iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** IX **Month of publication:** September 2017

DOI: http://doi.org/10.22214/ijraset.2017.9091

Shot Boundary Detection Using HSV Colormap for Content Based Video Retrieval

Kalpesh M Shah[1], Ramji M Makawana[2]
[1, 2] *Electronics & Communication Engineering Department- AI eSmart Solutions, CHARUSAT University*

Abstract: In this paper statistical features of color are used to describe the content of video frames. Frames are separated from video. Chrominance content of an image is extracted in HSV color map. Its First order and second order moments of HSV color map are calculated and compare with sequentially with other images using absolute difference. Scene change can be detected by finding boundaries which exist in between series of consecutive frames where two boundary frames are gradually or significantly changes either by its content or background of the frame. Videos from various categories like sports, news and cartoons are used to evaluate proposed method. Higher the absolute difference between consecutive frames point out video frames are changing drastically and so video cut may exist there. Few highest absolute differences boundaries are declared as shot boundary depending on length of a video or number of frames. It has been observed during experimental work, proposed technique offers very good results for all three different categories video. It finds sharp shot change very effectively. Recall and Precision are used as a detection measures for evaluating performance of proposed method.
***Keywords:* Video shot boundaries, HSV color map, Absolute difference, Recall and Precision**

I. INTRODUCTION

Rapid developments of digital video capturing and editing devices and technology have been producing an ample of digital video data every day. As a consequence, digital video data information keep on increasing everyday and so need of an efficient application arrives; such as video-on-demand or content based video retrieval. This technology of how to catalogue and retrieve videos on demand for future re-use becomes a need of society.

In automatic management of video database, it is practically inappropriate to use keywords to describe each video sequence because this annotation process asks for enormous human power; in addition keywords used cannot describe the subject properly. Hence, to fulfil this demand of retrieving similar videos from the database, we need a system Content based video retrieval which retrieves the same videos based on video content analysis.

Primary tasks of CBVR system are shot boundary detection, key-frame extraction. Shot boundary detection becomes the groundwork for video retrieval and management. Shot boundary detection is allowing the computer to discover the editing positions and produce the original shot sequences. The detected shots will become the basic query units in video retrieval systems. Video is a space and time varying structure captured by either single camera or multiple cameras. Video structure can be divided as video scenes, video shots, video frames and video key frames. All play a major role for video indexing and video retrieval applications. A video shot is a succession of frames captured by a single camera in a continuous run. Video shot is a group of sequence of similar action frames which carries almost same information and visual features such as colors, motions and textures.

The visual information of each shot of the video can be described by one or multiple frames, called key-frames. The number of key-frames cannot be predetermined because of content variation. In case of a static shot there is little content variation, so a single key-frame can describe the whole shot effectively, whereas in case of a high camera editing and object motion shot, we may need more key-frames for a better representation of a video content (Vasileios T. Chasanis Jan. 20093).

Video shot boundary detection is a primary step in video indexing and retrieval, and in general video database management. Shot boundary detection is used to segment a given video into its continuous shots in a way it was produced, and to classify and detect the different shot transitions (Don Adjeroh May. 20094). Different algorithms have been proposed, based on color histograms (Ballard 1991, Tanaka 19925), pixel color differences (H. Zhang 1993), color ratio histograms (Lee 1997), edges (R. Zabih 1999), and motion. In this proposed work, we studied the problem of video shot classification and separation using statistical features of HSV color map without calculating statistical threshold. Manual change of threshold is not a practical solution for shot boundary detection. Proposed approach does not rely on threshold value, which generally varies based on content of a video frames.

II. STATE OF THE ART

The transition takes place between two shots can be either abrupt (sharp) or gradual. The sharp or abrupt transition is generated because of a cut effect; in which frame f_i belongs to first shot and next frame f_{i+1} is correspond to the next shot have clear sharp discontinuity in its content. Gradual transition is generated using the application of video editing effects which covers several sequential frames or because of motion spread throughout the range of few frames, so that frame f_i belongs to first shot and frame f_{i+N} belongs to the second shot, and the *N-1* frames in between represent a gradual transformation between two frames. Algorithms detecting shot transitions are based on the calculation of the difference between values of a certain measured characteristics like color, texture, edge, motion etc. of two consecutive frames; this value is later processed via a threshold decision in order to determine the existence of a boundary or not (Jesús Bescós Apr. 2005). Shot boundary mainly has two basic steps.

1) Difference vector is calculated to find the similarity between two frames using features extracted from each image. It includes selection of features, function to evaluate inter-frame disparity like Euclidian distance, Chi-square difference, and Absolute difference.
2) For deciding shot boundary threshold level is calculated using certain statics based on extracted feature vector.
   Effectiveness of different algorithms is measured and compared using recall (R) and precision (P). It can be calculated using (1) and (2) respectively.

$$\Pr ecision = \frac{No.\ of\ relevant\ shots\ retrieved}{Total\ no.\ of\ shots\ retrieved} \qquad (1)$$

$$Recall = \frac{No.\ of\ relevant\ shots\ retrieved}{Total\ no.\ of\ shots\ in\ a\ video} \qquad (2)$$

A. Abrupt Shot Change (Cuts)

[1] Had implemented sharp transition detection method using 3×3 averaging filter to reduce camera motion and noise effects. The easiest way to detect two frames are significantly different in its content is to count the number of pixels that change in value more than some threshold. But, this method was found very sensitive to camera motion. Pixel to pixel matching based block-based comparisons was proposed by [2] using local mage characteristic unlike pixel to pixel comparison mentioned ahead. This approached had increased the robustness to camera and object movement. Each frame of a video is divided into b blocks that are compared with their subsequent blocks in next frame. These blocks comparison is calculated using likelihood ratio.

To reduce sensitivity to camera and object movements (A. Nagasaka 1995) used simplest approach of gray level histograms. An abrupt transition is confirmed if the absolute sum of histogram differences between two succeeding frames fi and fi+1 is greater than a predefined threshold level. One more easy and very effective approach proposed by [1] was to compare color histograms and used absolute difference to find dissimilarity.

[3] had used model based comparison which considered the video production system as a template. [4] used an Edge features for detection shot boundary where it looks for entering and exiting edge pixels.

B. Gradual Shot Changes

Accumulated differences between two frames of the gradual shot transition were calculated by using twin-comparison method. In an initial phase a high level of threshold is used to identify sharp cuts while in the second phase a lower value threshold is used to detect the first frame f_s of a gradual transition. This frame is then compared to subsequent frames and its difference value is accumulated. This accumulated comparison increases during a gradual transition. When the difference between consecutive frames decreases less than T_l , or when the accumulated comparison has increased to a value higher than T_h then end frame f_e of the gradual transition is detected. If the consecutive difference falls below T_l before the accumulated difference exceeds T_h, then the start frame f_s of gradual transition is dropped and the search of gradual transition remains continue.

The thresholds are highly depends on the content of a video. To overcome this problem, proposed an application of unsupervised clustering algorithm by considering video segmentation as a 2-class clustering problem; shot change and no shot change using the well known K-means algorithm explained by is used to cluster the frame dissimilarities. Two similarity measures based on color histograms were used: s2 statistics and the histogram difference.

[3] Proposed a efficient technique for gradual shots detection and classification, using fractal analysis and AIS-based classifier. They computed features of “vertical intercept” and “fractal dimension” of each frame which are calculated using Fourier transform

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue IX, September 2017- Available at www.ijraset.com

coefficients. Classifier based on Clonal selection algorithm was used. They evaluated their algorithm using TRECVID2006 benchmark dataset.

C. Distance Measures

[6] Extracted two types of features: spatiotemporal and compressed domain features. They used the Bhattacharyya measure defined by Eq. (3)

$$d(H1, H2) = \sqrt{1-\frac{1}{N\sqrt{\overline{H}_1\,\overline{H}_2}}\sum_i \sqrt{H1(i)H2(i)}} \quad (3)$$

Where, $\overline{H}_k = \frac{1}{N}\sum_j H_k(j)$ and N denotes the number of histogram

In order to evaluate inter-frame disparity (Jesús Bescós Apr. 2005) have used RGB color bands and applied a criterion consisting of a variant of the Pearson's Test. Color bands contain a similar amount of information for the computation of the objective similarity between two frames, it could be merged into a single distance function Eq. (4)

$$d_1^n[i] = \frac{d_1^n R[i] + d_1^n G[i] + d_1^n B[i]}{3} \quad (4)$$

Where $d_1^n Band[i]$ is the aforesaid variant of the *Pearson's Test* applied to the values that each color band takes for frames f_i and f_{i+n}.

[7] used the cosine measure for comparing the histograms of adjacent frames. The two N-dimensional vectors, A and B, represent the color signatures of the frames. The distance $D_{cos}(A, B)$ between vectors A and B is calculated using Eq. (5):

$$D_{cos}(A, B) = 1 - \frac{\sum_i^N a_i b_i}{\sum_i^N {a_i}^2 \sum_i^N {b_i}^2} \quad (5)$$

Where a_i and b_i are bins in A and B respectively.

III. THRESHOLD DECISION

Most of the algorithms comprise calculation of threshold after a disparity measure is computed, to identify video shot boundaries. At the point, where a frames difference value exceeds the threshold value is labelled as shot boundaries. [8] proposed a simple threshold which was automatically computed on the histogram difference measure presented by [1] . This computation is based on entropy maximization of the distribution of the ID measure. [1]. proposed a twin-comparison technique to identify gradual boundaries. Instead of only one, two thresholds are used. A low threshold is used to identify possible frames at the start of a gradual transition, and a high threshold is used to identify shot boundaries. In [9] four thresholds were used. That is, one distinguishes pixels with high optical flow one determines pixels violating the smoothness constraint, one threshold shot changes, and, finally, the last one is used to remove false alarms.

IV. PROPOSED METHOD

In the proposed method, shot boundary is extracted using statistical features calculated from HSV color map. Color image consists of luminance and chrominance. Here chrominance information is stored in the form of an HSV color map. We motivated to use HSV map because human neural processing system identifies the color as on HSV color map [9].

In this method frames are separated as *.bmp* format with its HSV color map. Color map for image of figure 1 is shown in figure 2. Table 1 is a sample of an HSV map. It is indicating 256 possible color shades with its chrominance level in form of HSV map. Mapping of RGB to HSV map is calculated using standard mathematical equations delineated in Eq. (6), (7), (8). As we can see from the color map each row represents the color of a bin. The row is composed of the three columns of the color space. The first column represents hue, second column represents saturation, and third column indicates value, thereby giving complete HSV map.

$$H = \cos^{-1}\left\{\frac{\frac{1}{2}[(R-G)+(R-B)]}{\sqrt[2]{(R-G)^2+(R-B)(G-B)}}\right\}$$

$$S = 1 - \frac{3}{R+G+B}\,[\min(R, G, B)]$$

$$V = \frac{1}{3}(R+G+B) \quad (6, 7, 8)$$

Table 1 Sample HSV Map of Figure 1.

H_q	S_q	V_q
0.87634	0.55357	0.21961
0.19298	0.39583	0.18824
0.15	0.19608	0.2
0.66288	0.49438	0.34902
0.65409	0.51456	0.40392
0.34	0.32051	0.30588
0.33333	0.35417	0.37647
0.50877	0.77869	0.47843
0.36667	0.33708	0.34902

Video is made up of sequence of images. Its frame rate is defined as number of images per second given in Eq. (9).

$$N = T \times R \tag{9}$$

N= Total number of frame in a video
R= Frame rate per second
T=Length of video in seconds

Figure 1. Sample Image

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue IX, September 2017- Available at www.ijraset.com

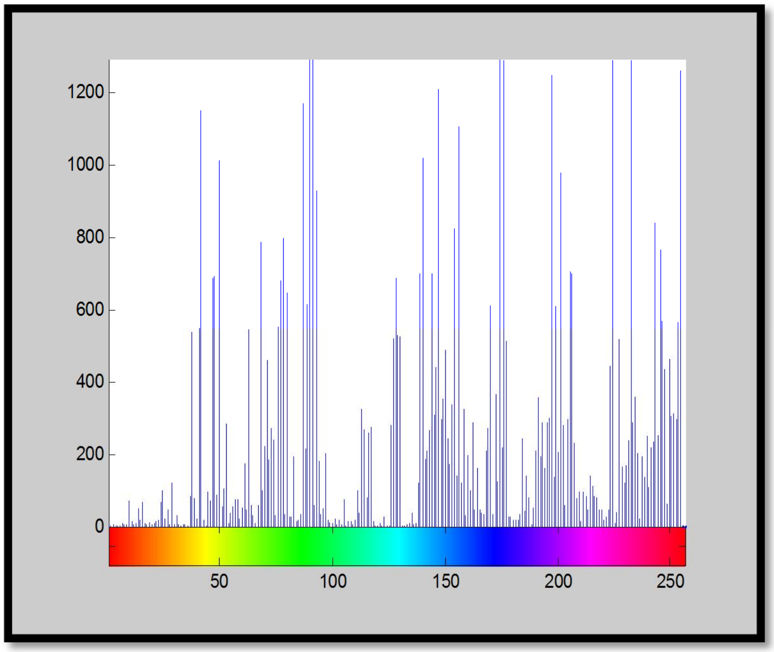


Figure 2. HSV histogram of image of Figure1

In this approach as shown in table 1 first and second order moments of HSV map has been calculated i.e. first and second order moments of each column is used as a color features and it can be found using equations Eq. (10), (11), (12), (13), (14), (15).

$$\mu_h = \frac{1}{256}\sum_{x=1}^{256} HSV(x,1)$$

$$\mu_s = \frac{1}{256}\sum_{y=1}^{256} HSV(y,2)$$

$$\mu_v = \frac{1}{256}\sum_{z=1}^{256} HSV(z,3) \qquad (10, 11, 12)$$

$$\sigma_h = \sum_{x=1}^{256} (HSV(x,1) - \mu_h)^2$$

$$\sigma_s = \sum_{y=1}^{256} (HSV(y,2) - \mu_s)^2$$

$$\sigma_v = \sum_{z=1}^{256} (HSV(z,3) - \mu_v)^2 \qquad (13,14,15)$$

V. EXPERIMENT RESULTS & DISCUSSION

The method was tested on video test set downloaded from youtube.com containing news, cartoon and cricket videos. All videos are containing significant camera effects like zoom-ins/outs, pans, and significant object and camera motion inside shot. Images are processed with its actual dimensions. Results are calculated on different types of video like cartoons, news, sports. We used

As shown in figures 3.graph of inter frame difference versus frame number for a video *sa2.avi*. It shows sharp cut or shot are differentiated by higher difference. These sharp cuts are generated because of camera switching from one shot to another and sometimes due to editing effects too. Figure 4. shows first six transitions detected as shot boundary for a video *sa2.avi*. As we can see during sharp transitions content are changing drastically. Same can be visualized in figure 3 too.

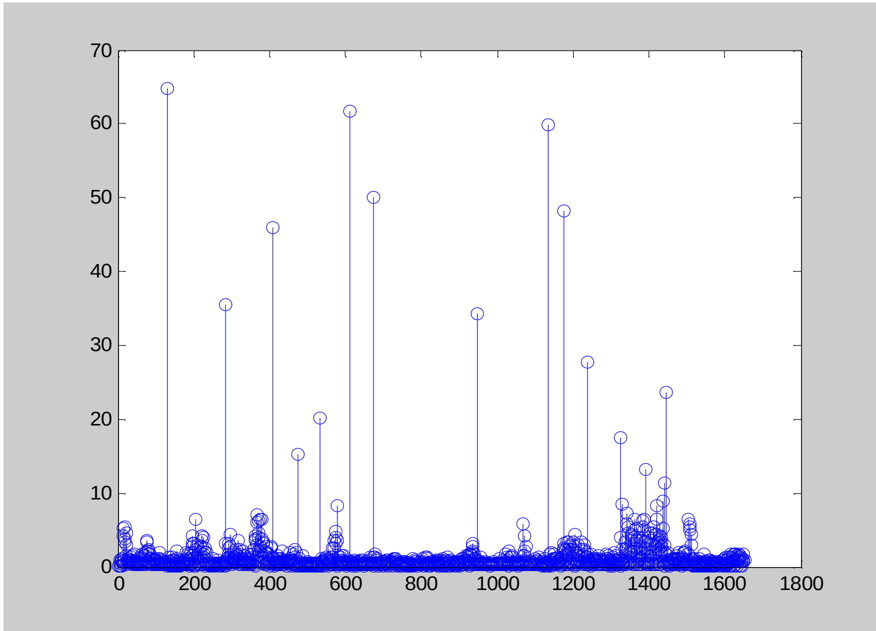


Figure 3. Frame Number Vs Frame Difference for *sa2.avi*

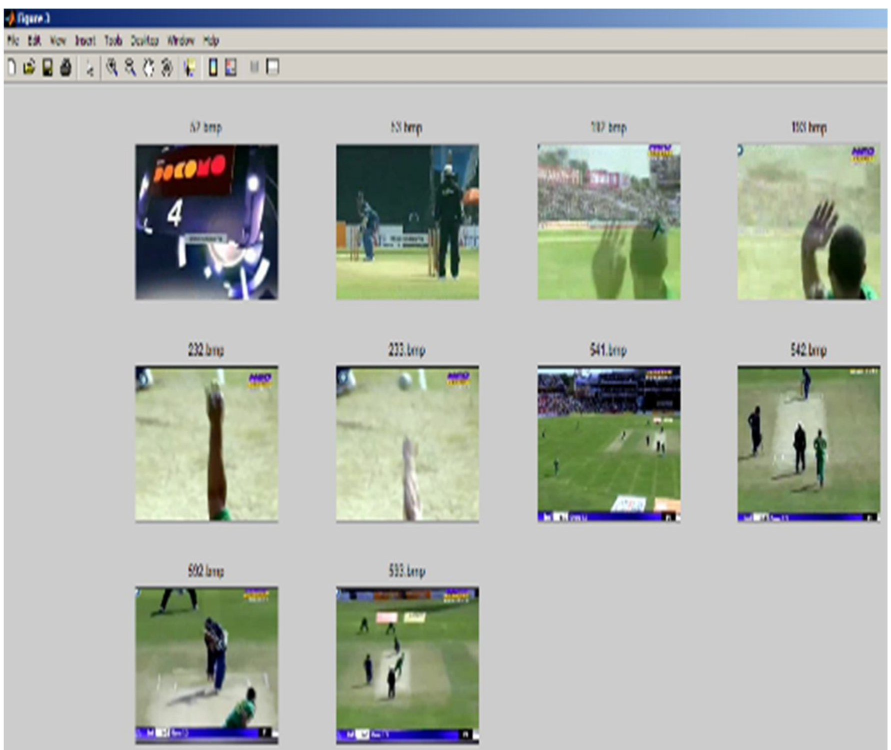


Figure 4. Shot Boundary detected for *sa2.avi*

There are certain shots in a video were shots are changing gradually. We observed method is finding out the sharp transitions as well performing well in gradual transition also. Most of the research has considered threshold for practical evaluation but threshold is changing depends on type, content of the videos. So in this method frame rate is used as a threshold to find shot boundary. Figure 5, 6, 7 and 8 show graph of Frame Number Vs Frame Difference and shot boundary detected for respective videos of a database. Results are shown in tabular and graphical form in Table 2 and 3 and in Figure 9.

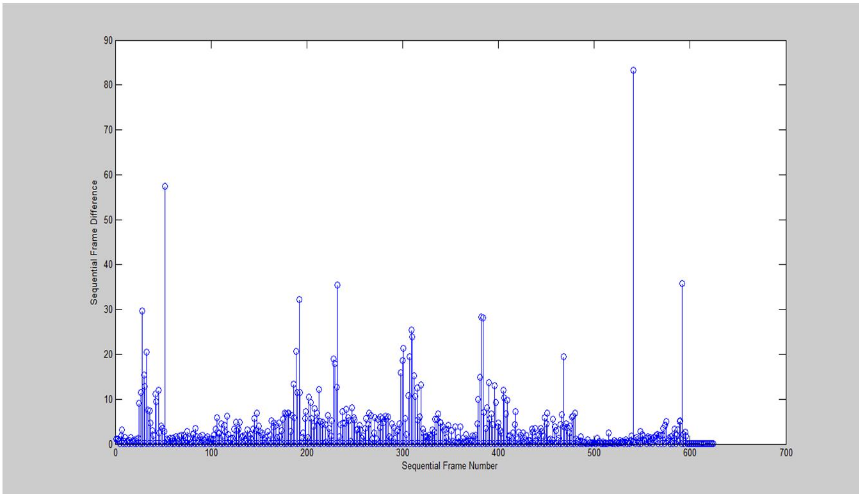


Figure 5. Frame Number Vs Frame Difference for *Test3.avi*

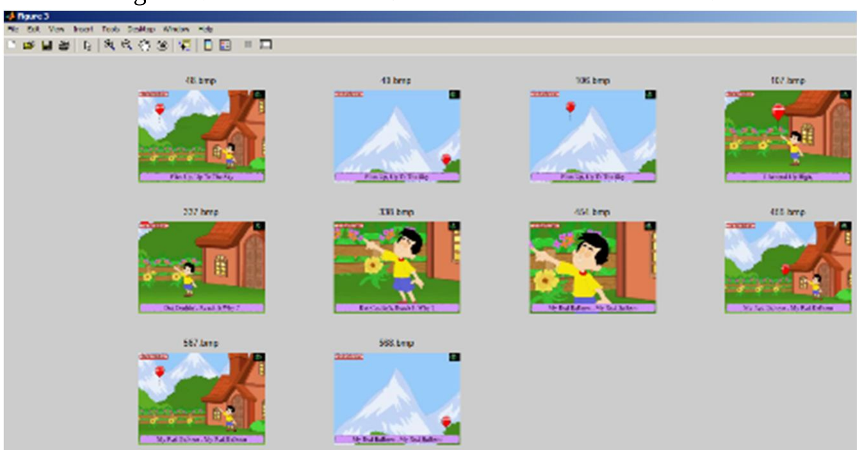


Figure 6. Shot Boundary detected for *Test3.avi*

Usually two measures recall and precision defined in (1) and (2) are used to measure the effectiveness of cut detection algorithm. We define these measures using %pass and % fail to retrieve a correct shot. RS defines retrieved shots, CS indicates correct shots and MS means missed shots.

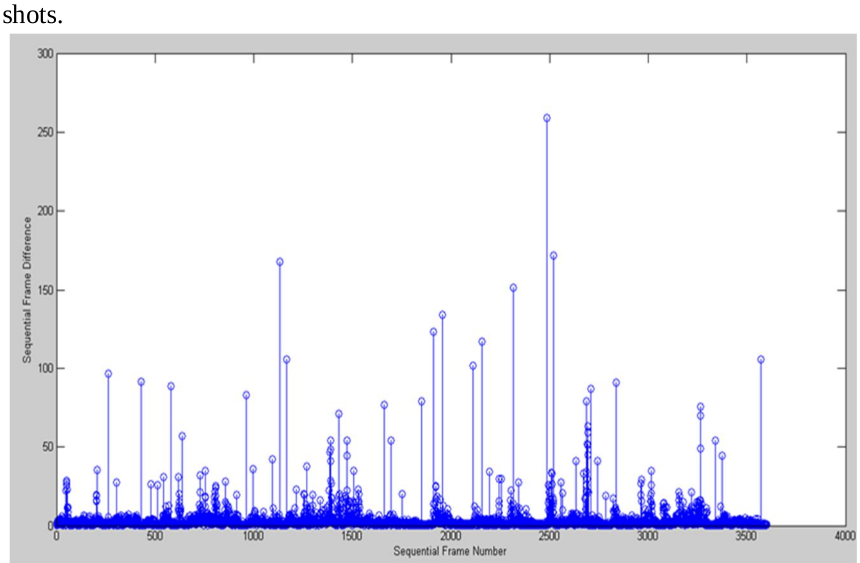


Figure 7.Frame Number Vs Frame Difference for *sa2.avi*

Figure 8. Shot boundary detected for *sa2.avi*

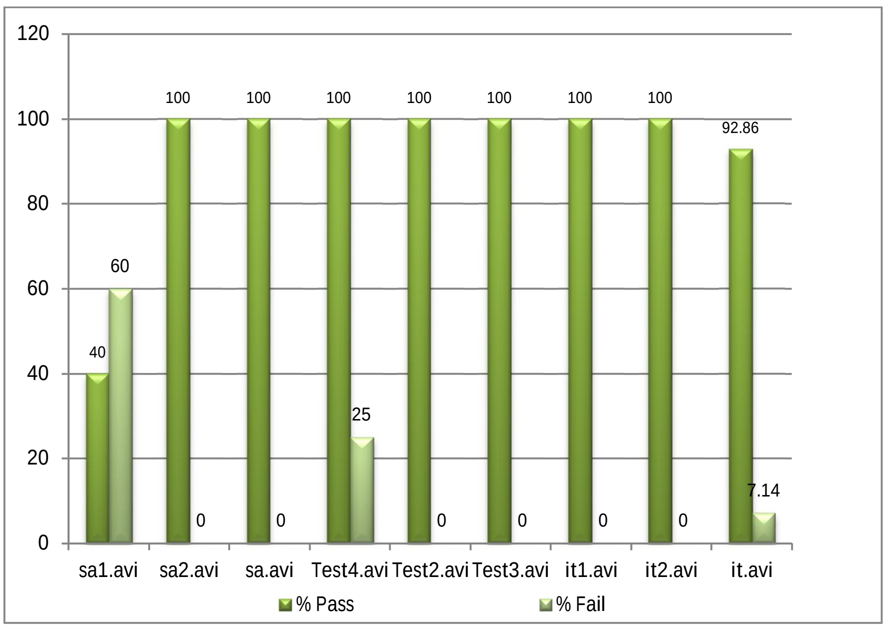


Figure 9. Experiment results

Table 2. Results of Shot boundary detection

Category	Video Name	Frames	RS	CS	MS
Sport	sa1.avi	598	5	2	3
	sa2.avi	598	5	5	0
	sa.avi	3602	12	12	0
Cartoon	Test4.avi	626	4	4	1
	Test2.avi	626	5	5	0
	Test3.avi	626	5	5	0
News	it1.avi	598	5	5	0
	it2.avi	598	5	5	0
	it.avi	4501	14	13	1

Table 3 Shot boundary evaluation

Category	Video Name	% Pass	% Fail
Sport	sa1.avi	40.00	60.00
	sa2.avi	100.00	0.00
	sa.avi	100.00	0.00
Cartoon	Test4.avi	100.00	25.00
	Test2.avi	100.00	0.00
	Test3.avi	100.00	0.00
News	it1.avi	100.00	0.00
	it2.avi	100.00	0.00
	it.avi	92.86	7.14

VI. CONCLUSION AND FUTURE WORK

Experiment work was carried out on different types of video like sports, cartoon, and news. HSV features were extracted from various types of videos and using its statistical features shot boundary was detected successfully. Proposed algorithm significantly improves results than other color histogram based methods. It consumes less time to detect boundary locations. During experiment work it has been observed that shot boundaries were clearly detected at the place were content or background was changing drastically or camera switching was taken place. In all categories of videos shot boundaries are detected with % pass rate more than 90%. In one of the sport category video pass rate obtained is less due to extreme gradual transitions. Gradual changes are detected in most of the cases. Frame rate i.e. assumption was camera shot must be of minimum one seconds gives very satisfactory results for both cut and gradual boundary detection. When boundary is detected but distance between previously detected boundary and current boundary is less than frame rate, it is considered as a continuation of previously detected shot. Because this in certain cases it was observed, though shot boundary was there but yet merging of consecutive shots taken place and so we missed the shot position there.

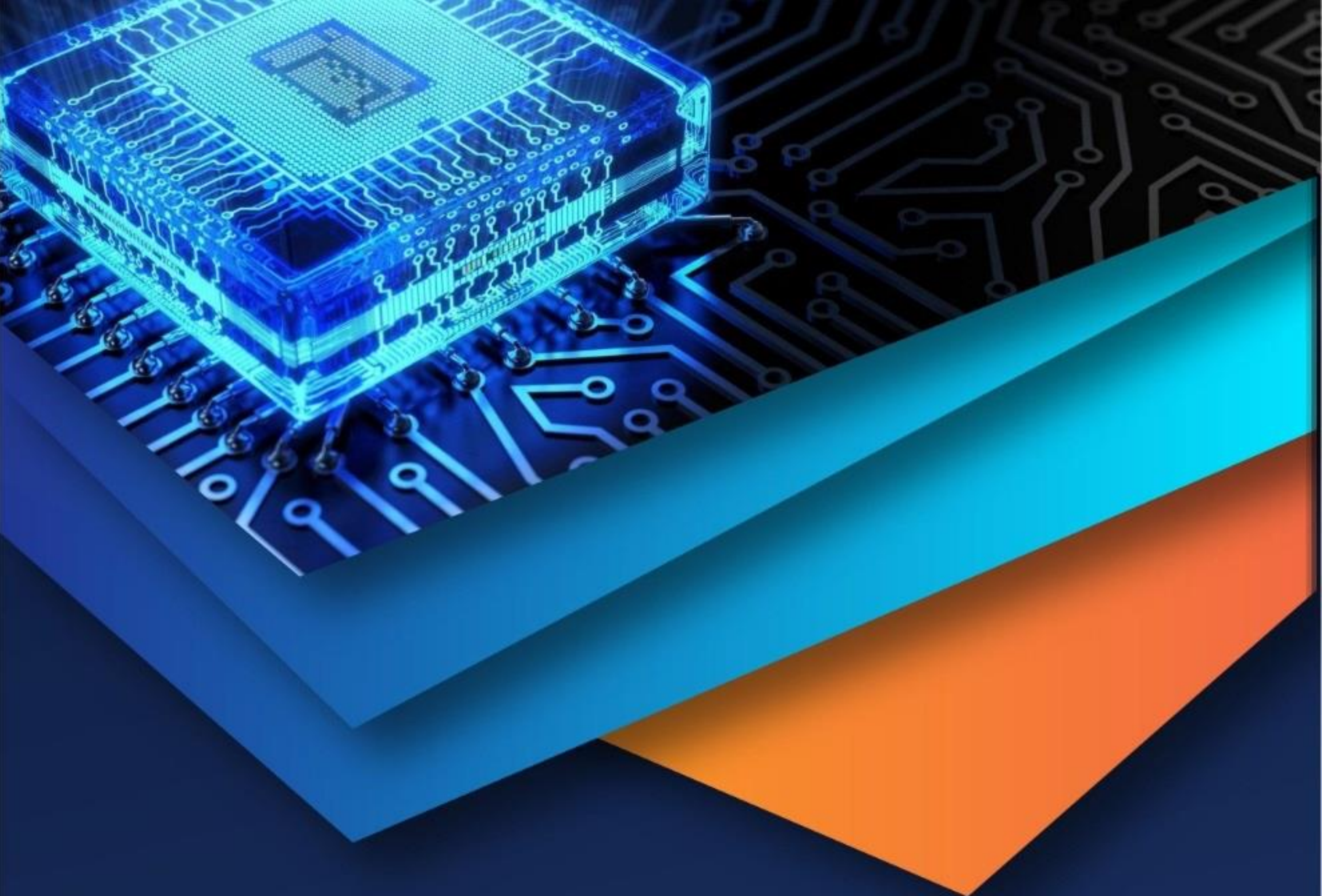
VII. ACKNOWLEDGMENT

Authors would like to thank Charutar Vidya Mandal and A.D. Patel Institute of Technology for supporting and providing resources for research work. Authors express his heartfelt gratitude for Electronics & Communication Engineering department.

REFERENCES

[1] Kankanhalli, and S. W. Smoliar H. Zhang, "Automatic partitioning of fullmotion video," vol. 1, no. 1, pp. 10-28, 1993.
[2] R Jain R Kasturi, "Computer Vision: Principles," pp. 469-480, 1991.
[3] Faeze Asdaghi, Mahmooh Fathy, Mohammad Reza Jahed-Motlagh Zeinab Zeinalpour-Tabrizi, "Gradual Shot Boundary Detection and Classification Based on Fractal Analysis," vol. 4, no. 8, 2010.
[4] D. A. Adjeroh and M. C. Lee, "Robust and efficient transform domain video sequence analysis: an approach from the generalized color ratio model," vol. 8, no. 2, pp. 182-207, 1997.
[5] J. Miller, and K. Mai R. Zabih, "A feature-based algorithm for detecting and classifying production effects," vol. 7, no. 2, pp. 119-128, 1999.
[6] Guillermo Cisneros, José M. Martínez, José M. Menéndez, and Julián Cabrera Jesús Bescós, "A Unified Model for Techniques on Video-Shot Transition Detection," vol. 7, no. 2, pp. 293-307, Apr. 2005.
[7] J. S. Boreczsky and L. A. Rowe, "Comparison of video shot boundary detection techniques," in SPIE, vol. 2670, Jan. 1996, pp. 170-179.
[8] Y. Tanaka A. Nagasaka, "Automatic video indexing and full video search for object appearances," pp. 113-127, 1995.
[9] A., Jain, R and Weymouth, T.E Hampapur, "Production model based digital video segmentation," pp. 9-46, 1995.
[10] R., Miller, J and Mai, K. Zabih, "A feature-based algorithm for detecting and classifying scene breaks," in International Multimedia Conference and Exhibition, California, 1995.
[11] A.M. Ferman, A.M. Tekalp B. Gunsel, "Temporal video segmentation using unsupervised clustering and semantic object tracking," pp. 592-604, 1998.
[12] T.N. Pappas, "An adaptive clustering algorithm for image segmentation," pp. 901-914, 1992.
[13] Mahmood Karimian, and Shohreh Kasaei Mostafa Tavassolipour, "Event Detection and Summarization in Soccer Videos Using Bayesian Network and Copula," vol. 24, no. 2, Feb. 2014.
[14] Alan Smeaton, Noel Murphy and Sean Marlow Colin O'Toole, "Evaluation of Automatic Shot Boundary Detection on a Large Video Test Suite ," 1999.
[15] J. Yu and M. D. Srinath, , 2001.
[16] L.-F. Cheong, "Computer Vision and Image Understanding," , 2004.
[17] Jordi Mas and Garbriel Fernandez, "Video Shot Boundary Detection based on Color Histogram".
[18] Junxiao Liu, "A Self-Adaptive Shot Boundary Detection Method Based on Neighbor Frame Difference and Non-neighbor
[19] Frame Difference," vol. 10, no. 7, pp. 23-30, 2015.

[20] Xuelog Li, Jun Feng, Dacheng Tao Xinbo Gao, "Shot based video retrieval with optical flow tensor and HMMs," Pattern recognition Letters , pp. 140-147, Mar. 2008.
[21] N. Nikolaidis, and I. Pitas C. Cotsaces, "Video shot detection and condensed representation. A review," vol. vol. 23, no. 2, pp. pp. 28-37, Mar. 2006.
[22] Zhe-Ming Lu and Yong Shi, "Fast Video Shot Boundary Detection Based," vol. 22, no. 12, Dec. 2013.
[23] Aristidis C. Likas, and Nikolaos P. Galatsanos Vasileios T. Chasanis, "Scene Detection in Videos Using Shot Clustering and Sequence Alignment," vol. 11, no. 1, Jan. 2009.
[24] M. C. Lee, N. Banda, and Uma Kandaswamy Don Adjeroh, "Adaptive Edge-Oriented Shot Boundary Detection," vol. 2009, no. Article ID 859371, May. 2009.
[25] M. J. Swain and D. H. Ballard, "Color indexing," vol. 7, no. 1, pp. 11-32, 1991.
[26] A. Nagasaka and Y. Tanaka, "Automatic video indexing and full video search for object appearances," 1992.
[27] Krishna K Warhade Shraddha C Nistane, "Gradual transition detection in movie videos: A challenge for shot boundary detection algorithm," International journal of Electrical, Electronics and computer science, vol. 1, no. 2, 2013.
[28] Kuk-Jin Yoon, and In So Kweon Seung-Hoon Han, "A new technique for shot detection and key frames selection in histogram space," in 12th Workshop on Image Processing and Image Understanding, 2000, pp. 475-479.
[29] N. Xie, L. Li, X. Zeng and S. Maybank W. Hu, "A Survey on Visual Content-Based Video Indexing and Retrieval," IEEE Transactions on Systems, Man, and Cybernetics, vol. 41, no. 6, pp. 797-819, Nov 2011.
[30] P. Geetha and V. Narayanan, "A Survey of Content-Based Video Retrieval," Journal of Computer Science, vol. 4, no. 6, pp. 474-486, 2008.
[31] N. Sebe, P. Gardner M. Lew, "Video indexing and understanding," Principles of visual information retrieval, pp. 163-196, 2001.
[32] Y. Ding, Y. Shi and Q. Zeng J. Li, "DWT-Based Shot Boundary Detection Using Support Vector Machine," in Information Assurance and Security, IAS '09, 2009, pp. 435-438.
[33] H. Y. Wang, and B. Zhang J. H. Yuan, "A formal study of shot boundary detection," Journal of Transactions on Circuits and Systems for Video Technology, vol. 17(2), pp. 168-186, February 2007.
[34] Sanghoon Sull and Sang Uk Lee Hyun Sung Chang, "Efficient video indexing scheme for content-based retrieva," IEEE Transactions on Circuits and Systems for Video Technology, vol. 9, no. 8, pp. 1269-1279, Dec 1999.
[35] A. Hanjalic, "Shot-Boundary Detection: Unraveled and Resolved?," IEEE Transactions on Circuits and Systems for Video Technology, vol. 12(02), pp. 90-150, 2002.
[36] W. WOLF H. H. YU, "A hierarchical multiresolution video shot Transition detection scheme," Journal of Computer Vision and Image Understanding, vol. 75(1/2), pp. 196-213, 1999.
[37] Y. W. Zhu, Q. B. Sun, and S. F Chang C. H Yeo, "A Framework for sub-window shot detection," , 2005, pp. 84-91.
[38] Duy DINH LE, Shini'ichi SATOH Thanh DUC NGO, "Scalable approaches for content based video retrieval," Progress in Informatics , vol. 11, pp. 31-39, 2014.
[39] Xu De, Gaun Tengfei, Wu Aimin, Lang Congyan Yang Xu, "3 DWT based motion suppression for video shot boundary detection," Lecture notes in Computer Science 2005.
[40] J Gu, H. Chen, Z. Tang T. Qin, "A fast shot-boundary detection based on K-step slipped window," in 2nd IEEE International Conference on Network Infrastructure and Digital Content, Sep. 2010, pp. 190-195.



10.22214/IJRASET

INDEX COPERNICUS 45.98

ISRA JIF IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL IMPACT FACTOR: 7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY