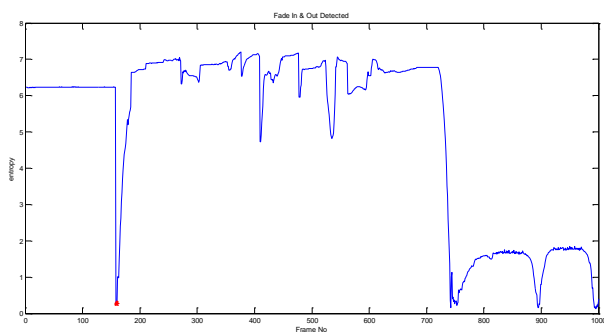


Figure.12. Entropy Graph for Fade in &amp; Out – Example - 1

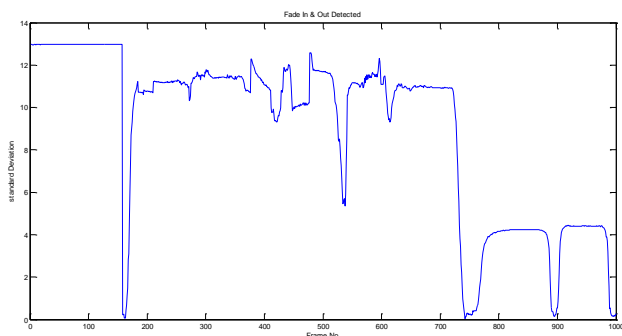


Figure.13. Standard Deviation Graph for Fade In &amp; Out – Example – 1

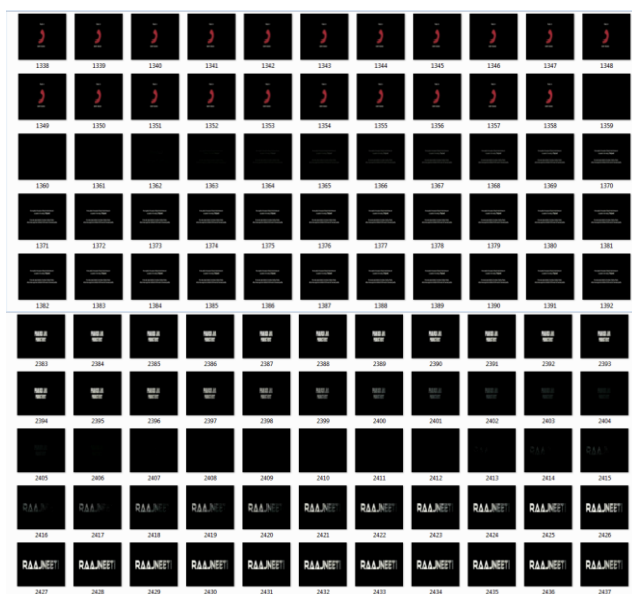


Figure.14. Fade in &amp; Fade out – Example – 2

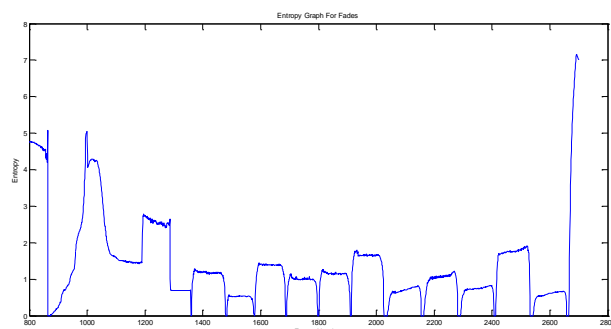


Figure.15. Entropy Graph for Fade in &amp; Out – Example – 2

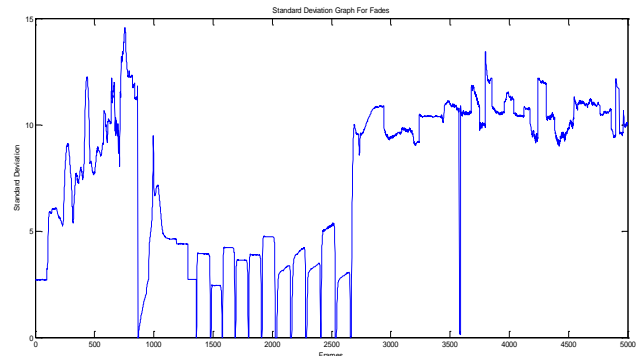


Figure.16. Standard Deviation Graph for Fade In &amp; Out – Example – 2

For Video, Fade In & Fade Out are the most common special effects which represent the transition of frames from a monotone image. Here in Figure 11 & 14 we have shown the examples for the Fade In & Fade Out effects. The graphs of Entropy & Standard Deviation are shown in above figures.

#### D. DISSOLVE:



Figure.17. Dissolve Frames – Example - 1

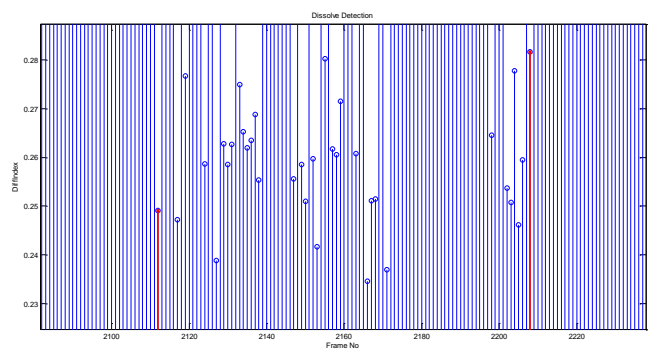


Figure.18. Possible Candidates Dissolves – Example - 1

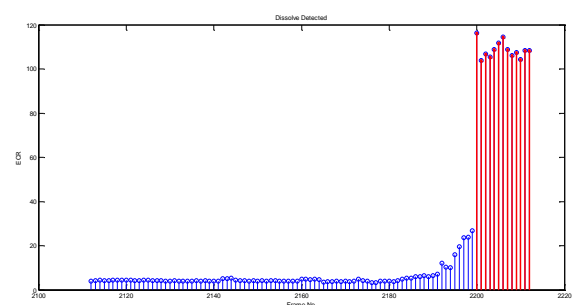


Figure.19. ECR Graph For Dissolve – Example - 1

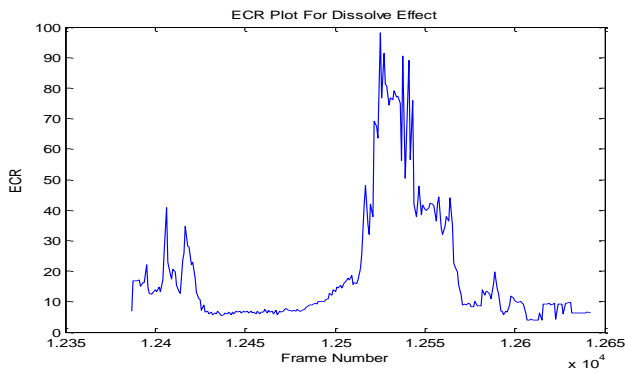


Figure.20. ECR Graph for Three Consecutive Dissolves

As described earlier, the effectiveness of shot boundary detection algorithm depends largely on the efficient Dissolve detection. Here the Figure 17 shows the few frames which are having dissolve transitions. Figure 19 clearly indicates the detection of the frames under the dissolve effect. During our experimental analysis we have come across a special case in which there are three consecutive Dissolve effects. The proposed algorithm has successfully detected this special case also. In Figure 20 we have shown the graph of ECR for this particular case.

Table 1. Result Table

Movie	Frame Number	Effect	Recall	Precision	F1
PanSinghT omar	1 to 15000	Cut	0.98	0.9	0.93
		Fade	1	1	1
		Dissolve	0.95	0.88	0.92
Rajneeti	1 to 20,000	Cut	0.98	1	0.98
		Fade	1	1	1
		Dissolve	0.93	0.9	0.91
Rajneeti	80000 to 100000	Cut	1	1	1
		Fade	N/A	N/A	N/A
		Dissolve	1	1	1
Swadesh	1 to 20000	Cut	0.96	0.92	0.93
		Fade	1	1	1
		Dissolve	N/A	N/A	N/A
Little Manhattan	10000 to 20000	Cut	1	1	1
		Fade	1	1	1
		Dissolve	1	1	1
Talk Shaow	1 to 10000	Cut	1	1	1
		Fade	1	1	1
		Dissolve	1	1	1
Tom & Jerry	1 to 7257	Cut	1	0.94	0.96
		Fade	1	1	1
		Dissolve	0.9	0.92	0.905
Medonna_4 Mins	1 to 7000	Cut	1	1	1
		Fade	1	1	1

		Dissolve	N/A	N/A	N/A
Cricket Match	1 to 20000	Cut	1	1	1
		Fade	1	1	1
		Dissolve	0.8	0.73	0.76

Here in Figure 21 we have shown the final retrieval images. The very first image shows the query image and the rest are the first 20 retrieved key frames. As the images shows the key frames only we can say that we have successfully retrieved the top 20 shots from the movie video based on the query image.

#### E. RETRIEVAL:



Figure.21. First 20 Retrieved Key Frames For Query Image



## V. CONCLUSION

In this paper we have proposed a unique feature based adaptive approach for the video segmentation & indexing. We have used minimum feature based approach & defined a single unique feature for strong and weak cut detection by which we can increase the computation speed and the efficient detection of different types of cuts. The fade in & fade out effects are detected by the entropy and standard deviation of frames in adaptive manner. The proposed method for the dissolve detection shows very promising results. Also we have tried to step ahead towards the query image based video retrieval. Experimental results of different types and size of videos show that the proposed approach gives very decent analysis and detection of video shot boundary as well as video retrieval.

In future work we will try to improve the retrieval algorithm by introducing the adaption parameters in the algorithm.

## ACKNOWLEDGMENT

It is a pleasure to acknowledge the assistance received from several individuals during the preparation of this paper & research work. We are greatly indebted to Mrs. Sheetal D. Tank who is a head librarian at Atmiya Institute of Technology & Science, Rajkot for supporting me and providing us the access of various Journals & Research Articles.

## REFERENCES

- [1] <http://royal.pingdom.com/2012/01/17/internet-2011-in-numbers/>
- [2] Seung-Hoon Han, Kuk-Jin Yoon, and In So Kweon, 2000. "A new technique for shot detection and key frames selection in histogram space." 12th Workshop on Image Processing and Image Understanding, pp 475-479
- [3] Analysis and Verification of Shot Boundary Detection in Video using Block Based  $\chi^2$  Histogram Method by 1Naimish.Thakar, 2Prof.P.I.Panchal, 3Upesh Patel, 4Ketan Chaudhari, 5Santosh.Sangada in International Journal of Advances in Electronics Engineering
- [4] HYBRID APPROACH FOR SHOT BOUNDARY DETECTION FROM UNCOMPRESSED VIDEO STREAM 1Ketan Chaudhari, 2Santosh Sangada, 3Upesh Patel, 4Prof.J.P. Chaudhari, 5Prof.P.I.Panchal in International Journal of Advances in Electronics Engineering
- [5] Shot boundary detection via similarity analysis by Matthew Cooper Jonathan Foote, John Adcock, and SandeepCasi FX Palo Alto Laboratory Palo Alto, CA USA <http://www.fxpal.com>
- [6] A. Hanjalic, Shot-boundary detection: unraveled and resolved?, Circuits and Systems for Video Technology, IEEE Transactions on 12 (2) (2002) 90-105.
- [7] SVM-based shot boundary detection with a novel feature by kazunorimatsumotomasakinaitokddir&d laboratories, inc. 2-1-15 ohara, fujimino-shi, saitama 356-8502, japan {matsu, hoashi, naito, fsugaya}@kddilabs.jpkeiichirohoashifumiakisugaya
- [8] Costas Cotsaces "Video Shot Boundary Detection and Condensed Representation: A Review" Student Member, IEEE, Nikos Nikolaidis, Member, IEEE, and Ioannis Pitas, Senior Member, IEEE.
- [9] Comparison of Automatic Shot Boundary Detection Algorithms by Rainer Lienhart1, Microcomputer Research Labs, Intel Corporation, Santa Clara, CA 95052-8819 Rainer.Lienhart@intel.com
- [10] Advanced and Adaptive Shot Boundary Detection A. Miene, A. Dammeyer, Th. Hermes, and O. Herzog
- [11] Comparison of video shot boundary detection techniques John S. Boreczky Lawrence A. Rowe
- [12] A novel shot boundary detection framework Wujieheng, Jinhui Yuan, Huiyi Wang, Fuzong Lin and Bo Zhang
- [13] A. Anjulan, N. Canagarajah, Invariant region descriptors for robust shot segmentation, in: Proceedings of the IS&T/ SPIE, 18th Annual Symposium on Electronic Imaging, California, USA, January 2006
- [14] D.G. Lowe, Distinctive image features from scale-invariant key points, Internat. J. Comput. Vision 60 (2004) 91-110.

**Viral B. Thakar** has completed his graduation with specialization in Electronics & Communication from the esteemed Saurashtra University, Rajkot.

He is currently pursuing his post graduation from the Charotar University of Science & Technology, Changa in the field of Communication System Engineering. His current research focuses on the field of Video Retrieval & Wireless Sensor Networks.

**Chintan B. Desai** has completed his graduation with specialization in Electronics & Communication from the esteemed Veer Narmad South Gujarat University, Rajkot.

He is currently pursuing his post graduation from the Charotar University of Science & Technology, Changa in the field of Communication System Engineering. His current research focuses on the field of Video Retrieval, Speech Processing & Wireless Sensor Networks.

**S K Hadia** has received the BE degree in E & C engineering from Bhavnagar University and M.E. degree in Communication System Engineering from Gujarat University. He is currently working as Associate Professor in E & C dept., Charotar Univesrity of Science and Technology, Changa and also doing Ph.D. form Charotar University of Science and Technology, Changa. His area of interest is in Wireless Communication Systems, Microelectronics and Optimization