World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:10, No:7, 2016

Shot Boundary Detection Using Octagon Square Search Pattern

J. Kavitha, S. Sowmyayani, P. Arockia Jansi Rani

Abstract**—**In this paper, a shot boundary detection method is presented using octagon square search pattern. The color, edge, motion and texture features of each frame are extracted and used in shot boundary detection. The motion feature is extracted using octagon square search pattern. Then, the transition detection method is capable of detecting the shot or non-shot boundaries in the video using the feature weight values. Experimental results are evaluated in TRECVID video test set containing various types of shot transition with lighting effects, object and camera movement within the shots. Further, this paper compares the experimental results of the proposed method with existing methods. It shows that the proposed method outperforms the state-of-art methods for shot boundary detection.

Keywords**—**content-based indexing and retrieval, cut transition detection, discrete wavelet transform, shot boundary detection, video source.

I. Introduction

CONTENT Based Video Retrieval (CBVR) mainly focuses on automatic indexing, retrieval and management of video data. Video shot detection plays major role in digital video analysis. It is used as basics in various emerging applications such as video structure analysis, video summarization, video indexing and video retrieval. In film-making and video production, a shot is a series of frames, which runs for an uninterrupted period of time [1]. A video shot is defined as a sequence of frames captured uninterruptedly by a single camera [2].

Various levels of video structures are frame, shot, scene, etc. Among these, shot level is considered to be an appropriate stage for browsing and content based retrieval. Shots are combined during the editing process using transition techniques. The transitions are used to express emotion, ideas, etc. The basic transition types are abrupt and gradual. The abrupt transition refers the sudden change from one shot to the other. It is also called as cut transition. The gradual transition refers the gradual change between two shots, which mean a series of frames that belong to both the first and second video shot. This type of transitions is classified into dissolve, wipe, fade in and fade out. The most basic type of shot transition is cut transition which is used in the most of the video editing process. So, this work is used to find out cut transition in the video.

The researchers have developed many shot boundary detection techniques to detect the shot transitions in the video. The shot boundary detection process affects its performance due to the presence of the illumination effects, object or camera motion and special effects. H. Zhang et al. [3] presented a comprehensive study of a partitioning system that detects segment boundaries. In [4], the authors have compared the content of two consecutive frames using the edge change ratio (ECR). Shahraray [5] uses block matching and motion estimation to detect shot boundaries. A block-matching process is performed for each block in the first image to find the best fitting region of the second image. The video models [6] are used to classify the transition effects used in video and to design automatic edit effect detection algorithms. The performance of the various shot-change detection algorithms that use color histogram, MPEG data and block marching methods are evaluated and characterized by [7]. Lienhart [8] presented a survey of different core concepts underlying the different detection schemes for the three most widely used video transition effects: hard cuts, fades and dissolves.

Hidden Markov model (HMM) based shoat detection method is developed by Zhang et al. [9]. HMM is built to each type of transitions for identifying shape. Also, statistical comer change ratio and HSV color histogram is detected for identifying appearance. Kawai [10] proposed a shot boundary detection algorithm based on the block matching motion estimation. Adaptive edge oriented framework for shot boundary detection [11] was developed by Don et al. Shiguo Lian [12] developed an automatic video temporal segmentation based on multiple features. This algorithm computes the frame difference or similarity by such simple features as pixel difference and histogram differences adopts motion-based difference to resist camera or object movements in the same shot and uses the flash detection to avoid false positives caused by light changes or flashes.

Domnic et al. [13] proposed a shot boundary detection algorithm which extracts feature vectors from GLCM of the Hilbert transformed video. Choudhury et al. [14] introduced a new approach for graph based video clustering using modularity maximization. This method is used to detect shots in a video. The accurate edge localization property of Kirsch operator is used to extract features such a mean and standard deviation that is used for shot boundary detection work [15].

In this paper, shot boundary is detected using octagon square search pattern. The rest of the paper is organized as follows: Section II describes the system architecture of the proposed shot boundary detection technique and Section III discusses the design of the proposed shot boundary detection technique. In Section IV, experimental results and analysis are presented and finally, concluding remarks are furnished in Section V.

J. Kavitha, S. Sowmyayani and P. Arockia Jansi Rani are with Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Tirunelveli, India (e-mail: vsskavitha@gmail.com, sowmyayani@gmail.com, jansimsuniv@gmail.com).

II. System Architecture

This is the modification work of shot boundary detection based on DWT and Texture Filtering [17]. First, the input video is selected from the video source. Then, it is split into frames. Color, edge, motion and texture features of each frame are extracted using predefined functions. For detecting abrupt transitions, adjacent frame differences (both left and right) and similarities are compared. Then weights of the feature are calculated based on the significance of the individual feature. The transition detection method decides a frame which exists in boundary or non-boundary according to the weight of the features. Fig. 1 depicts the overall system architecture of the presented method.

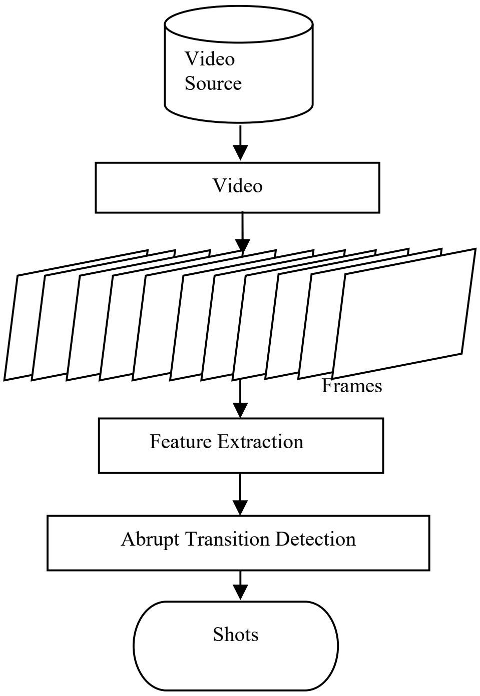

Fig. 1 System Architecture of the proposed work

III. The Proposed Method

A. Color Feature Extraction

Color is one of the important feature in shot boundary detection. In each frame, absolute difference of color histogram (F_C) is calculated for previous frame (x_i-1) to current frame (x_i) and current frame (x_i) to next frame (x_i+1). Previous frame (x_i-1) to current frame (x_i) is termed as left direction. Current frame (x_i) to next frame (x_i+1) is termed as right direction. Then the similarity between the left direction and right direction is calculated using Euclidean distance measure. This process is started from the second frame. Fig. 2 shows the computation of color feature value of the frame. Frames are denoted by f_1, f_2, f_3…f_{n-1}, f_n. Color histogram feature of each frame is termed as Cf_i.

- Left direction(LD): $f_{i-1}$and f_i
- Right direction(RD): f_i and f_{i+1}

$$Cf_i = \sqrt{|F_C(f_{i-1}, f_i) - F_C(f_i, f_{i+1})|^2} \quad (1)$$

B. Edge Feature Extraction

The first level Discrete Wavelet Transform (DWT) has four sub band images namely HL, LH, HH and LL. Among them first level diagonal detail coefficient (HH) of each frame has enough edge information. So, it is used in this method for reducing computation time. In each HH band, block based sobel edge detection algorithm is used for calculating edge feature. The edge feature of each frame (F_E) is calculated from left and right direction adjacent frames. Then, the similarity between the left direction and right direction is calculated using Euclidean distance measure. This process is started from the second frame and the block value is four. Edge feature of each frame is termed as Ef_i.

$$Ef_i = \sqrt{|F_E(f_{i-1}, f_i) - F_E(f_i, f_{i+1})|^2} \quad (2)$$

C. Motion Feature Extraction

The first level DWT diagonal detail coefficient (HH) is used for motion vector estimation. Octagon square search pattern is applied for detecting motion estimation [16]. A block-matching process is performed for each block in the first frame to find the best fitting region of the second frame. The motion vector of each frame (F_{mv}) is calculated from left and right direction adjacent frames. Then the similarity between the left direction and right direction is calculated by using Euclidean distance measure. This process is started from the second frame and block value is 16. Motion feature of each frame is termed as Mf_i..

$$Mf_i = \sqrt{|F_{mv}(f_{i-1}, f_i) - F_{mv}(f_i, f_{i+1})|^2} \quad (3)$$

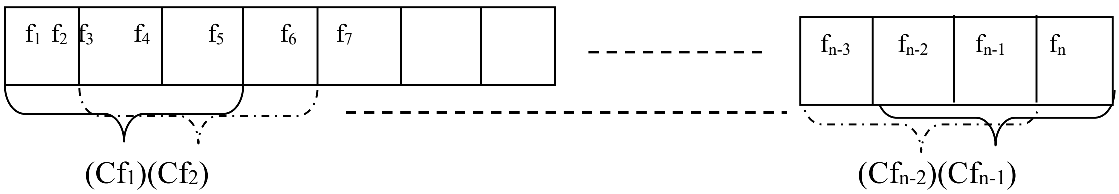

Fig. 2 Computation of color, edge and motion feature value of the frame

D. Texture Feature Extraction

The speed-up robust feature (SURF) is a detector and descriptor of points of interest. The descriptor is used to detect the range of matching between two images. Initially, first frame in the video is act as reference frame. The SURF feature points (Tf_i) of the reference frame and each frame in the video are extracted ($\mathfrak{T}$). If they are very different, then there may be a shot boundary and the latter frame became a reference frame. The non-matching features are denoted as nMF. If the matching ratio is high, then shot boundary does not exist between the two frames, it is referred as matching features (MF). Sometimes the motion with in the shot affects the shot detection process. So, the reference frame is changed at every 50th frame in the video.

$$\mathfrak{T} = \begin{cases} 1 & if\ (Tf_i \in nMF) \\ 0 & if\ (Tf_i \in MF) \end{cases} \quad (4)$$

E. Abrupt Transition Detection

The accuracy of the transition detection process is influenced by the weight value. Each weight feature is calculated based on the individual frame feature. The mean value of color and edge features of the current frame and the next frame is computed for α value. The β value is calculated according to the value of motion vector values. The normalization process is performed for each feature to represent the feature values in the range of 0 to 1. Weight values of α, β, Cf_i and Ef_i are used in transition detection process for detecting shot boundary. They are discussed in Section IV *C*. The transition detection method decides a frame which exists in boundary or non-boundary according to the weight of the features.

$$\alpha = \frac{1}{n}\sum_{i=1}^{n} Cf_i + Cf_{i+1} + Ef_i + Ef_{i+1} \quad (5)$$

$$\beta = \frac{1}{n}\sum_{i=1}^{n} Mf_{i-1} + Mf_i + Mf_{i+1i} \quad (6)$$

$$\Phi = \begin{cases} transition & if\ (\alpha_i > \omega_1, \beta_i < \omega_2, Cf_i > \omega_3, Ef_i > \omega_4, Tf_i \in nMF) \\ nontransition & otherwise \end{cases} \quad (7)$$

IV. Experimental Results

The main objective of this work is to identify proper transition from the given video. In this section, dataset used in the shot boundary detection scheme, performance measures and comparison of the proposed method with the existing methods are presented.

A. Dataset and Evaluation Criterion

To evaluate the performance, the proposed shot boundary detection scheme is tested with a variety of test videos, as listed in Table I. The first four videos are selected from the TRECVID 2001 test data and the remaining two videos are from the TRECVID 2007 dataset [19]. The TRECVID 2001 videos are downloaded from 'open-video project' [18]. All these video sequences have 690 shot transitions and various effects like object motion, camera motion, illumination and zooming. There are 1,72,961 frames and a duration of 1:47:29 seconds in total. To examine the performance of the proposed method, precision, recall and F1 score are used. The precision, recall and F1 score are calculated using:

$$Precision(P) = \frac{N_C}{N_C + N_F} \quad (8)$$

$$Recall(R) = \frac{N_C}{N_C + N_M} \quad (9)$$

$$F1\ Score = \frac{2 \times P \times R}{(P+R)} \quad (10)$$

Where N_C is the correct number of shot boundaries that are detected, and N_M is the number of missed shot boundaries and N_F is the number of false shot boundaries that are detected. The proposed method is developed in MATLAB (R2013a).

B. Parameter Selection

The performance of the proposed method is affected by four feature weights. The feature weights are allotted based on the significance of the extracted features. The range of w_1 is from 0.7 to 0.83. The range of w_2 is from 0.53 to 0.94. The range of w_3 is from 0.9 to 0.98. And the range of w_4 is from 0.3 to 0.8. Most of the videos give best result if weight values are in the ratio such that w_1 =0.83, w_2=0.83, w_3=0.93 and w_4=0.4.

TABLE I
DESCRIPTION OF TRECVID VIDEO DATASET

Video	File	Size (MB)	Run Time (mm:ss)	Frame Count	**Cuts**
NASA 25th Anniversary Show – Seg 5	Anni005	66.9	6:19	11364	38
NASA 25th Anniversary Show – Seg 9	Anni009	72.4	6:50	12307	38
A&S Reports Tape 5-Report 264	Nad 57	62.7	6.57	12781	44
The Great Web of Water	Bor08	251	28.07	50569	380
TRECVID 2007 dataset	BG_2408	240.7	25.27	35892	101
TRECVID 2007 dataset	BG_9401	335.6	33.53	50049	89

C. Comparison with Recent Techniques

The F1 measure of the proposed work is shown in Fig. 3. The work gives best result to the video files such that Anni005, nad57, and BG2408. From Fig. 3 it should be noted that the F1 measure of the proposed method is 0.95 for the anni005, nad57 and BG_2408 video files. To prove the accuracy of the presented method, it is also compared to the recent shot boundary detection methods like Kirsch directional derivatives [15] and modularity optimization method [14]. It is shown in Figs. 4 and 5. Figs. 4 and 5 depict that the average F1 measure for the presented work is high when compared with that of the other recent works.

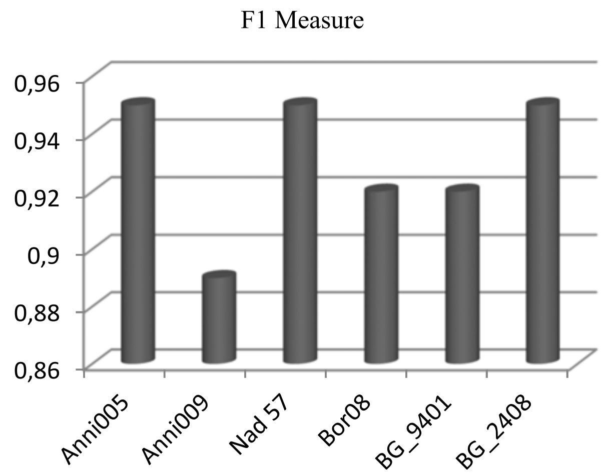

Fig. 3 F1 Measure of the proposed method

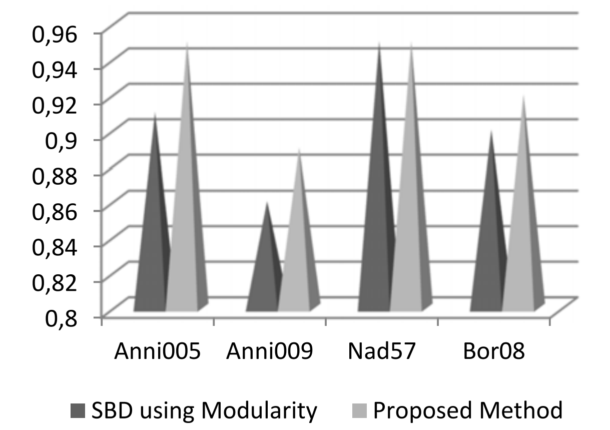

Fig. 4 Comparisons of F1Measure between the proposed method and SBD using Modularity method

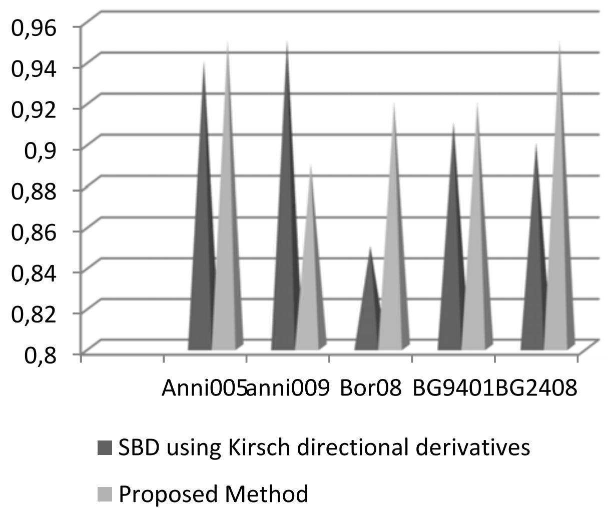

Fig. 5 Comparisons of F1 score between the proposed method and the SBD using Kirsch directional derivative.

V. Conclusion

A new shot boundary detection system using Octagon square search pattern has been designed and developed for the automatic shot detection in the given video. The Experimental results are tested on TRECVID datasets and validated using the performance measures such as precision, recall and F1 score. From the results the proposed shot boundary detection method achieves better F1 Score.

Acknowledgment

We would like to thank the National Institute of Standards and Technology (NIST) for providing TRECVID dataset.

References

[1] Sklar, Robert. *Film: An International History of the Medium.* (London): Thames and Hudson, (c. 1990). p. 526.

[2] Alan Hanjalic, “Shot Boundary detection: Unraveled and Resolved?”, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 12, No. 2, February 2002.

[3] H. Zhang and S. Kankanhalli, “Automatic partitioning of full-motion video,” ACM Journal of Multimedia Systems, vol. 1, pp. 10–28, Jan. 1993.

[4] R. Zabih, J. Miller, and K. Mai, “A feature-based algorithm for detecting cuts and classifying scene breaks,” in Proc. ACM Multimedia ’95, San Francisco, CA, pp. 189–200, 1995.

[5] B. Shahraray, “Scene change detection and content-based sampling of video sequences,” *Proc. SPIE*, vol. 2419, pp. 2–13, Apr. 1995.

[6] A. Hampapur, R. Jain, T. Weymouth, Production model based digital video segmentation, J. Multimedia Tools Appl. 1 (1) (1995) 9–46.

[7] U. Gargi, R. Kasturi, and S. H. Strayer, “Performance characterization of video-shot-change detection methods,” IEEE Trans. Circuits, Systems, Video Technology, vol. 10, pp. 1–13, Feb. 2000.

[8] R. Lienhart, “Reliable transition detection in videos: A survey and practitioner’s guide,” Int. Journal of Image and Graphics, vol. 1, pp. 469–486, Sept. 2001.

[9] Weigang Zhang, et. al., “Video Shot Detection Using Hidden Markov Models with Complementary Features,” In Pro-ceedings of the First International Conference on Innovative Computing, Information and Control. Vol.3. http://doi.ieeecomputersociety.org/10.1109/ICICIC.2006.549, 2006.

[10] Y. Kawai, H. Sumiyoshi, and N. Yagi. “Shot Boundary De-tection at TRECVID 2007,” In TRECVID 2007 Workshop, Gaithersburg, 2007.

[11] Don, A., Uma, K. Adaptive edge-oriented shot boundary detection. *EURASIP Journal on Image and Video Processing* 2009.

[12] Shiguo Lian, “Automatic video temporal segmentation based on multiple features,” *Soft Comput.*, vol. 15, no. 3, pp. 469–482, 2011.

[13] G. G. Lakshmi Priya , S. Domnic, “Transition Detection Using Hilbert Transform and Texture Features”, American Journal of Signal Processing, 2012, 2(2): 35-40.

[14] Choudhury, A., Medioni. G.: "A framework for Robust Online Video Contrast Enhancement Using Modularity Optimization," *Circuits and Systems for Video Technology*, IEEE Transactions on, 2012, (22), 9, pp. 1266 – 1279.

[15] B.H. Shekar and K.P. Uma, “Kirsch directional derivatives based shot boundary detection: an efficient and accurate method”, Procedia Computer Science, pp. 565-571, 2015.

[16] Sowmyayani. S., Arockia Jansi Rani, P., “Block based Motion Estimation using Octagon and Square Pattern”, Int. Journal of Signal Processing., Image Processing and Pattern Recognition, vol. 7, no. 4, pp.317-324, 2014.

[17] Kavitha. J, Sowmyayani. S., Arockia Jansi Rani. P., “Shot Boundary Detection Using DWT and Texture Features”, IEEE International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE’16).

[18] Open Video Project (OVP) (online) http://www.open-video.org/.

[19] TRECVID Dataset (online) http://trecvid.nist.gov/.