Journal of Engineering and Applied Sciences 12 (Special Issue 10): 8987-8993, 2017
ISSN: 1816-949X

Shot Boundary Detection Based on Dynamic Candidate Segment and Motion Estimation

Isra'a Hadi and Hikmat Z. Neima
College of IT, University of Babylon, Hillah, Iraq

Abstract: Nowadays, multimedia applications are widely extended and efficient methods for video analysis are required. Video Shot Boundary Detection (SBD) is an essential step for video analysis, indexing, retrieval and summarization. SBD is the process of finding the location of boundaries between video shots. Many SBD methods have been proposed, however, most of them focus on increasing the detecting accuracy which logically leads to expensive computation time. In this study, a fast and efficient SBD method is proposed. The proposed method has two stages, namely, shot boundary identification and verification. In the first stage, input video is reduced spatially and temporally. For the spatial reduction, Extend Salient Region (ESR) method is used to reduce the process only on the important part within each frame. While Dynamic Candidate Segment Selection (DCSS) is utilized in terms of temporal reduction. In DCSS, only the frame segments that highly probable to have shot boundary is passed for the verification stage. In the verification stage, Edge Count Ratio (ECR) is applied to obtain frame local feature in order to minimize the false detected boundaries. The proposed method showed good results in regard to detecting accuracy and computation time.

Key words: Shot boundary detection, candidate segment, extended salient region, edge detection, twin comparison, motion estimation

INTRODUCTION

The rapid evolution in digital video technology coupled with a substantial advance in computer performance has resulted in an explosion of digital video data. Many applications such as video-on-demand, distance learning surveillance, utilize video data. This has spurred for development of tools for efficient indexing, searching, browsing and retrieval (Hannane *et al*., 2016; Koprinska and Carrato, 2001).

Video structural analysis is a fundamental step for further video content analysis. A shot is a consecutive sequence of frames captured by a single camera. The shot is the basic unit in the video file. Video Shot Boundary Detection (SBD) is the most common process for video structural analysis. SBD is the process of temporally segmenting the video stream into its building blocks (shots) (Yuan *et al*., 2007).

Depending on a transition between shots, shot boundaries can be classified into two types, namely Cut Transition (CT) and Gradual Transition (GT). In CT, the transition occurs in case of an abrupt change in some specific features between two successive frames. In GT, on the other hand, the transition is gradual. GT can be further classified into dissolve, fade and wipe. In a dissolve, the last frames of the current shot are temporally overlapped with the first frames of the next shot. Fade, occurs as a smooth change in brightness of frame till it turns into blank frame (fade out) and the blank frame turns into the next shot (fade in). In wipe, the appearing and disappearing shots are interchanged in intermediate frames until the appearing shot totally replaces the disappearing one (Boreczky and Rowe, 1996; Cotsaces *et al*., 2006).

During the last decade, shot boundary detection has granted a considerable research attention and many methods have been presented to analyze this problem. The basic idea behind SBD is to find the discontinuity of visual content. Most of SBD methods extract one or more features from each frame and then compare between two successive frames in order to decide the presence of boundaries (Gao and Ma, 2014).

In this study, an efficient method for boundary detection is proposed. In the proposed method, accuracy is ensured by optimal use for feature extraction and adaptive threshold. Whereas the computation cost is decreased by skipping frames that probably have no boundary and process only informative frames. Following subsections present fundamental information that is related to the proposed shot boundary detection method.

Extended Salient Region (ESR): In terms of still image, a salient region is in high importance among image parts. The salient region is often used in the field of image

Corresponding Author: Isra'a Hadi, College of IT, University of Babylon, Hillah, Iraq

matching. In this context, the most important part in each image will be compared with the corresponding important part in another image in order to determine the ration of matching between two being processed images. However, to improve the accuracy of the matching process, taking only the salient region into account is insufficient. This is because that the background has an influence on the accuracy. So, extension the salient region by including the surrounding regions will ensure higher accuracy. Combination the most salient region with the surrounding regions is called Extended Salient Region (ESR) (Zhang *et al*., 2017).

Candidate segment selection: Theoretical and practical analyses show that the difference between any two successive frames are the largest in cut transition while these differences are the smallest between two successive frames belong to the same shot. Differences between two successive frames on the other hand are in the middle in case of gradual transition (Huo *et al*., 2016).

From the above observation, it is clearly noticeable that there are many frames have the smallest differences especially within the same shot. Hence, in order to speed up the process of shot boundary detection these frames can be avoided in the detection process. For the sake of excluding these non-boundary frames, only segments of the frame that might have boundaries will be further investigated to locate the shot boundaries. These segments are called candidate segments. The process of eliminating non-boundary frames ensures reduction of the computational complexity in the next steps of the shot boundary detection task (Lu and Shi, 2013).

Candidate segments are selected based on the fact that consecutive frames within the same shot have high correlations. Thus, if the first and the last frame of a segment exhibit high similarity this can be considered as an indication that the segment has no boundary inside and hence can be skipped in the further processes. On the other hand if the first and last frames of a segment exhibit low similarity, it can be concluded that these frames are located in two different shots and hence the segment is marked as candidate segment (Li *et al*., 2009).

Lu and Shi (2013) and Li *et al*. (2009), the video sequence is partitioned into several segments in segment length of 21 frames. After ward, for each segment, calculate the similarity between its first and last frames, so that, the segment will be declared as a candidate or discarded segment based on a comparison of the similarity against a certain threshold. If the similarity threshold, the current segment will be announced as a candidate otherwise, the segment is non-candidate one. The major issue in the process of selection of candidate segments is the number of frames in each segment. Again in both methods (Lu and Shi 2013; Li *et al*., 2009) each segment contains 21 frames to select the candidate segments based on the similarity of first and last frames in the segment being checked. Analysis of these two methods shows an issue regarding this fixed partition strategy. The issue is that both methods assume that each segment contains only one boundary.

Practical analysis of several videos indicates that some shots are composed of <10 frames and hence, segment of 21 frames might have more than a boundary. Therefore, the fixed partition strategy has a chance to miss a considerable number of shot boundaries.

In this study, a Dynamic Candidate Segments Selection (DCSS) method is utilized. In DCSS method, a video sequence is dynamically partitioned into segments. Different segments have a different number of frames. DCSS method modifies the length of the current segment based on the previous segment length and the similarity measure of the current frames. Similarity measure that is used in DCSS method is a correlation. The skipping interval is obtained by the following (Eq. 1-4):

$$d_j = \sum_{k=1}^{i-1} \frac{1}{i-1} \text{Corr}\,(D_{di-1}, D_k)\, d_k \tag{1}$$

$$D_k = \sum_{j=1}^{k} d_j \tag{2}$$

$$\text{CorrR}\ (i, i+d_j) = \frac{\min(C_{i.\,i+1}, C_{i.\,i+d_j})}{\max(C_{i.\,i+1}, C_{i.\,i+d_j})} \tag{3}$$

$$\text{Corr}(f, f+1) = \frac{\sum_{i=0}^{N-1}\sum_{j=0}^{M-1} (X_f(i, j)-\mu_f)\,(X_{f-1}(i, j)-\mu_{f-1})}{\delta_f\ \delta_{f+1}} \tag{4}$$

Where:

$C_{i.\,i+1}$ = The Correlation between Frames F_i and F_{i+1}

$C_{i.\,i+dj}$ = Means the Correlation between Frames F_i and F_{i+dj}

d_j = The current skipping interval

d_k = The previous skipping interval

D_k = The summation of skipping intervals values before the similarity measure (CorrR) is degraded below a threshold

Larger Correlation Ration (CorrR) indicates that the similarity between F_i and F_{i+i} is closed to similarity between F_i and F_{i+s} which means that frames F_i and F_{i+dj} are similar and consequently the interval started in F_i and ended in F_{i+dj} can be declared as non-candidate segment and hence it is avoided in the further processes. Small value of CorrR, on the other hand, announces occurrence of difference between F_i and $F_{i+dj}{}^{+1}$, F_{dj}, so that, the

segment with interval (F_i-F_{i+dj}) is considered as a candidate segment that has high probability to contain a shot boundary. Afterward, the candidate segment will be divided into two parts F_i, $F_{i+dj/2}$ and F_i, $F_{i+dj/2}$ and the CorrR is calculated one more time to discard a half of candidate segment interval. This process is called bisection partition which ensures more computation complexity reduction in the subsequent steps. The initial skipping interval is set to 40 based on the fact of the time of human visual reduction is about 1-2 sec. So, for a video sequence with 30 frame per sec FPS, it is reasonable to initially skip 40 frame (1.3 sec).

Motion estimation: Motion Estimation (ME) is the process by which estimate the motion difference between adjacent frames in a video sequence. ME plays a crucial role in video compression. ME process determines motion vector which represents the motion activity. In video coding motion vector is transferred coupled with the previous frame. In video decoding, on the other hand, the motion vector is used in the process of motion compensation. For the sake of estimating motion vector, two categories of motion estimation algorithm can be used. The first category research on the pixel level. This category is more accurate, however, it requires intensive computation time. The second category relies on blocks instead of pixel level. Block-based motion estimation methods which are usually called Block Matching (BM) are the most popular methods due to their simplicity and balancing between effectiveness and computation time (Diaz-Cortes *et al*., 2017). In block matching algorithms, the current Frame F_t is partitioned into equal-sized block each block of size N×N pixels. Each block in F_t has an area of search in the previous Frame F_{t-1}. The motion vector is found when the difference measurement metric is minimized.

There are several block matching algorithms have been presented. An exhaustive algorithm is the most accurate algorithm, however, it is expensive in terms of computation time. In order to reduce the computation time, several block matching algorithms have been developed such as Simple and an Efficient Search (SES), Diamond Search (DS), Three Step Search (TSS), Four Step Search (4SS) and other search algorithms (Barjatya, 2004). Three Step Search (TSS) is found to be efficient algorithm in regard to matching performance and computation time as well (Santamaria and Trujillo, 2012). In this study, TSS algorithm is utilized for motion vector estimation.

Literature review: Gao and Ma (2014) claim that the most information is found in the center of frame and in the pixels around the frame center. The proposed method speeded up the process of SBD in two directions, namely temporally and spatially. In terms of temporally redundancy several frames are adaptively skipped at each time. Whereas spatial redundancy is performed by concentrating on the information pixels which are found in the center of frame. In the proposed method, each frame is divided into non-overlapping sub-regions.

Sub-regions in the frame are defined as non-focus region which are the most external sub-regions, focus regions are the sub-regions around the center of frame. While the sub-regions located in between focus regions and non-focus regions are defined as second focus region. For the sake of finding the similarity matrix, only the focus region and down-sampled second region are included while the non-focus region is discarded.

An approach for precise detecting of shot boundary has been presented by Gao and Ma. The proposed approach is composed of four modules namely Frame-Skipping Module (FSM), Change Detection Module (CDM), Abrupt Boundary Module (ABM) and Gradual Boundary Module (GBM). FSM outputs video sequence differs from the input video by skipping several frames to provide less number of frames for further processing. In the second module, CDM classifies the provided frames from FSM into C_{pass}, $C_{gradual}$ or C_{abrupt}. C_{pass} means the frame will be ignored for further processing as it has no possibility to be a shot boundary. C_{abrupt} and $C_{gradual}$ indicate the frame is probably a cut or gradual transition and they require further processing. In order for that to be done, global histogram is utilized in case of cut transition while block-based histogram is found for gradual transition.

The difference either in global histogram or in block-based histogram is compared with two threshold values T_{high} and T_{low}. If the difference $>T_{low}$, these two successive frame are declared as C_{pass}. If the difference less than T_{high} and $<T_{low}$, these frames are signed as $C_{gradual}$. While if the difference $<T_{high}$, there is possibility to be abrupt boundary, so, this frame is marked as $C_{gradual}$.

Hannane *et al*. (2016), the proposed method has taken to the consideration challenges such as object motion, camera rotation which cause false shot detection. Researcher extract SIFT keypoints and edge SIFT keypoints from the video frame. Edge SIFT keypoints of objects in the frame are robust against illumination and scale variance. In addition to that, edge SIFT keypoints are of significant importance in the video information. Histogram of SIFT-point distribution is found and this will be used to measure the differences between two successive frames.

Lu and Shi (2013), a fast technique for shot boundary detection has been proposed. The proposed technique

has two main steps. The first step is applying Singular Value Decomposition (SVD) on matrix which is composed of frame color histogram for R, G and B color components. SVD derives a low dimensional feature space from high dimensional feature space. The second step is candidate segment selection in which non-boundary frames are discarded and the remaining frames have possibility to be as boundaries. Cosine distance is utilized to find the similarity between two successive frames.

Candidate segment selection and transition pattern analysis were used to detect shot boundaries by Tippaya *et al*. (2015). As a global feature in this approach, color histogram has been employed. Speed Up Robust Feature (SURF) was utilized as a local descriptor in combination with color histogram. Pearson Correlation Coefficient (PCC) was applied as a distance measurement between histograms of two successive frames. In the next step of the proposed method, apply local maxima detection and area under curve calculations to analyze candidate segment in order to analyze the transition pattern.

MATERIALS AND METHODS

Proposed method: The following block diagram shows the proposed method Fig. 1. As it is clear from the block diagram that the proposed method has three stages, namely, preprocessing, SB identification and SB verification.

Preprocessing: In order to focus on the most important contents in each frame of the video sequence, Extended Salient Region (ESR) is utilized to extract only the frame contents that are located in the center of the frame. This stage ensures minimizing of a frame data to be further processed rather than taking the whole frame. The salient region is the region in the image (frame) that located in the center of that image. This observation is theoretically logical due to the cameraman ordinary tries to focus on the important area that is needed to be captured and thus cameraman definitely locates the important area in the center of his camera's lens. Extended salient region has been proposed to reduce the chance of missing some of important content. This is done by including the objects in the center and a part of surrounding background. Consequently, a considerable number of the image (frame) pixel will be excluded in the next step of the process. Logically, minimizing the number of frame pixels that are required to be further processed leads to speed up the whole process.

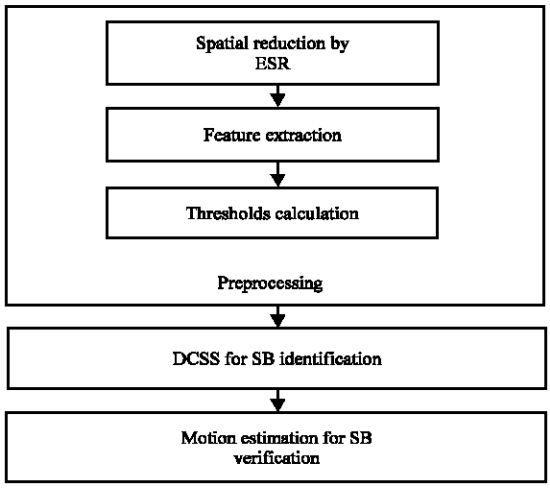

Fig. 1: Block diagram of the proposed method

Feature extraction: Image feature extraction is the process of detecting certain features of interest within an image and then represent these features for further processing. For the video sequence, extracted features depict the visual aspects of the video content. In this study, two features are extracted from each frame in the video sequence. The first feature is global while the second feature is local. The global is the color histogram of the frame. The color histogram is a representation of image colors distribution. The local feature extracted in the proposed method is edge of each frame. From each two adjacent frames, edge count ratio is calculated using Eq. 5.

For frames n and n-1, entering and exiting edge pixels are defined as follows. Entering edge pixels (X^{in}_n) are the number of edge pixels in frame n which have a distance, from the closest edge pixel in frame n-1 is larger than a certain predefined distance. Whereas exiting edge pixels (X^{out}_n) are the fraction of edge pixels in frame n-1 that have a distance more than a certain distance away from the closest edge pixel in frame n (Lew, 2001):

$$ECR = \max (X^{in}_n \mid \delta_n, X^{out}_{n-1} \mid \delta_{n-1}) \quad (5)$$

Where:

X^{in}_n = Entering edge pixels in frame n
X^{out}_n = Exciting edge pixels in frame n-1
δ_n = The number of edge pixels in frame n
δ_{n-1} = The number of edge pixels in frame n-1

Adaptive threshold calculation: A good threshold value selection plays a crucial role in SBD algorithm performance. Most of the video SBD methods try to use

a suitable similarity threshold value to accurately determine the boundaries in the video sequence (Ko *et al*., 2006). Fixed threshold either be longer than the difference between successive frames which leads in missing of considerable number of boundaries or be very small and hence, it is sensitive to small difference even the differences resulted from noise, illumination change or even small motion of objects within the same frame.

Consequently, adaptive threshold is suitable to overcome this issue into some context. Most of SBD methods that use an adaptive threshold are dependent on the two well-known statistical metrics, namely mean and standard deviation. However, as video sequence normally is composed of thresholds of frames and hundreds of shots which are vary in contents, threshold produced based on global mean and standard deviation will be less effective than the threshold produced based on local mean and standard deviation.

Li *et al*. (2009) and Fanan and Khobragade (2016), video sequence is partitioned into groups each of 200 frames and local threshold is calculated based on global and local mean in addition to local standard deviation. Local threshold is calculated using Eq. 6:

$$T_L = 1.1\mu_L + 0.6\left(\frac{\mu G}{\mu L}\right)\delta_L \quad (6)$$

Lu and Shi (2013), the local threshold is found for each 200-frame group using the following Eq. 7:

$$T_L = \mu_L + a\left(1+\ln\left(\frac{\mu_G}{\mu_L}\right)\right)\delta_L \quad (7)$$

Where:

μ_L = Lcal mean
μ_G = Global mean
δ_L = Local standard deviation
a = A small constant

This equation is claimed to be better than the threshold calculated using previous Eq. 1 that is justified as it is smaller and misses less of true candidate segments.

In this study, two threshold values were used. The first one is called large threshold Th_l and the second threshold value represents small threshold Th_s. These two thresholds are given by:

$$T_b = a\times\mu_L + b\times\log\left(\frac{\mu_G}{\mu_L}\right)\times\delta_L \quad (8)$$

$$T_s = c\times\mu_L + d\times\log\left(\frac{\mu_G}{\mu_L}\right)\times\delta_L \quad (9)$$

where, a = 1.1, b = 0.25 c = 0.8 and d = 0.001.

DCSS: In the previous study, ESR is used for spatial reduction. In DCSS, the video sequence is temporally reduced by discard the frame segments that probably have no boundary while preserving the segments that are potentially have shot boundaries within them. Instead of fixed segment interval, DCSS checks segments with different intervals based on the ratio of a difference between the frame being processed and its adjacent frame and the difference between the current frame and the frame positioned at the end of the interval. Current interval calculation relies on that ratio and the previous interval. A number of frames in intervals is between 5 and 40. In order to prevent the successive intervals to be >5 frames, we check the number of frames in each interval occasionally. If more than two intervals are found to have small frames, the next interval is forced to be longer. This step prevents intervals from being stuck in a small number of frames.

When the interval is determined, a binary search is applied on two levels. The interval is divided into two parts and the difference is calculated for both parts. The part that has a larger difference will be passed to the next level of binary search whereas another part of the further process. In the second part of the binary search, the same process is repeated in order to minimize the number of frames in each segment of frames to be deeply investigated in the verification stage.

Shot boundary verification: After the use of DCSS, candidate frame segments that potentially contain shot boundaries are passed to verification stage. In verification stage, each segment of candidate segments is deeply investigated in order to the verify existence of shot boundary. This stage is necessary, since, the result of the previous stage may wrongly determine a segment as a candidate segment. The false positive of the previous stage is eliminated in the verification stage. For the sake of false positive alleviation, each candidate segment is entirely checked against two threshold values, T_b and T_s. If the difference between two adjacent frames exceeds T_b, the current position will be declared as a cut transition. For gradual transition detection a twin threshold approach is employed.

The difference is checked against two thresholds T_b and T_s. If the difference is found to be less then T_b longer than and T_s, a potential gradual transition is announced. Then continue in checking while the

difference exceeds T_b. Motion in video sequence resulted from object or camera motion causes the difference between two adjacent frames to be high, although, they are within the same shot. This observation needs to be dealt with. For that purpose, in the proposed method, the motion activity is calculated via calculating of the motion vector. If the motion activity is found to be high, T_b, T_s are raised in a certain proportion in order to avoid the motion from being taken as a real difference and sub sequentially as a shot boundary. Three step search is efficient block matching algorithm used for motion estimation. After getting the motion vector, it is simple to measure the motion activity using a standard deviation of the motion vector. If the standard deviation of motion vector is found to be larger than 2.9, this is a cue of active motion. Thus, we raise both of T_b and T_s.

RESULTS AND DISCUSSION

In order to evaluate the proposed shot boundary detection method, we have tested the method on six of standard videos. The videos are downloaded from a free video data website (open-video-project) while the description of each one of these videos is obtained from the TRECVID 2001 dataset. The videos used for the testing stage are described in Table 1. Videos were selected to have a variety of transition type cut and gradual (dissolve, FIO). Two basic metrics are usually used to measure the performance of shot boundary detection algorithms. These are precision and recall. In general, Precision is defined as the proportion of relevant returned information by the system. Recall, on the other hand is defined as the proportion of all the relevant information are returned by the system (Xu *et al*., 2016):

$$\text{Recall} = \frac{N_C}{N_C+N_F} \quad (10)$$

$$\text{Precision} = \frac{N_C}{N_C+N_M} \quad (11)$$

Where:

N_C = The number of shot boundaries that are correctly detected

N_M = The number of shot boundaries that are missed in detection process

N_F = The number of shot boundaries that are falsely detected

Table 2 shows the results of recall and precision obtained from implementing the proposed method on the described dataset. Noticing the results in Table 2, for cut transition, recall and precision values are very good as well as the recall in case of gradual transition. Results of recall of V5 and 6 in terms are excellent, however, their precision were degraded because of the few number of gradual transition in both video sequences.

Table 1: Dataset video description

		Transition		
Video file	Frames	Cut	Gradual	Total
V1	11362	38	27	65
V2	12306	37	65	102
V3	48450	231	11	242
V4	12510	45	19	64
V5	7401	13	1	14
V6	7419	12	2	14

Table 2: Results of proposed method

	Cut		Gradual	
Video file	Recall	Precision	Recall	Precision
V1	0.89	0.87	0.81	0.73
V2	0.94	0.85	0.89	0.70
V3	0.98	0.95	0.90	0.66
V4	0.93	0.84	0.78	0.71
V5	1.00	0.86	1.00	0.50
V6	0.91	0.84	1.00	0.66

CONCLUSION

Video shot boundary detection is a basic step towards video analysis. In this study, an efficient video shot method has been proposed. We can conclude that the use of global feature coupled with local feature results in better accuracy compared with use of only global feature. Video sequences, normally have motion within. This motion is either caused by objects motion or camera motion or even by both of them sometimes. Occurrence of this motion may be considered as a difference and consequently reflects an occurrence of shot boundary. Therefore, an attention should be paid for this aspect, so that, the accuracy will be better.

REFERENCES

Barjatya, A., 2004. Block matching algorithms for motion estimation. IEEE Trans. Evol. Comput., 8: 225-239.

Boreczky, J.S. and L.A. Rowe, 1996. Comparison of video shot boundary detection techniques. J. Electron. Imaging, 5: 122-128.

Cotsaces, C., N. Nikolaidis and I. Pitas, 2006. Video shot detection and condensed representation: A review. IEEE. Signal Process. Mag., 23: 28-37.

Diaz-Cortes, M.A., E. Cuevas and R. Rojas, 2017. Motion Estimation Algorithm using Block-Matching and Harmony Search Optimization. In: Engineering Applications of Soft Computing, Diaz-Cortes, M.A., E. Cuevas and R. Rojas (Eds.). Springer, Switerzland, ISBN:978-3-319-57812-5, pp: 13-44.

Fanan, N.G. and A.S. Khobragade, 2016. A fast robust technique for video shot boundary detection. Proceedings of the 2016 Online International Conference on Green Engineering and Technologies (IC-GET'16), November 19, 2016, IEEE, Coimbatore, India, ISBN:978-1-5090-4557-0, pp: 1-6.

Gao, G. and H. Ma, 2014. To accelerate shot boundary detection by reducing detection region and scope. Multimedia Tools Appl., 71: 1749-1770.

Hannane, R., A. Elboushaki, K. Afdel, P. Naghabhushan and M. Javed, 2016. An efficient method for video shot boundary detection and keyframe extraction using SIFT-point distribution histogram. Intl. J. Multimedia Inf. Retrieval, 5: 89-104.

Huo, Y., Y. Wang and H. Hu, 2016. Effective algorithms for video shot and scene boundaries detection. Proceedings of the 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS'16), June 26-29, 2016, IEEE, Okayama, Japan, ISBN:978-1-5090-0807-0, pp: 1-6.

Ko, K.C., Y.M. Cheon, G.Y. Kim, H.I. Choi and S.Y. Shin *et al*., 2006. Video shot boundary detection algorithm. In: Computer Vision, Graphics and Image Processing, Kalra, P.K. and S. Peleg (Eds.). Springer, Berlin, Germany, ISBN:978-3-540-68301-8, pp: 388-396.

Koprinska, I. and S. Carrato, 2001. Temporal video segmentation: A survey. Signal Process. Image Commun., 16: 477-500.

Lew, M.S., 2001. Principles of Visual Information Retrieval. Springer, London, UK., ISBN:1-85233-381-2, Pages: 359.

Li, Y.N., Z.M. Lu and X.M. Niu, 2009. Fast video shot boundary detection framework employing pre-processing techniques. IET Image Process., 3: 121-134.

Lu, Z.M. and Y. Shi, 2013. Fast video shot boundary detection based on SVD and pattern matching. IEEE Trans. Image Process., 22: 5136-5145.

Santamaria, M. and M. Trujillo, 2012. A comparison of block-matching motion estimation algorithms. Proceedings of the 7th Colombian Congress on Computing (CCC), October 1-5, 2012, IEEE, Medellin, Colombia, ISBN:978-1-4673-1475-6, pp: 1-6.

Tippaya, S., S. Sitjongsataporn, T. Tan, K. Chamnongthai and M. Khan, 2015. Video shot boundary detection based on candidate segment selection and transition pattern analysis. Proceedings of the 2015 IEEE International Conference on Digital Signal Processing (DSP), July 21-24, 2015, IEEE, Singapore, Asia, ISBN:978-1-4799-8059-8, pp: 1025-1029.

Xu, J., L. Song and R. Xie, 2016. Shot boundary detection using convolutional neural networks. Proceedings of the Visual Communications and Image Processing (VCIP'16), November 27-30, 2016, IEEE, Chengdu, China, ISBN:978-1-5090-5317-9, pp: 1-4.

Yuan, J., H. Wang, L. Xiao, W. Zheng and J. Li *et al*., 2007. A formal study of shot boundary detection. IEEE. Trans. Circuits Syst. Video Technol., 17: 168-186.

Zhang, J., S. Feng, D. Li, Y. Gao and Z. Chen *et al*., 2017. Image retrieval using the extended salient region. Inf. Sci., 399: 154-182.