Video Retrieval – An Improvised Algorithm for Movies Using Novel Feature

Krutarth A. Mehta[1], Sandip B. Chotaliya[2]

[1] ME EC Student, Noble group of institution,junagadh,Gujarat,India
[2] Assi. Prof. in Nobel Group Of Institutes, Junagadh, Gujarat,India

---***---

Abstract - *Video Retrieval can be implemented by various processes which contain many techniques & methods. It has been claimed that these methods perform consistently on the videos like television broadcasting of different sports events, talk shows & news telecast. Movies contain a huge amount of informative data which varies in a random manner and so it requires highly robust protocols for automatic shot boundary detection as well as retrieval. The aim of this research is to generate an adaptive & robust algorithm for segmentation & classification of movie videos and get good results of retrieval. In this paper, we have proposed novel feature for shot boundary detection which can detect abrupt transitions like cuts and special effects like fades and dissolves in different movies. This paper present a minimum features based algorithms so; it will give more efficiency of the shot detection and retrieval.*

Key Words: Abrupt transition, Feature reduction, Minimum features based algorithm, Video Retrieval

1. INTRODUCTION

In the era of multimedia, video has become one of the most important mediums in our daily lives. 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 is [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. It has become a necessity for the management of this very huge video data; sophisticated video database systems are highly demanded to enable efficient browsing, indexing and retrieval, which is totally human dependent now a days and so it takes more time and lacks speed of automation and diverged by too much human subjectivity. It demands an approach which can handle automatic indexing and retrieval directly based on videos content, and by which we can provide efficient search with satisfactory responses to the scenes and objects that the user seeks.

Video is a combination of the frames, shots and scenes. In this paper, we will explain some of the ways to have efficient retrieval of the videos using minimum Features based algorithm [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 combination of frames, shots and scenes. The shot boundary detection is the process of dividing 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 shot to act as a shot's representative. Than the comparison between this representative frame and other query frame will be done in different manner for the retrieval. There are so many methods available like gray scale or color histogram, pixel comparison, edge detection, some of the predefined models, etc.

This paper is having following sections. Section-II gives the brief idea about the literature review. The proposed algorithm is explained and discussed in section III. The performance analysis & results are mentioned in section IV. Section V is the conclusion and the future enhancements.

2. LITERATURE REVIEW

From the analysis of the different research material regarding the video segmentation we conclude that there are mainly two types of transitions in any video sequence: Abrupt & Gradual. The abrupt transition includes CUT transition while gradual transition includes fades; wipe dissolve and other special effects.

A cut or abrupt transition defines immediate changes between current frame and previous one. This paper defines two different types of cuts: Hard Cuts & Soft Cuts. Hard cuts are similar to the conventional type of cuts and it is used for shot boundary detection. Soft cuts is used for shot boundary detection and also used to identify that whether consecutive shots are part of same scene or not. So that we can say that soft cuts play very important role & they play a more significant role in the video retrieval [2, 3]. Fig.1, 3 shows the few examples of these types of effects.

The other type of transition is called gradual transition and it is also known as Special Effects because they are by product of the different Linear & Nonlinear combination of the consecutive frames of two different shots. Most of the movie videos contain fades, dissolves and wipe type of special effects [2, 3, and 4]. Fig. 6 and 9 shows the example of these types of effects.

A Fade is combination of two effects: Fade-in and Fade-out. A Fade-in means to gradually start a shot from a monotone screen, while a fade-out is nothing but ending a shot into a monotone image. Refer Fig. 6.

Dissolve is a term used to describe a transition effect in which one video clip (or one image) is gradually fads out while another image simultaneously replaces the original one. It is also called as video dissolve, film dissolve, or linear light blend. The overall objective is for the image or clip to have seamlessly transition from one image to the other in a fluid motion. 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 are 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 proposed 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 this document 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 in [3- 5,7,10,11,12], we conclude that all the methods are effective for cut detection but their response on the gradual transition is not 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.
Video retrieval is basically divided into two subsections which are Video segmentation & indexing and Retrieval.

2.1 CUT Detection

As mentioned in earlier section a CUT is characterized as an abrupt change in the frame sequence. We are considering to following individual features for each frame to identify the Hard Cuts and Soft Cuts.

2.1.1 Discrete Cosine Transform

The discrete cosine transform (DCT) helps separate the image into parts (or spectral sub-bands) of differing importance (with respect to the image's visual quality). The DCT is similar to the discrete Fourier transform: it transforms a signal or image from the spatial domain to the frequency domain. In our research area 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 anN x N transform matrix. The elements of A are

$$Ai,j = Ci \cos\frac{(2j+1)i\pi}{2N} \qquad (1)$$

$$\text{Where } Ci = \begin{cases} {1/N}^{1/2} & i = 0 \\ {2/N}^{1/2} & i > 0 \end{cases}$$

Therefore the DCT can be written as

$$Yxy =$$

$$Cx\, Cy \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} Xij \cos\frac{(2j+1)y\pi}{2N} \cos\frac{(2i+1)x\pi}{2N}$$ (2)

2.1.2 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 mi meets the following conditions

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

Using above equation for individual H, S & V components of image we have come to the resultant HSV Histogram is the combination of all three individual Histograms. 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 Mi of a histogram mj is defined as:

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

2.1.3 Edge Change Ratio

The Edge Change Ratio is defined as follow. Let σn be the number of edge pixels in frame n, Xnin and Xn-1out are the number of entering and exiting pixels in the frame then calculating equation (5) will give the value of Edge Change Ration in between 0 to 1.

$$ECRn = \max\left({Xn^{in}}/{\sigma n}, {Xn-1^{out}}/{\sigma n - 1}\right)$$ (5)

Our algorithm is having main two supporting features is grey level histogram and grey image correlation. These individual features are 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

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

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

2.2 Fade Detection

Using Entropy feature we can detect easily fade in/out effect. 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

$$E = -sum(p * \log_2 p)$$ (8)

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. Fig. 7 show the Entropy graph for the fade in and fade out effects for the same video sequence.

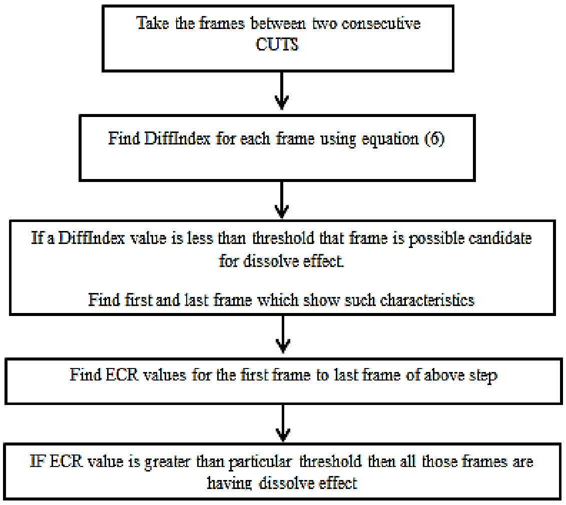
Apart from this we have also calculated the standard deviation of every frame. By combining the results of standard deviation and entropy we can accurately find the monotone frame which act as a key parameter for fade in or fade out detection.

2.3 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 pivotal.

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 Fig. 9 shows the example frames which are having dissolve effect. Fig. 10 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. Fig. 11 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. Fig. 12 shows the result for one special case of three dissolves effect in Rajneeti.avi.

3. ADAPTIVE APPROACH

The adaptive threshold for the above difference index is calculated using the following method.
Divide all the frames in fix group size (Clustering Or Windowing).
For each group calculate mean & standard deviation of the DiffIndex.
Using equation (7) calculate the threshold value for CUT detection.

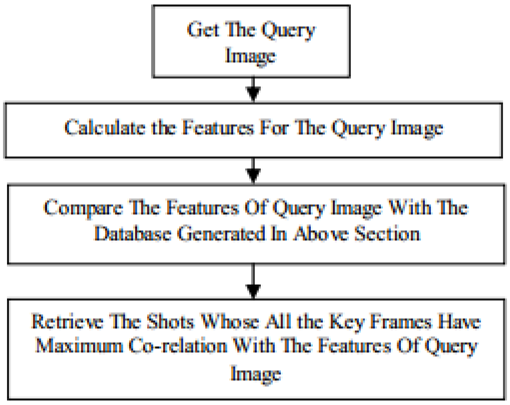
$$Threshold = mean + a * standard\ deviation \tag{7}$$

Where 'a' is an integer constant.
If for a particular frame difference index is greater than threshold announce it as CUT.

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. Using this kind of clustering based adaptive approach we can get very comforting results for the CUT detection.

4. 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.



5. Results

The performances of the implemented algorithms are evaluated based on the recall and precision criteria [8]. 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.

$$Recall = \frac{Correct}{Correct + Missed} \tag{9}$$

$$Precision = \frac{Correct}{Correct + False\ Positives} \tag{10}$$

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

The proposed algorithm for video shot boundary detection is evaluated on 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. If we compare the results to the previous proposed algorithms in [3,4 & 5] the proposed algorithms shows a significant improvement.

5.1 CUT Detection:

5.1.1 Hard Cuts:

Fig -1: Hard Cuts

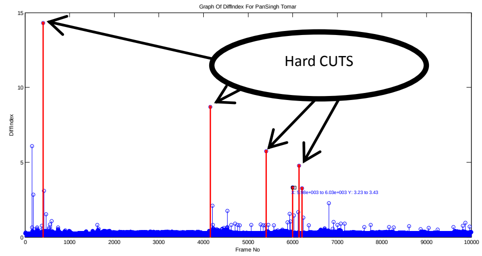


Fig -2: Difference Index for Hard Cuts

5.1.2 Soft Cuts:

Fig -3: Soft Cuts

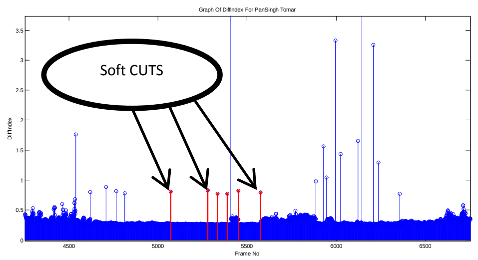


Fig -4: Difference Index for Soft Cuts

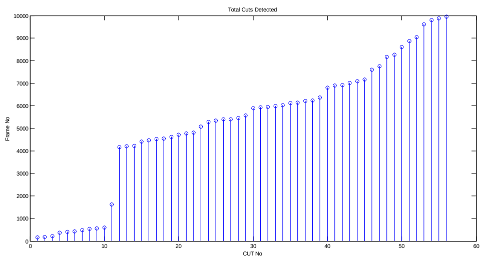


Fig -5: Total Cuts Detected

5.2 Fade In & Out:

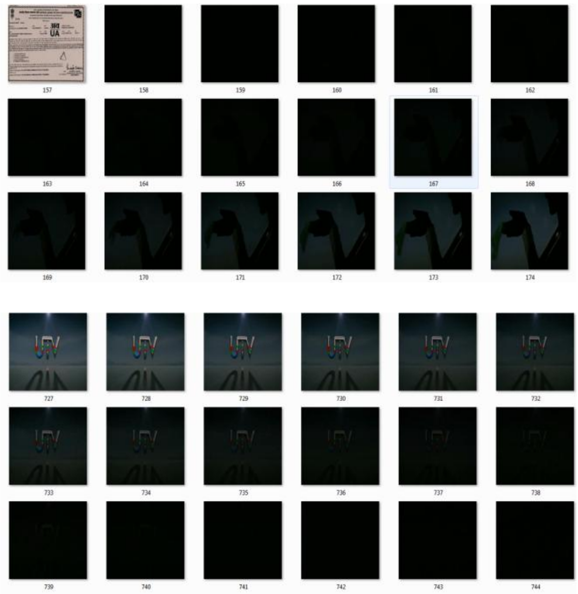


Fig -6: Fade In & Fade Out

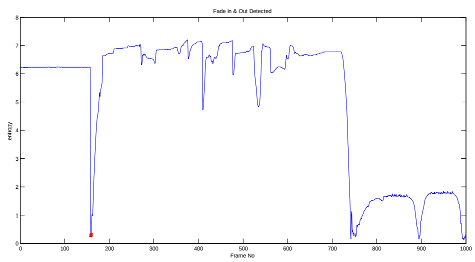


Fig -7: Entropy Graph for Fade In & Out

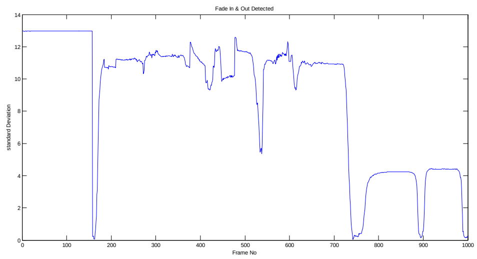


Fig -8: Standard Deviation Graph for Fade In & Out

5.3 Dissolve:

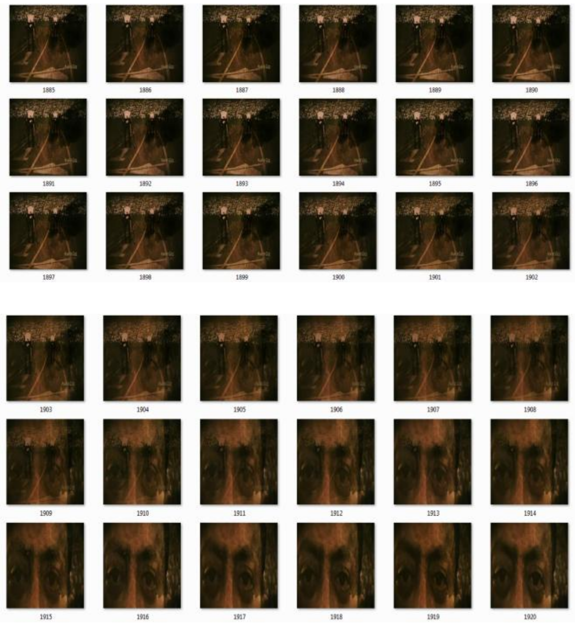


Fig -9: Dissolve Frames

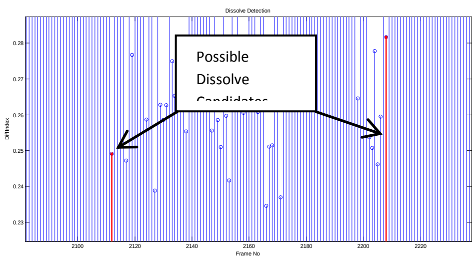


Fig -10: Possible Candidates Dissolves – 1

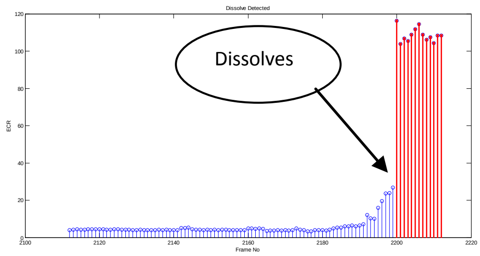


Fig -11: ECR Graph For Dissolve

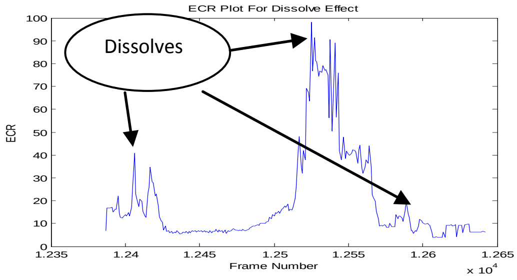


Fig -12: ECR Graph for Three Consecutive Dissolves

Table -1: Result Table

Movie	Frame No.	Effect	Recall	Precision	F1
PanSingh Tomar	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 13 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.

5.4 RETRIEVAL:

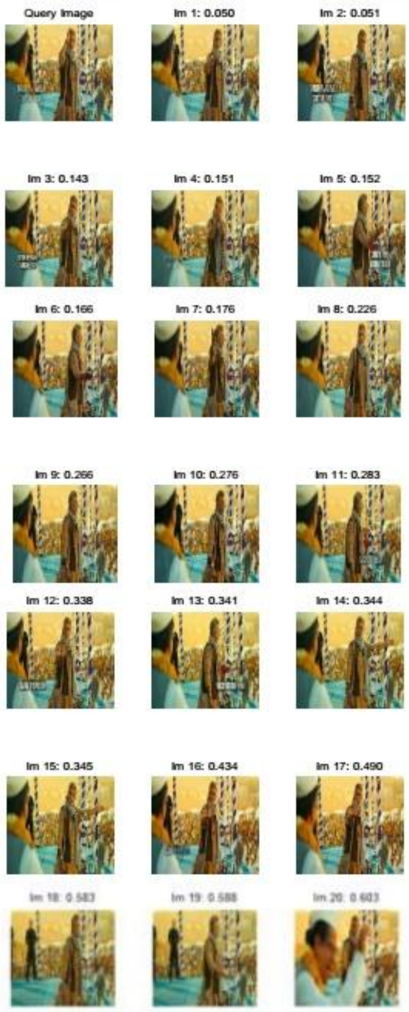


Fig -13 First 20 Retrieved Key Frames For Query Image

6. CONCLUSIONS

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

REFERENCES

[1] http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360_ns827_Networking_Solutions_White_Paper.html
[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 χ2 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,andIoannis 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] 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

BIOGRAPHIES

Mr Krutarth A Mehta has completed his graduation with specialization in Electronics & Communication from the esteemed R.K.College of engineering and technology, Rajkot.
He is currently pursuing his post graduation from the Noble group of institutions, Junagadh.
His current research focuses on the field of Video Retrieval.

Prof. Sandip B. Chotaliya has completed his graduation with specialization in Electronics & Communication from the Esteemed Godavari College Of Engineering, Jalgaon.
He has also completed his post graduation with specialization in Wireless & Communication (2009-2011) from the Gyan Vihar University(Deemed Unversity), Jaipur.
He is a Assistant Professor In Nobel Group Of Institutions , Junagadh

