Journal of Engineering and Applied Sciences 12 (21): 5388-5393, 2017
ISSN: 1816-949X

Video Shot Boundary Detection: A Comprehensive Review

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

Abstract: Video shot is the basic component of video file. Shot Boundary Detection (SBD) is the process of segmenting video sequence into its shots. SBD is a very important step for further analysis in video processing. Various methods for detecting video shot boundaries have been proposed. Although, the importance of the step of boundary detection, few studies on shot boundary detection have been published, these studies have focused on limited range of SBD methods. In this study, an extended study of SBD is presented. This study classifies shot boundary detection methods based on the level they work on and most recent methods are investigated as well.

Key words: Boundry, detection, shot boundary, component, methods, limited

INTRODUCTION

The rapid evolution in digital video technology coupled with substantial advance in computer performance have 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. Shot boundary detection is the most common process for video structural analysis. Shot is a consecutive sequence of frames captured from a single camera. Shot boundary detection is the process of temporally segmenting the video stream into its building blocks (shots) (Yuan *et al*., 2007). Depending on transition between shots, shot boundaries can be classified into two types, namely Cut Transition (CT) and Gadual Transition (GT). In CT, the transition occurs in case of 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 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 (Cotsaces *et al*., 2006; Boreczky and Rowe, 1996).

Methods for shot boundary detection have been discussed in several review papers. Koprinska and Carrato (2001) researchers have presented a survey on existing temporal video segmentation that work on compressed and uncompressed video stream.

Yuan *et al*. (2007), a formal study of SBD is presented. Researchers proposed a general framework of SBD and they discussed several challenges of that framework. State-of-art methods of SBD are reviewed by Cotsaces *et al*. (2006). Researchers have discussed several of relatively recent methods. Boreczky and Rowe (1996) a comparison of SBD techniques is presented compared to older studies on SBD, this comparison has taken a relatively wide range of test material.

Despite there are several reviews of SBD are published in literatures, these reviews have not taken into account recently published SBD methods. The objective of this study is to provide a classification of SBD methods including the most recent ones.

Shot boundary detection: As prementioned in section 1, video shot boundary detection, sometimes called temporal video segmentation is a fundamental step in structural video analysis. SBD is a very important step for further video stream analysis.

Normally, the difference or dissimilarity between two successive frames is measured in order to detect shot boundary. Despite there are other categories, based on the level that is used to detect the difference, SBD methods can be mainly classified into three categories, namely, pixel, block and histogram comparisons.

Pixel comparison: The measurement of pixel difference is considered as the simplest way. The difference of two successive frames is detected if number of pixels that change in value is more than a predefined threshold. Gray

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

camera movement highly effects the result of this method (Boreczky and Rowe, 1996). Equation 1 was used to find the absolute sum of difference in pixel values is:

$$D(i,i+1) = \frac{\sum_{x=1}^{x}\sum_{y=1}^{y}\sum_{c|P_i(x,y,c)|} - P_{i+1(x,y,c)}}{XY} \quad (1)$$

Where:

i and i+1 = Two successive frames

P_i (x, y, c) = The color component (c) of the pixel (x, y) in frame i

Despite of simplicity of the methods of pixel comparison, they suffer from the lack of distinguishing large change in a small area from small change in a large area (Koprinska and Carrato, 2001). Srilakshmi and Sandeep (2015), Structural Similarity (SSIM) Index has been utilized to detect shot boundary. Structural similarity is calculated using Eq. 2:

$$SSIM(x,y) = \frac{(2\mu_x+\mu_y)(2\sigma_{xy})}{(2\mu_x^2+\mu_y^2)(\mu_x^2+\mu_y^2)} \quad (2)$$

Where:

x, y = Two successive frames

σ_{XY} = Co-variance between frame x and y

Researchers considered μ as an estimation of the frame luminance while standard deviation as an estimation of contrast in the frame. The dissimilarity index is obtained using Eq. 3:

$$DSSIM(X,Y) = \frac{1}{1\text{-}SSIM} \quad (3)$$

The proposed method is applicable to detect both cut transition and gradual transition. For cut transition, video is segmented into frames. Then and for each frame, mean, variance and covariance are calculated as next step, SSIM and DSSIM are calculated between successive frames. Plotting the dissimilarity graph measure versus the frame number in the forth step. Finally, shot boundaries are the sudden transitions in the plot.

For gradual transition, same fashion is used except no need to calculate mean, covariance, SSIM, DSSIM. Instead, calculate standard deviation of each frame in video. Afterward, plot the graph of standard deviation versus the frame number. The gradual decrease intensity is fade-out while the gradual increase will represent fade-in.

Madhusudhan and Hegde (2015) statistical methods like mean and standard deviation have been utilized to detect video shot boundaries. The proposed method segments video file into frames as a first step. In the second step, a preprocessing such as convert color to complexity of computation. Frame difference between each pixel in F_i and corresponding pixel in F_{i+1} is given as follow:

$$F_K = F_i - F_{i+1} \quad (4)$$

As the frame difference in any video sequence follows Gaussian distribution, researchers utilized statistical features like mean and standard deviation. These statistical features give useful information regarding video segments change. Gaussian distribution is obtained when frames belong to the same video sequence. On the other hand when frames belong to different video sequence which means presenting of shot boundary, the distribution will be random.

After calculation of mean and standard deviation for each frame, a threshold will be calculated as T = 2* SD+Mean. The next step in the proposed method is to find the local maxima that are larger than the threshold. Lastly, the frame number is detected to find the local maxima and mark these frames as boundaries of shots.

Sun and Wan (2014) a novel metric for SBD has been proposed. In the proposed approach, researchers presented a new view to capture the similarity between two adjacent frames without feature extraction. Their view depends on the value of pixels in adjacent frames. If the first frame is not a boundary, the value of the pixel in the first frame will be remained at the same place in the next frame or it may slightly deviate in space in the next frame. Thus, searching in a small region for a pixel with a value closed to the pixel value in the first frame it will be most probably to find it. However, if the first frame represents a boundary frame, then image content in the next frame will be considerably different from the image content in the first frame. Therefore, for most pixels in first frame, there are no similar pixel values in the next frame.

Based on the prementioned view, researchers have proposed a method for shot boundary detection in both cases of Cut Transition (CT) and Gradual Transition (GT). The proposed method dealt with special aspects such as change in illumination and camera movement.

Block comparison: A new algorithm for gradual shot detection has been proposed by Yoo *et al.* (2006). The proposed algorithm is based on the fact that says: most gradual curves is characterized by the distributions of variance of information of edge in the sequences of frames. Whereas, parabolic shapes will be appeared in case of dissolve.

In the first step of the proposed algorithm, average edge of sequence of frames is obtained by applying sobel edge detector. The average edge frames are divided into

nine sub-block and the variance of each sub-block will be then compared with the variance of full frame. Then extract sequence of feature that showing parabolic variance curve. In order to obtain smoothing curve, opening operation, a morphological operation will be applied on the extracted feature sequence. Finally, the local minimum of sliding window is then selected as a point for gradual detection.

Hanjalic (2002) an algorithm for both cut and gradual boundary detection has been proposed. The proposed algorithm is based on the concepts of probability and statistic. Researchers utilized motion compensation features in order to compute the discontinuity values.

In this research, each frame in video is divided into number of nonoverlapping blocks. Afterwards, searching for corresponding blocks in adjacent frame. Matching between two frames will be done based on block-matching criterion. Which is the sum of absolute differences of block-wise average values of YUV color components. The maximum displacement was selected to be 4 pixels for the purpose of experiments. The priori probability takes into consideration only information of shot length. While the probability density function takes into account the range that the discontinuity value falls on. Researchers developed likelihood function by analysis the relationship between likelihood and probability function. This likelihood function ensures minimizing the average of error probability which in turn leads to maximizing the quality of detector for cut transitions detection, adjacent frames are compared based on their blocks whereas for detecting gradual transitions instead of comparing adjacent frames, select two frames that have minimum shot length to be compared. For gradual transition which is stretched along some frames, the detection was done by comparing frames from the beginning and from the end of boundary.

Histogram comparison: The comparison of histogram is widely utilized to detect boundaries compared to block and pixel comparisons. Color computed from two adjacent frames is usually employed to detect shot boundary.

A new method for shot boundary detection in both cases, i.e., cut boundary and gradual boundary has been proposed by Li *et al.* (2016). The idea in the proposed method is to deal with both gradual boundary and cut boundary in the same fashion. This is achieved by converting the gradual boundary to cut boundary. The proposed method has three-stage process. In the first stage, color histogram is used in order to HVS color space is used among color spaces according to its closeness to human feeling. The normalized HVS color histogram is extracted from each frame as its feature. Euclidean distance is used here to measure the distance between color histogram of two successive frames. In case of cut boundary, euclidean distance usually is a local maximum while in case of gradual boundary it is slightly higher than the value of the mean. In order to, detect the gradual boundary, fine-level color histogram is used in the second stage. This was decided to, accumulate the small change in difference in color histogram between two successive frames. Researchers claimed that five-level is appropriate to detect the gradual change and it is not suitable to do more than five levels. This is because that the shot with five frames will be lost. And the meaningful shot has five frames at least. In first and second stages, candidate cut change and gradual change are detected, respectively. In the third stage, a voting mechanism is used to make the final decision regarding the detected candidate changes whether they are exist or not.

Mahmoud *et al.* (2013), an approach for generating video summary has been proposed. As shot boundary detection is the first step in such process, researchers utilized a shot boundary detection method which is based on color histogram. HVS color histogram was used to get color histogram to extract color feature.

HSV (Hue, Saturation, Value) is a color space that provides an intuitive color representation and it is closed to perceiving of color by human. Bins are used in this approach are 32 bins of H, 4 bins of S and 2 bins of V.

In the proposed approach, the Bhattacharyya distance is employed to calculate the distance between histograms in two successive frames:

$$\text{Bhattacharyya distance} = \sum_{i=0}^{n} \sqrt{\sum P_i \cdot \sum Q_i} \quad (5)$$

The shot is determined by comparing the distance obtained from Eq. 5 with a preset threshold. The threshold used here is 0.97 which is established through experiments. If the distance of two consecutive frames is less than the prespecified threshold (0.97), then a shot boundary is detected.

The advantages that make researchers to use Bhattacharyya distance are:

- It is a self-consistent and this ensures that the minimum distance between any two points is the straight line between them
- It is independent from the width of bins
- It is independent from how bins are divided
- It is not affected by the data distribution of data across histogram

Candidate segment selection and transition pattern analysis were used to detect shot boundaries by

Tippaya *et al.* (2015). In this approach, a combination of local and global features was implemented for video temporal representation.

As a global feature in this approach, color histogram has been employed due to providing a good tradeoff between complexity and accuracy. In the proposed approach, three types of color histogram were used, namely RGB, normalized intensity histogram and normalized RGB histogram.

Pearson Correlation Coefficient (PCC) was applied as a distance measurement between histograms of two successive frames. High PCC means high correlation between these two frames which indicates that two frames belong to the same shot while low value of PCC indicates that these frames may belong to different shots. The dissimilarity used here is as inverse of PCC. Since, histograms of two frames are similar sometimes, the dissimilarity value will be very low and consequently this will result in missed shot boundaries. To overcome this limitation, researchers have combined local features due to their ability to represent video frames. Speed Up Robust Feature (SURF) was used as a local descriptor in combination with color histogram in order to enhance the performance of boundary detection process. Video sequence is divided into using a predefined threshold. Afterwards, for each segment, mean and standard deviation values are calculated. These values will then be used to find adaptive threshold. Depending on the result of the comparison between the threshold and dissimilarity result, the dissimilarity segment will be represented as a candidate that may contain boundary of video shot.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. Keypoint feature matching combined with color histogram have been utilized shot boundaries by Lee and Kolsch (2015). Firstly, for each frame in video sequence, proposed method extracts descriptive features and similarity between two consecutive frames is calculated. Secondly, build groups of frames and similarity of two groups is obtained. Descriptive features from each frame are:keypoint features which are descriptors of frames based on their appearance. Keypoint used in proposed method is designed to be robust to image transformation such as brightness, rotation and scaling. Color block histograms which is obtained in RGB color space. Each frame is divided into some blocks and color histogram for each block is calculated in separate. Similarity between two frames depends on the number of descriptors that are matched. Researchers have considered two features as matched in case of their distance to the nearest distance ratio is larger than a preset threshold. In order to detect gradual transition, groups of frames will be taken to the account to calculate the similarity. In this case, a graph theory was employed. The similarity used here is min-max cut algorithm as it can find continuity of intra-group and discontinuity of inter-group. Number of frame in each group was determined based on Fibonacci sequence and this obviously reduces the time of computation.

MATERIALS AND METHODS

Specified SBD methods: In addition, to a forementioned classes of comparisons, i.e., pixel, block and histogram comparisons, there are methods for shot boundary detection differ in way they are based on. This section presents several of these methods.

Damnjanovic *et al.* (2007), spectral clustering has been used to detect shot boundary. The proposed method was based on the assumption that the shot boundary is considered as a global feature rather than local feature. General information about shot is gathered from the information about each frame during the time. Using the same fashion information about shot boundary is extracted through two successive shots interaction.

For the purpose of similarity matrix calculation, a sliding window of 100 frames was used rather than taking the whole video frames. This will definitely reduce the number of operations. The structure of video inside sliding window was described using three eigenvalues and each one of these eigenvalues gives an indication of possible choice of boundary of shot. Candidate shot boundary is found based a threshold which is calculated depending on normalized cut value.

As a last step in the proposed method, mean and standard deviation for normalized cut of adjacent frames will be calculated. The objective function is then compared with a threshold to decide whether shot boundary is presented or not presented in the current window.

A method for shot boundary detection based fuzzy correlation measurement between two consecutive frames has been proposed by Chakraborty *et al.* (2015). This method was only applied to detect cut transition. The proposed method is divided into three stages. In the first stage a preprocessing stage, video frames are extracted and then converted from RGB color space into gray scale image in order to reduce the time complexity. As it is well known that the fuzzy membership function is a fundamental step in any fuzzy-based process. Therefore, the value of fuzzy membership was found in the second stage. For the sake of fuzzy membership estimation, researchers assumed that each frame as a fuzzy set and pixels in that frame are considered as members of that fuzzy set. The fuzzy membership value of each pixel is calculated based on its hue value which in turn is based on pixel's intensity value. Equation 6 was used to calculate fuzzy membership value:

$$h(p)=\frac{gray(p)}{Max(I)-Min(I)}\ Max(I)\neq Min(I) \quad (6)$$

Where:

h(p) = Hue value of the pixel p
gray(p) = Gray scale intensity of the pixel p
Max(I) = Maximum Intensity value in image (Frame) I
Min(I) = Minimum Intensity value in image (Frame) I

In the last stage, the fuzzy correlation between successive frames was calculated. The frame is observed to be either a boundary or non-boundary based on comparison of fuzzy correlation value and a preset threshold. If the fuzzy correlation value was less than the threshold, then the frame is declared as a shot boundary.

Convolutional Neural Network (NCC) based framework for shot boundary detection has been presented by Xu *et al.* (2016). The proposed framework is divided into three stages.

In the first stage, candidate segments selection is done. Candidate segments are defined as segments that probably contain boundaries. This step is involved to avoid manipulation the whole video sequence which in consequent reduces the time complexity. Correlation measurement was employed here to select candidate segments. In this stage, segment is considered to have 21 frames each. The segment is decided to be candidate if the correlation value is less than a threshold.

Convolutional neural network was utilized in the second stage to extract features to be used in calculation of similarity measurement. Researchers justified use of CNN to its ability to filter the noise presented in background and to effectively represent information.

Finally and as the third stage for each candidate segment, feature vectors are extracted and the similarity will be tested based on the vectors among between segments. The threshold used in this study is 0.9.

In conclusion, the proposed framework has good performance in terms of evaluation metrics compared to state-of-art methods. The method is applicable on both types of transition, cut and gradual.

RESULTS AND DISCUSSION

Performance evaluation: In order to evaluate shot boundary, twobasic metrics are usually used. 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 informationare returned by the system Xu *et al.* (2016):

$$Recall=\frac{N_C}{N_C+N_M} \qquad Precision=\frac{N_C}{N_C+N_F} \quad (7)$$

Where:

N_C = Number of shot boundaries that are correctly detected
N_M = Number of shot boundaries that are missed in detection process
N_F = Number of shot boundaries that are falsely detected

CONCLUSION

Shot boundary detection is a very critical step video processing. Its importance comes from the sensitivity of segmenting video sequence into its basic units (video shot) for further video analysis. In this study, a review of shot boundary detection methods is presented. This paper classifies state-of-art methods into three categories based on the level they work on. Pixel, block and histogram comparison are the main three categories of these methods. It is noticed that use of one of state-of-art methods has relatively poor performance in terms of recall and precision while when a combination of more than one method is employed, the performance was better. Distant from the state-of-art method, several specific methods were described in this review. Different concepts such as soft computing and pattern recognition have been utilized in the proposed methods. The performance of these methods is obviously better than state-of-art methods.

REFERENCES

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

Chakraborty, B., S. Bhattacharyya and S. Chakraborty, 2015. An unsupervised approach to video shot boundary detection using fuzzy membership correlation measure. Proceedings of the 5th International Conference on Communication Systems and Network Technologies (CSNT15), April 4-6, 2015, IEEE, Gwalior india, ISBN:978-1-4799-1798-3, pp: 1136-1141.

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

Damnjanovic, U., E. Izquierdo and M. Grzegorzek, 2007. Shot boundary detection using spectral clustering. Proceedings of the 15th European Conference on Signal Processing 2007, September 3-7, 2007, IEEE, Poznan, Poland, ISBN:978-839-2134-04-6, pp: 1779-1783.

Hanjalic, A., 2002. Shot-boundary detection: Unraveled and resolved?. IEEE. Trans. Circuits Syst. Video Technol., 12: 90-105.

Hannane, R., A. Elboushaki and K. Afdel, 2016. Efficient video summarization based on motion sift-distribution histogram. Proceedings of the 13th International Conference on Computer Graphics Imaging and Visualization (CGiV) 2016, March 29-April 1, 2016, IEEE, Agadir, Morocco, ISBN:978-1-5090-0811-7, pp: 312-317.

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

Lee, K. and M. Kolsch, 2015. Shot boundary detection with graph theory using keypoint features and color histograms. Proceedings of the 2015 IEEE Winter Conference on Applications of Computer Vision (WACV), January 5-9, 2015, IEEE, Waikoloa Village, Hawaii, ISBN:978-1-4799-6684-4, pp: 1177-1184.

Li, Z., X. Liu and S. Zhang, 2016. Shot boundary detection based on multilevel difference of colour histograms. Proceedings of the 1st International Conference on Multimedia and Image Processing (ICMIP), June 1-3, 2016, IEEE, Beijing, China, ISBN:978-1-4673-8941-9, pp: 15-22.

Madhusudhan, M.V. and C. Hegde, 2015. Video shot boundary detection using methods. Proceedings of the 2015 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE), December 19-20, 2015, IEEE, Bengaluru india, ISBN:978-1-4673-8785-9, pp: 52-56.

Mahmoud, K., N. G hanem and M. Ismail, 2013. Vgraph: An effective approach for generating static video summaries. Proceedings of the 2013 IEEE International Conference on Computer Vision, December 2-8, 2013, IEEE, Alexandria, Egypt, ISBN:978-1-4799-1670-2, pp: 811-818.

Srilakshmi, B. and R. Sandeep, 2015. Shot boundary detection using structural similarity index. Proceedings of the 5th International Conference on Advances in Computing and Communications (ICACC), September 2-4, 2015, IEEE, Bengaluru, India, ISBN:978-1-4673-6993-0, pp: 439-442.

Sun, J. and Y. Wan, 2014. A novel metric for efficient video shot boundary detection. Proceedings of the 2014 IEEE Conference on Visual Communications and Image Processing, December 7-10, 2014, IEEE, Lanzhou, China, ISBN:978-1-4799-6140-5, pp: 45-48.

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 Conference on Visual Communications and Image Processing (VCIP) 2016, November 27-30, 2016, IEEE, Chengdu, China, ISBN:978-1-5090-5317-9, pp: 1-4.

Yoo, H.W., H J. Ryoo and D.S Jang, 2006. Gradual shot bundary dtection using localized edge blocks. Multimedia Tools Appl., 28: 283-300.

Yuan, J., H. Wang, L. Xiao, W. Zheng, J. Li, F. Lin and B. Zhang,, 2007. A formal study of shot boundary detection. IEEE Trans. Circuits Syst. Video 17: 168-186.