2015 IEEE Power, Communication and Information Technology Conference (PCITC)
Siksha 'O' Anusandhan University, Bhubaneswar, India.

A Texture Based Method for Scene Change Detection

T. Kar
School of Electronics Engineering
KIIT University, Bhubaneswar
e-mail: tkarfet@kiit.ac.in

P. Kanungo
Dept of Electronics &Telecommunication Engg.
C.V. Raman College of Engineering, Bhubaneswar
e-mail: pkanungo@gmail.com

Abstract— **Due to the growth of internet and multimedia technology, there is a tremendous need for an efficient method of automatic video retrieval and storage process. Automatic annotation is one of the key solutions for the automatic retrieval or search of a particular scene based frames or a particular scene based videos from the huge video database. Video segmentation based on shot boundary detection (Cut detection) is the fundamental step for the content based text annotation or video data analysis. In this paper, a new automatic cut detection method is proposed based on the local binary pattern feature. The proposed algorithm is tested with six test videos and its efficacy is validated with few existing popular approaches.**

Keywords-Shot Boundary Detection, Local binary Pattern, texture feature, Video segmentation.

I. INTRODUCTION

Due to the growth of internet technology and the popularity of social sites video sharing like uploading, downloading of video and content based video searching is increasing exponentially[1]. Due to the huge amount of videos it is a hurculian task for manual indexing and annotation. Therefore automatic text annotation is one of the key solutions for the content based video searching, retrieval and storage. Video segmentation [2] i.e. classifying the scene of a video into different shots is the initial step of any automatic annotation. Abrupt transition and gradual transition are the two basic transitions available in most of the videos. Abrupt transition, known as hard cut, is a sudden transition between two consecutive shots. If the transition is smooth and it exists for more than two frames then it is defined as gradual transition. Gradual transitions are classified as fade or dissolve. Various features used for shot boundary detection includes pixel intensity based method, histogram based method, edge oriented method, motion based approach, statistical feature based techniques that are reported in the literature in last two decades.

Pixel based shot detection are more simpler, accurate and popular in comparison to compressed domain shot detection. Histogram based method [3,4], block based histogram method [5], and pixel differencing method[6], and edge change ratio based cut detection[7] are the few cut detection methods, which are popular due to their simplicity. Recently Lakshmipriya et al.[8], proposed a shot boundary detection method based on edge strength using block based orthogonal features of a frame which handles the fast motion and lighting effects but fails in case of explosion events in a scene. Amel et al. [9] used the motion activity to detect the shot boundaries which can handle the camera motion as well as the object motion in a video. Whereas the experimental result shows that the highest performance in terms of 'Recall' is 84.21% in football videos and lowest performance is for films videos that is 66.66%.

Threshold is used as a parameter to distinguish between transition between two shots. Therefore threshold selection is one of the crucial and challenging task for the efficiency[10] of any shot detection algorithm. Jialei et al.[11] used the mutual information as a feature to evaluate the distance between two consecutive frames and used SVM to detect a threshold for localizing the shot boundary. Jiang et al.[12] proposed a dual detection model based on pre-detection and re-detection process. Uneven block colour histogram difference and pixel value difference are used as a distance feature whereas an adaptive binary search is used to locate the shot boundary from the distance features in the pre-detection step. In the re-detection process a scale invariant feature is used to refine the detected boundaries to reduce the false detection which improves the precision rate. Among all these methods histogram based cut detections are simplest and efficient whereas most of the time all these histogram based methods fail to eliminate false cuts. False cuts increase because of having variations of the histogram of two similar scene or two different frame of same shot. In this paper a new feature histogram based cut detection method is proposed to increase the efficiency of cut detection process.

The rest of the paper is organized as follows. A description of few related work is reported in section II. The texture description using local binary pattern is described in section III. Proposed cut detection method and its algorithm reported in section IV. Simulation results and discussions followed by conclusions are given in section V and VI respectively.

II. RELATED WORK

Various shot boundary detection algorithms has been proposed in literature. Edge change ratio based method, Absolute sum histogram based method pixel wise intensity difference method are most popular cut detection algorithms due to their simplicity and efficacy. Therefore in this section these methods are explained and the

978-1-4799-7455-9/15/$31.00 ©2015 IEEE

performances of these methods are also compared with the proposed method in section V.

A. Edge Change Ratio based shot boundary detection

Edge change ratio (ECR) based cut detection is proposed by Zabih et al.[7]. In this approach edge feature is extracted for each and every frame of a video. An edge change ratio between two consecutive frames is evaluated as follows in (1).

$$ECR_t = \max\left(\frac{E_{t+1}^{in}}{\sigma_{t+1}}, \frac{E_t^{out}}{\sigma_t}\right) \quad (1)$$

Where, σ_t is the number of edge pixels in frame t, E_{t+1}^{in} and E_t^{out} are the number of entering and exiting edge pixels in the $(t+1)^{th}$ and t^{th} frame respectively. If ECR value is greater than a threshold TH, a cut is declared between t^{th} and $(t+1)^{th}$ frame. The method works efficiently if there are very slow motion objects in a particular shot, whereas the efficacy of this method decreases in presence of fast object motion.

B. Absolute Sum based Histogram differencing method

It is one of the simplest method which uses the histogram feature for the cut detection in a video. This histogram based method [4] computes the gray level histogram difference of the two images known as absolute sum of histogram difference (ASHD). The ASHD between two consecutive frames is evaluated as in (2).

$$ASHD_{t,t+1} = \sum_{i=0}^{255} \left| h_t(i) - h_{t+1}(i) \right| \quad (2)$$

Where, h_t *and* h_{t+1} are the histograms of t^{th} and $(t+1)^{th}$ frames respectively. If the $ASHD_{t,t+1}$ in (2) is above a threshold value then a shot boundary is detected in between t^{th} and $(t+1)^{th}$ frame. This method also fails for the fact that two different images can have same histogram as histogram representation ignores the spatial distribution of images.

C. Pixel-wise intensity difference method

The sum of absolute intensity difference [6] (ASID) between two corresponding pixels of two consecutive frames is calculated as in (3). This ASID is compared against a threshold for detecting a cut between two consecutive frames. This method fails for large object motion and camera motion.

$$ASID_{t,t+1} = \sum_{i=0}^{R} \sum_{j=0}^{C} \left| f_t(i,j) - f_{t+1}(i,j) \right| \quad (3)$$

Where f_t and f_{t+1} are the t^{th} and $(t+1)^{th}$ frame of a video, R and C are the number of rows and columns of a frame.

Among all these methods ASHD is the simplest method and it considers only the gray level distribution information. Hence many times the performance of ASHD method degrades which motivated us to use the texture information instead of gray level information as texture feature represents the spatial correlation of pixel intensities in a selected region.The Local binary pattern (LBP) is one of the feature used for texture representation. This featured histogram of the texture information will have different histogram even if the graylevel histogram are same for two different scenes.

III. TEXTURE DESCRIPTION USING LOCAL BINARY PATTERN

Local binary pattern (LBP) is one of the important features used for texture analysis. Most of the literature used circular neighborhood structure to evaluate the binary pattern[13]. Ojala et.al [14] used 8 neighboring pixels for evaluation of LBP to capture the texture feature. In this paper a 3x3 neighborhood structure as shown in Fig. 1 is considered for the evaluation of local binary pattern(LBP) feature.

Illustration of LBP and spatial co-rrelation feature:

Let's consider a 3x3 neighborhood pixel centered at (x_c, y_c) coordinate having gray value of g_c =6 as shown in Fig. 1(a). Let the position of each neighboring pixel is defined as a variable "p". The neighboring pixel in bold face having gray value 6 has p=0. Similarly the position of each pixel is increased by 1 in clockwise direction. The gray value of each position are g(0)=6, g(1)=5, g(2)=2, g(3)=3 g(4)=8, g(5)=7, g(6)=9, g(7)=5. The local binary pattern for Fig. 1 (a) is given in Fig. 1 (b) where the binary pattern is **1** 0 0 0 1 1 1 0.

6	5	2
5	6	3
9	7	8

1	0	0
0		0
1	1	1

1	2	4
128		8
64	32	16

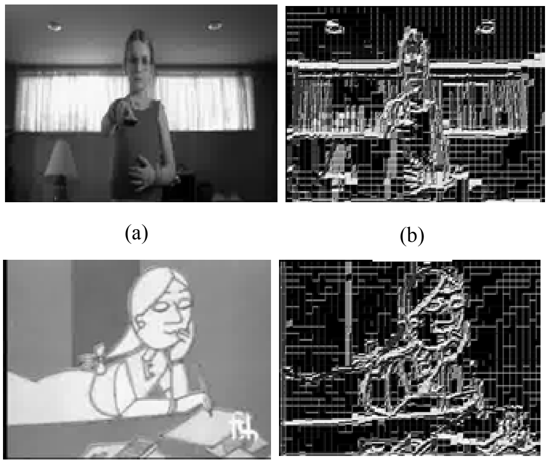
Fig 1. (a) 3x3 neighborhood, (b) LBP binary pattern: clockwise from the bold facet **1** (top left corner)i.e 10001110, (c) decimal value of each place.

$$LBPF\ (x_c, y_c) = \sum_{p=0}^{7} S(p) 2^p \quad (4)$$

$$\text{Where, S(p)=}\begin{cases} 1 & g(p) \geq g_c \\ 0 & g(p) < g_c \end{cases} \quad (5)$$

The spatial correlation feature of Fig. 1(a) is LBPF (x_c, y_c) = 1+16+32+64=113. Where g_c is the gray value in the 3x3 neighborhood.

The most important properties of LBP are it is invariant to any monotonic gray level change and it is computationally simple. It's importance in applying to video shot detection is that it is illumination invariant [13,14]. But LBP fails for flat image areas such as sky where the gray value of the neighbouring pixels are very close to the value of the centre pixel. The local binary pattern feature (LBPF) image is generated using (4) which is shown in Fig. 2 for two different frames of two different videos.



(a) (b)

(c) (d)

Fig 2. (a)A gray scale frame no 1406 of littlemiss sunshine, (b)LBPF image of (a), (c) A gray scale frame no 101 of cartoon ekchidiya video, (d) LBPF image of (c).

IV. Proposed Method

The efficacy of most of the methods reported in section II decreased in case of rapid object motion or slight illumination variation. The most simplest and efficient method is the ASHD based cut detection but it also failed when two different scene have same histogram, because histogram information does not contain any neighborhood or spatial relation. Motivated by the spatial correlation property of LBP, a new method known as "Absolute Sum Local Binary Pattern Histogram Difference" (ASLBPHD) is proposed to detect the cuts in a video more efficiently.

A. ASLBPHD based cut detection

In this proposed method, the texture based feature is extracted using local binary pattern. Instead of using the gray level histogram a new feature histogram based on texture information is generated from the texture image . This texture image is created based on LBP which is shown in Fig. 2. The extracted feature helps to minimize the dissimilarity between two frames of same shot and maximizes the dissimilarity between two frames of two different shots. Using this feature a LBPF image using (4) is generated and the histogram of this LBPF image i.e. LBPH is evaluated. The sum of absolute LBPH difference is evaluated using (6) which is used as a similarity measure to detect a cut between two consecutive frames.

$$ASLBPHD_{t,t+1} = \sum_{i=0}^{255} \left| LBPH_t - LBPH_{t+1} \right| \quad (6)$$

Where, $ASLBPHD_{t,t+1}$ is the absolute sum difference between the LBP histogram of tth frame and (t+1)th frame of a video respectively. Using this $ASLBPHD_{t,t+1}$ values and a threshold value TH, cuts are detected using (7).

$$C(t) = \begin{cases} 1 & ASLBPHD_{t,t+1} \geq TH \\ 0 & otherwise \end{cases} \quad (7)$$

Where, t is the frame number, if C(t)=1 then there is a cut existing at tth frame.

B. The Proposed Algorithms

Step1: Convert the RGB color frames of the video into gray scale frames.

Step2: Convert each gray level frame into LBP feature frame.

Step3: Generate LBP feature histogram (LBPFH) from the LBP feature frames of step2.

Step4: Evaluate the absolute sum LBPFH difference of consecutive frames using (6) and evaluate C using (7) for all frames.

Step5: If for the tth frame, C(t)=1 then declare a cut at the tth frame.

V. Simulations and Discussions

The cut detection is basically used for text annotation or video indexing . The efficacy of the cut detection may not be limited to a particular database rather the efficacy should be evaluated for all kind of videos. Therefore in this paper five different varieties of videos i.e Littlemiss Sunshine(LM), The Big Bang Theory(TBBT), Ekchidiya(Ekchi), Masoom and Before Sunrise(BS) are considered for testing and validation of the proposed cut detection method. Littlemiss Sunshine is an english movie which has strong light variation in it. The Big Bang Theory is a sitcom video which has been used by many literature earlier. "Ekchi" is a cartoon video which has less textural information. "Masoom" is a hindi film song clipping and "Before Sunrise" is an english movie having high speed object motion. This list of videos along with the number of frames, duration and ground truth cuts are tabulated in Table 1. All these videos are uncompressed AVI format with a spatial resolution of 320x240 and gray level resolution of 256. The proposed method is compared with ASHD, ECR and PID based method using three major performance criteria i.e recall, precision and F1 measure[15]. performance measures are defined as in (8), (9), (10) respectively.

$$\mathrm{Re}call(R) = \frac{D_{TC}}{T_C} \quad (8)$$

Where D_{TC} is the number of detected true cut by the algorithm . T_C is the actual number of true cuts in a video which is equal to the ground truth cuts tabulated in Table 1.

$$\mathrm{Pr}ecision(P) = \frac{D_{TC}}{D_C} = \frac{D_{TC}}{D_{TC} + F_C} \quad (9)$$

Where D_C is the number of cuts detected and F_C is the number of false cut.

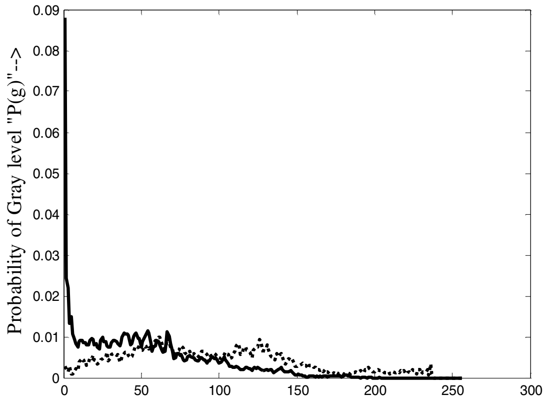
$$F_1 = \frac{2 * R * P}{R + P} \quad (10)$$

where F_1 is the harmonic average of R and P.

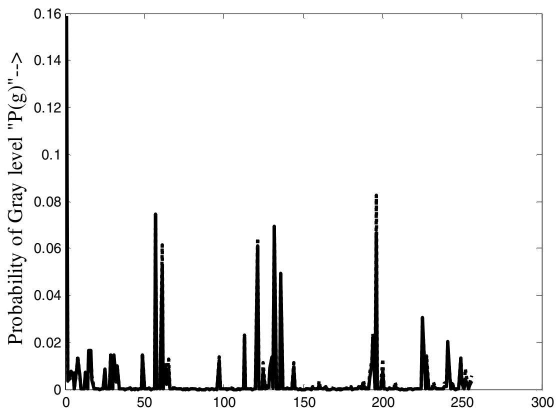
TABLE I. DESCRIPTION OF THE VIDEO SEQUENCE

Video Sequence	Number of frames	Duration (sec)	Ground Truth cuts	Threshold
LM	4000	133	31	0.23
TBBT(Sitcom)	3000	95	32	0.16
Ekchi (cartoon)	3000	100	39	0.26
Masoom(hindi song)	2000	66	9	0.23
BS	2000	66	12	0.18

The proposed method captures the texture information rather than the intensity of the pixel. Therefore the histogram differencing method which considers the intensity information has more false cut and high missed cut where as in proposed method the number of false cuts and missed cuts are substantially low in comparison to the other cut detection algorithms. Hence the efficacy of proposed algorithm is best in many test videos like LM, TBBT and BS videos. Fig. 3(a) shows the histograms of frame 103 and 104 of cartoon video “ekchidiya”.

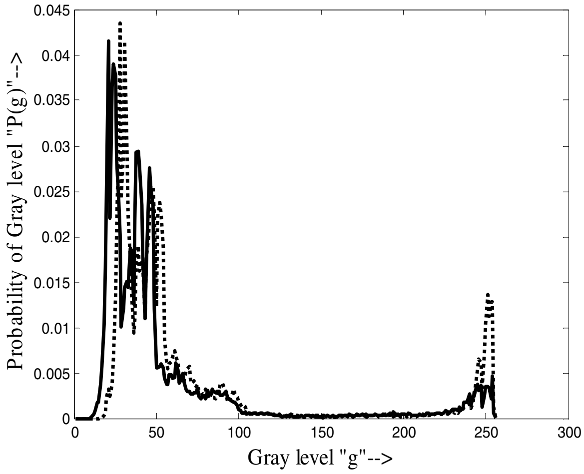


(a)

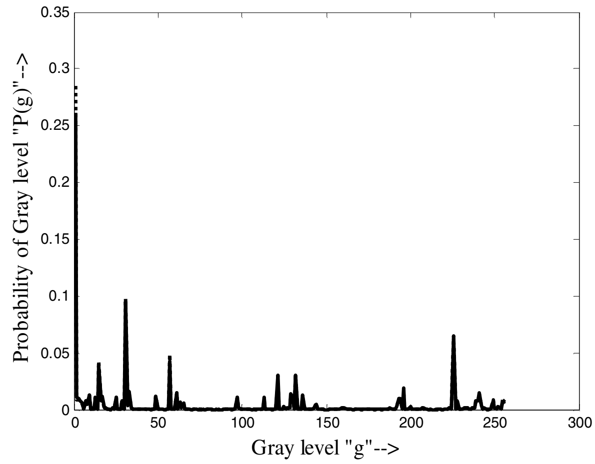


(b)

Fig 3. Comparison of histograms of frame 103 and 104 of littlemiss Sunshine. (a) Graylevel histogram (b)LBPF histogram



(a)



(b)

Fig 4. Comparison of histograms of frame 1429 and 1430 of Before Sunrise: (a) Gray level histogram (b) LBPF histogram.

During the ground truth evaluation it has been found that no true cut exists between these two frames where as in the Fig. 3(a) it has been found that there is a huge variation between the histogram of these two frames. Hence ASHD method detects a false cut between these two frames. But LBPF histogram of these two frames are very close to each other which is shown in Fig. 3(b). Hence LBPFH based method does not detect a cut in between these two frames. Similar study is carried out for the video “BS” for frame 1429 and 1430. It has been found that the ASHD detects a false cut between these two frames whereas LBPFH based approach does not detect a false cut, which is shown in Fig. 4(a) and Fig. 4(b) respectively. Therefore the number of false cut is very less in the proposed method in comparison to ASHD, ECR and PID methods. The performance measure Recall, Precision, and F1 for all videos with ASHD, ECR, PID and the proposed ASLBPFH based method is shown in Fig.5, 6, and 7 respectively.

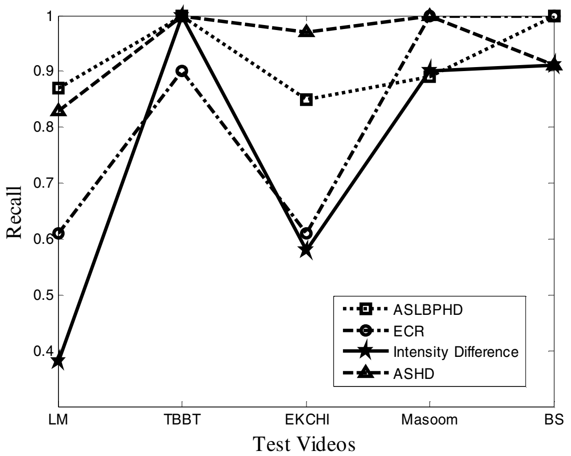


Fig 5. Recall measure(Normalised Scale)

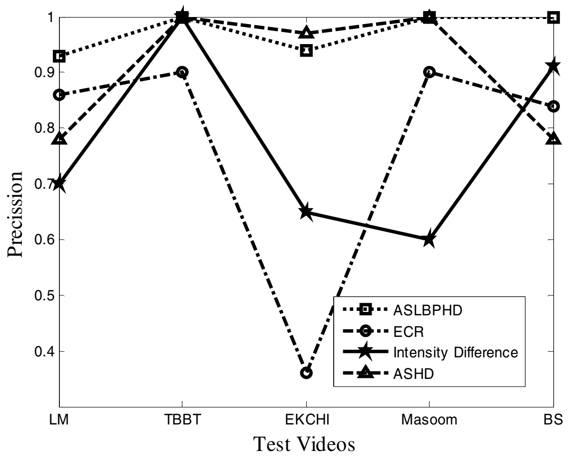


Fig 6. Precission Measure(Normalised scale)

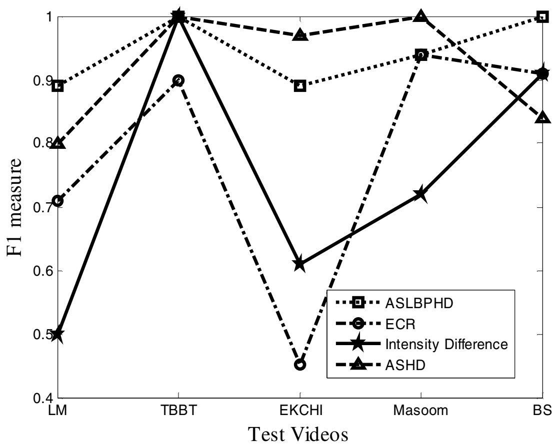


Fig 7. F1 measure(Normalised scale)

In the performance measure the proposed method has the highest recall measure in comparison to other methods for LM, TBBT and BS videos. But for "Ekchi" and "Masoom" it has the second highest recall measure. Therefore average true cut detection is higher in case of the proposed method in comparison to other methods. In precision measure the proposed method has the highest performance for four videos and second highest for "Ekchi" video. Therefore average false cut detection is less in the proposed algorithm in comparison to other approaches. Similar observations are also made for F1 measure. Therefore the proposed method is more efficient in the detection of hard cut for all varieties of videos. It has been observed that the proposed method failed in case of cartoon "Ekchidiya" video because of negligible texure informattion.

VI. CONCLUSIONS

The performance of most of the cut detection methods depends on the number of detected true cuts and false cuts . Further to improve the performance of any cut detection algorithm the objective is to increase the number of true detected cuts and simultaneously decreasing the number of false cuts. In the present work three different cut detection methods ASHD, ECR, and PID methods are extensively studied and their performances are evaluated. It has been observed that ASHD based cut detection outperforms ECR and PID. Whereas the performance of ASHD can be improved by reducing the no of false cuts and missed cut. In this paper a new cut detection method is proposed based on the texture feature extracted from the local binary pattern. It has been found that the proposed method out performs the ECR and PID in all the test videos whereas the proposed method outperforms the ASHD in most of the videos except synthetic videos and a song video of film "Masoom". In future work our objective is to increase the true cut detection and gradual transition detection capturing the rotational invariance feature of the local binary pattern and evaluation of an automatic threshold.

REFERENCES

[1] Murat Birinci,Serkan Kiranyaz,"A Perceptual Scheme for Fully Automatic Video Shot Boundary Detection," signal processing: image communication,pp.410-423,vol.29,2014.

[2] Linda G. Shapiro, George C.Stockman, computer vision,Printice Hall, ISBN:0-13-0307-963, 2001.

[3] U. Gargi, R. Kasturi and S.H. Strayer,"Performance characterisation of video shot change detection methods," IEEE Trans. on Circuits and system for Video Technology, CSVT-10(1),pp.1-13,2000.

[4] Rainer Lienhart, "Comparison of automatic shot boundary detection algorithm,"Image and video processing VII, in proc. of SPIE 3656-29,1999.

[5] G.G. Lakshmi Priya, S. Domnic, "Video Cut Detection using Block based Histogram Differences in RGB Colour Space," International conference on Signal and Image Processing, pp.29-33,2010.

[6] Boreczky JS, Rowe LA,"Comparison of video shot boundary detection techniques",storage and retrieval for still image and video databases IV,in Proc. of the SPIE 2670, san Jose,CA,USA,p.170-179,1996.

[7] R.Zabih, J.Miller and K.Mai, "A feature based algorithm for detecting and classifying scene breaks," ACM multimedia 95, San Fransisco, CA, pp.189-200,1995.

[8] G. G. Lakshmi Priya, S. Domnic."Edge Strength Extraction using Orthogonal Vector for Shot Boundary Detection," Procedia Technology(6),pp.247-254,2012.

[9] Abdelati Malek Amel, Ben Abdelali Abdessalem and Mtibaa Abde llatif "Video Shot boundary detection using motion activity Descriptor, Journal of Telecommunications,"vol.2(1),pp.54-59,2010.

[10] H. Lu and Y. Tan, "An effective post refinement method for shot boundary detection," IEEE Trans. Circuits system. Video Technol. **15**(11).p. 1407-1421,2005.

[11] Jialei Bi, Xianglong Liu, Bo Lang, "A Novel Shot Boundary Detection Based On Information Theory using SVM" 4th

International Congress on Image and Signal Processing,pp.512-516, 2011.

[12] Xinghao Jiang, Tanfeng Sun, Jin Liu,Juan Chao,Wensheng Zhang. "An Adaptive Video Shot Segmentation Scheme Based on Dual-detection model," Nurocomputing,vol.116,pp.102-111,2013.

[13] T. Ojala, M. Pietikainen, and D. Harwood, "A Comparative Study of Texture Measures with Classification Based on Feature Distribution," Pattern Recognition,vol. 29,no. 1, pp. 51-59,1996.

[14] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns"IEEE Trans. Pattern Analysis and Machine Intelligence. Vol. 24, no. 7, pp.971-987, July 2002.

[15] Krishna K. Warhade, Shabbier N. Merchant and U. B. Desai,"Performance Evaluation of Shot Boundary Detection Metrics in the presence of Object and Camera Motion," IETE journal of research,vol 57(5),pp.461-466,2011.