A Review on Shot Boundary Detection

[*]Ms. Pragati Ashok Kevadkar, [#]Prof. Jadhav dattatray A.

[*]PG student, [#]Professor, Department of Electronics and Telecommunication, JSPM's ICOER, Wagholi, Pune, Maharashtra, India. [*]kevadkarpa28@gmail.com, [#]prof.jadhavda@gmail.com

Abstract - Nowadays, a large number of video datasets available online. There is rapid growth in the multimedia data, because of the improvement in the data acquisition, storage and communication technologies, supported by improvement in audio and video signals. As the dataset is large, peoples find out useful information from the video. This can be facilitated by Content-Based Video Retrieval (CBVR) Methods. CBVR can be divided into three steps, segment the video, feature extraction and video retrieval based on the query. From the above step, segmentation is an important step. In the segmentation process, the video is divided into shots that consist of continuous action in the time domain. In this paper, different approaches for video shot boundary detection will be critically reviewed.

Keywords —Video Retrieval, Shot boundary detection, video processing, CBVR

I. INTRODUCTION

In today's world, multimedia and web development technologies grow faster with the advancement of the internet. Due to this the multimedia content also increases rapidly. This leads to an increase in the problem of effectively managing the video data. On the internet, video is the most consumed data type. It consumes a large storage space and contains voluminous information [1]. The video contains large information because it consists of the frame (image), text and audio [2]. The human brain can gather most of the visual information that can process faster than another form i.e. audio and text. Hence video facilitated easy communication medium among the individuals [3].

In the last couple of decades, performance capability, storage capacity and the number of recording devices have increased rapidly, resulting in the active data uploading and watching videos at inconceivable rates [4]. The best example is YouTube which is the second most popular video-sharing website. Statistics show that 500 hours of videos were uploaded every minute in 2019 and it is more than 300 hours of videos per minute in 2016. This growth is because of the individual and companies sharing their content through Video Sharing Website (VSW) to increase their audience. On the other side, mobile technology increases day by day, as a result of the prevalence of mobile technology [5] which motivates the upload videos to social media. The availability of video editing software on computers enables users to edit, altering, the video content. In addition to this, uploading the video does not need skilled programmers that results in the duplication of video content.

The rapid increase in the amount of video data has led to the improvement in the technique for process and storing data [6]. A manual search is required to retrieve the appropriate image. To address this problem, the image search like facilities are provided in Google image search but the video search engine is yet unavailable. This is the main motivation of the research [7].

The image and video data to be processed, transmit and analyze using the advance techniques. A shot can be defined as the frame that has been captured uninterruptedly and continuously run of a single camera. This can be processed by identifying and developing improved Shot Boundary Detection (SBD). Video shot transition is divided into the abrupt and gradual shot transform. Abrupt transition is also called the hard cut is the sudden change in the consecutive frame of the video while the gradual transition is the automated transition used to detect the transition difference between consecutive shots in the video. A gradual transition is of three types: fade, dissolve and wipe.

A. A fade:

Fade transition is of two types, Fade-in and Fade-out. Fade in is occur when an image displayed from the black image while fade out appears when images faded to a black screen or dot. This effect lasts after a few frames.

B. A dissolve:

This transition is the asynchronous occurrence of fade in and fades out. These two effects are layered for a fixed period.

C. A wipe:

A wipe is an imaginary line between the clearing of the older scene and displaying a newer scene. It can occur over more frames. Mainly, the SBD algorithm is processed Feature extraction, similarity measurement and decision making about the presence/absence of a cut.

In this paper, the different approaches for SBD have been presented. This paper is organized as The methods used for

 DOI : 10.35291/2454-9150.2020.0200

shot boundary detection have been presented in section II. In section III, the metrics used for SBD have been explained. In section V, the paper has been concluded.

II. SHOT BOUNDARY DETECTION

In this section, different methods for SBD has been explained.

A. Features Based Techniques

In the features based technique, features are calculated either from the frame or region of the frame. The features are explained below.

1) Color and Luminance: Features include Luminance and color where the average of grayscale luminance and color pixels are calculated [8-9], histogram features.

2) Histogram: Histogram represents the character of the image. It is easier to calculate and invariant to rotation, translation, and zooming of the frame [10].

3) Edges: These features are based on the edge of the ROI of the frame. Edges can be used to extract the Region Of Interest (ROI) of the object. These features are invariant to the illumination changes, the motion of the camera, and the visual observation of humans. These features are computationally costlier, highly noise sensitive and high dimensional

4) Transform Domain Coefficient Features: The transform like Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT) are used to extract the texture features from the frame. The pro of this method is it is easy to calculate the fast processing. The only disadvantage is these features are variant to zooming.

B. Spatial Domain Feature

The region of interest selection and extract the features from that ROI acting a significant role in the shot change detection. Luminance like features are extracted from each pixel of the frame and it is used for shot detection. In some approaches, the frame is divided into the block and the set of features is calculated per block. Spatial domain features are invariant to the small object motion [11]. This method has some disadvantages like it required high computational complexity and instability of the algorithm. Also, it shows poor performance while performance measurement between two similar shots.

C. Temporal Domain

The shot boundary can be detected in the temporal domain, where the temporal window of the frame is used to detect the shot change. It is also called a temporal domain of continuity metric. The dissimilarity between the consecutive frames can be measured by the discontinuity metrics [11]. This approach may fail when activity varies significantly. This problem can be resolved by considering all frames within the temporal window [10] and Comparing it with frame-by-frame discontinuity metric or by measuring the discontinuity metric of the window. In a different method, one or more statistical features are calculated for the whole shot and compare with the next frame for consistency [10, 11]. If there is the existence of variation within shots, statistical features then the calculation for an entire shot may not be effective. In a different approach, the complete video is taken as consideration to measure its characteristics for detecting shot change [12]. This method has some disadvantages: The system may fail for video with variations occurs within the shot and the features extracted within that area may not effective for the decision of SBD.

D. Shot Change Detection Technique

There are different techniques to detect shot change.

5) Thresholding: In this method, the values of the calculated features are compared with the fixed threshold [10, 11]. The selection of value is one of the key problems. The threshold value for one frame may not be applied to another frame.

6) Adaptive Thresholding: The generalization of the approach cannot be achieved using manual thresholding. The reported problem can be solved by calculating threshold value differ based on average discontinuity in the temporal domain [13]

7) Probabilistic Detection: SBD can be done by modeling the pattern of detailed types of shot transitions and then altering the shot estimation assuming their exact probability distributions [9, 14].

8) Trained Classifier: In this approach, the model is trained for shot detection by providing the images to the classifier. The trained classifier will classify the frame into "shot change" and "no shot change" classes [15]. These approaches are more efficient than the previous approaches but required higher computational complexity.

III. VARIOUS SHOT DIFFERENCE MEASUREMENT APPROACH

From the survey of the different papers and approaches, it is observed that previous work is mostly done to detect automatic shot detection in video, but recently approached mainly focused on gradual transition detection. The techniques used for these approaches are mainly, pixel differencing, statistical differencing, histogram differencing, color and luminance differencing, edge differencing and motion vectors.

The measurement approach for the shot difference is explained in detail below.

 DOI : 10.35291/2454-9150.2020.0200

E. Pixel Comparison

The pixel difference between two successive video frames or the percentage of pixels that have been changed in successive frames is compared. This approach is sensitive to fast object and camera movement, camera panning or zooming. The pixel differences (PD) metric is defined as,

$$PD(i) = \frac{1}{PQ}\sum_{x=1}^{P}\sum_{y=1}^{Q}|(f[x,y,i] - f[x,y,i+1])| \quad (1)$$

where P is the number of rows in the frame and Q is the number of columns in the image.

F. Statistical Based Difference

The frames are grouped into small regions and the statistical feature of each pixel within the group is calculated of each successive frame. In [16], Kasturi et al. calculate the standard deviation and mean of the gray level pixels of the various region of the image. The advantage of this method is it is noise-tolerant but has a disadvantage of slow computation speed due to complex statistical computation.

G. Likelihood Ratio

The likelihood ratio compares the statistical features of the corresponding block or region between two successive frames. If the likelihood ration of the transition is greater than the defined threshold value, then it is considered as the region is changed. This method minimizes the problem of false detection because of camera motion. Likelihood Ratio is defined as,

$$LHR(i) = \frac{\left[\left(\frac{\sigma_i+\sigma_{i+1}}{2}\right)+\left(\frac{(\mu_i-\mu_{i+1})^2}{2}\right)\right]^2}{(\sigma_i \times \sigma_{i+1})} \quad (2)$$

Where LHR is a likelihood ratio between two consecutive regions; where

$$\mu_i = \frac{1}{PQ}\sum_{x=1}^{P}\sum_{y=1}^{Q} f[x,y,i] \quad (3)$$

and

$$\sigma_i = \left[\frac{1}{PQ}\sum_{x=1}^{P}\sum_{y=1}^{Q}(f[x,y,i] - \mu_i)^2\right]^{\frac{1}{2}} \quad (4)$$

H. Histogram Difference

The histogram shows the nature of the image. In this method histogram of the successive frame is calculated and compared with each other's histogram. If the difference of histogram (bin-wise) among the histograms becomes larger than the predefined threshold, then the SBD is detected. The color histogram change rate is used as an evaluation metric by Ueda et. al. [17].

Histogram difference is defined by

$$HD(i) = \sum_{j=1}^{G}|(H_i[j] - H_{i+1}[j])| \quad (5)$$

Where, $H_i[j]$ and $H_{i+1}[j]$ represent the histogram for the i^{th} frame and $(i+1)^{th}$ frame, respectively, and j is the G possible gray levels.

I. Region-Based Histogram Differences

In this method, the image is divided into regions or blocks. Boreczky et al. [18] segregate each image into 16 blocks of a 4x4 mask. For each image, a 64-bin histogram is measured for each region. A Euclidean distance is calculated between the histogram of the successive block. If the distance is greater than the predefined threshold value, region count for that region is incremented. If the count is large than the threshold value then declared as the SBD.

J. Edge Based Difference

In this method, the edges of consecutive frames are extracted. Then, the newly added edges and disappeared edges were observed. If the difference is greater than some threshold value then declared as the shot boundary. Zabih et al. [19] compared the edges of the frames to detect the shot boundary.

K. Motion Vectors

Ueda et al. [17] and Zhang et al. [20] used motion vectors for measurement of the shot is zoom or pan using a block matching algorithm.

L. Adaptive Thresholding

Instead of local or predefine threshold, the adaptive threshold is the better option to enhance the shot change detection precision. The similarity function can be evaluated using histogram similarity and equivalent contextual region (ECR).

M. Chi-square Test

Nagasaki and Tanka [22] experimented using the histogram and pixel difference metrics and it is observed that the histogram metrics show the effectiveness. They observed the best results obtained by classifying the images into 16 equal regions, using a Chi-square test on the RGB histogram of those regions. Chi-square (CS) test is defined as.

$$CS = \sum_{j=1}^{G}\frac{|(H_i[j] - H_{i+1}[j])|^2}{H_{i+1}[j]} \quad (6)$$

Where, $H_i[j]$ denotes the histogram value for the i^{th} frame and j is one of the G possible gray levels.

IV. DISCUSSION

The previous methods are mostly based on the calculation of the statistical parameters like pixel comparison, likelihood ratio, histogram differencing and chi-square test. But these approaches are mostly an environment dependant.

In almost all algorithms reviewed in this approach, abrupt transitions were effectively detected than the gradual transitions. These methods show false detection in some cases such as lighting effect, large object motion near to the camera, and fast camera motion.

Hence there is a need for a robust, accurate and generalized algorithm to detect the shot boundaries. This can be achieved by the following approach.

The main distinguishing factor in new presented approach is the Dilated DCNN cell which is summarized as four 3D 3x3x3 convolution operations. This approach results in a significant reduction in trainable parameters as compared to standard 3D convolutions with the same field

 DOI : 10.35291/2454-9150.2020.0200

of view.

It is used to expand the receptive fields of the network exponentially with linear parameter accretion. This results in acquiring more contextual information and increases the accuracy of state-of-the-art shot boundary detection systems. Multiple DDCNN cells stacked on top of each other followed by spatial max-pooling layer forms a Stacked DDCNN block.

V. CONCLUSION

SBD is a critical step for video processing. The important step for the video shot detection system is the segmentation of video into shots. In this paper, a review of methods of SBD is presented. The method of this paper is mainly classified into three classes based on the input to the segmentation. This method has been divided based on the Pixel, Block and histogram. It is noticed that the use of previous methods has relatively poor performance in terms of evaluation metrics while a combination of more than one algorithm may perform better than the existing one.

In recent years, the deep learning approached attracts researchers due to its robustness and generalization. The deep learning approaches are more accurate, environmentally independent, hence in the future, the SBD algorithm can be implemented using deep learning algorithms like Convolutional Neural Network (CNN) to detect the transitions in the scenes. This could improve the performance of the system.

REFERENCES

[1] Priya, R.; Shanmugam, T.N. "A comprehensive review of significant researches on content-based indexing and retrieval of visual information. Front. Comput. Sci. 2013, 7, 782–799.

[2] Yuan, J.; Wang, H.; Xiao, L.; Zheng, W.; Li, J.; Lin, F.; Zhang, B. A formal study of shot boundary detection. IEEE Trans. Circ. Syst. Video Technol. 2007, 17, 168–186.

[3] Palmer, S.E. Vision Science: Photons to Phenomenology; MIT Press: Cambridge, MA, USA, 1999.

[4] Del Fabro, M.; Boszormenyi, L. "State-of-the-art and future challenges in video scene detection: A survey", Multimedia Syst. 2013, 19, 427–454.

[5] Gonzalez-Diaz, I.; Martinez-Cortes, T.; Gallardo-Antolin, A.; Diaz-de Maria, F. Temporal segmentation and keyframe selection methods for user-generated video search-based annotation. Expert Syst. Appl. 2015, 42, 488–502.

[6] Fayk, M.B.; El Nemr, H.A.; Moussa, M.M. Particle swarm optimization based video abstraction. J. Adv. Res. 2010, 1, 163–167.

[7] Parmar, M.; Angelides, M.C. MAC, "REALM: A Video Content Feature Extraction and Modelling Framework", Comput. J. 2015, 58, 2135–2170.

[8] Campisi, P., Neri, A., Sorgi, L.: Automatic dissolve and fade detection for video sequences. In: Proc. Int. Conf. on Digital Signal Processing, vol. 2, pp. 567–570 (2002)

[9] Lelescu, D., Schonfeld, D.: Statistical sequential analysis for real-time video scene change detection on the compressed multimedia bitstream. IEEE Trans. on Multimedia 5(1), 106–117 (2003)

[10] Cernekova, Z., Kotropoulos, C., Pitas, I.: Video shot segmentation using singular value decomposition. In: Proc. 2003 IEEE Int. Conf. on Multimedia and Expo, vol. II, pp. 301–302 (2003)

[11] Heng, W.J., Ngan, K.N.: An object-based shot boundary detection using edge tracing and tracking. Journal of Visual Communication and Image Representation 12(3), 217–239 (2001)

[12] Boreczky, J.S., Rowe, L.A.: Comparison of video shot boundary detection techniques. In: Storage and Retrieval for Image and Video Databases (SPIE), pp. 170–179 (1996)

[13] Campisi, P., Neri, A., Sorgi, L.: Automatic dissolve and fade detection for video sequences. In: Proc. Int. Conf. on Digital Signal Processing, vol. 2, pp. 567–570 (2002)

[14] Hanjalicb, L.: Shot-boundary detection: Unraveled and resolved. IEEE Transaction Circuits and Systems for Video Technology 12(2), 90–105 (2002)

[15] Lienhart, R.: Reliable dissolve detection. In: Storage and Retrieval for Media Databases. Proc. of SPIE, vol. 4315, pp. 219–230 (2001)

[16] Kasturi, R., Jain, R.: Dynamic vision. In: Kasturi, R., Jain, R. (eds.) Computer Vision: Principles, pp. 469–480. IEEE Computer Society (1991)

[17] Ueda, H., Miyatake, T., Yoshizawa, S.: IMPACT: an interactive natural-motion-picture dedicated multimedia authoring system. In: Proc. CHI 1991, pp. 343–350 (1991)

[18] Boreczky, J.S., Rowe, L.A.: Comparison of video shot boundary detection techniques. In: Storage and Retrieval for Image and Video Databases (SPIE), pp. 170–179 (1996)

[19] Zabih, R., Miller, J., Mai, K.: A feature-based algorithm for detecting and classifying scene breaks. In: Proc. ACM Multimedia 1995, pp. 189–200 (1995)

[20] Zhang, H.J., Kankanhalli, A., Smoliar, S.W.: Automatic partitioning of full-motion video. Multimedia Systems, 10–28 (1993)

[21] Acharjee, S., Chaudhuri, S.S.: A New Fast Motion Vector Estimation Algorithm for Video Compression. In: ICIEV 2012, pp. 1216–1219 (2012)

[22] A Nagasaka, and Y Tanka, "Automatic video indexing and full-video search for object appearance", Visual Database Systems II, E Knuth, and L Wegner Editors., Elsevier Science Publishers, pp. 113-27, 1992.

[23] Sadiq H. Abdulhussain, Abd Rahman Ramli, M. Iqbal Saripan, Basheera M. Mahmmod, Syed Abdul Rahman Al-Haddad and Wissam A. Jassim, "Methods and Challenges in Shot Boundary Detection: A Review", Entropy 2018, 20(4), 214, 2018.

[24] Tomas Soucek, Jaroslav Moravec, Jakub Lokoc, "TransNet: A deep network for fast detection of common shot transitions", Computer Vision and Pattern Recognition, 2019.

 DOI : 10.35291/2454-9150.2020.0200