Video Shot Boundary Detection: A Review

Gautam Pal[1], Dwijen Rudrapaul[2], Suvojit Acharjee[3], Ruben Ray[4], Sayan Chakraborty[5], and Nilanjan Dey[6]

[1] Dept. of CSE, TIT Agartala, India
[2] Dept. of CSE, NIT Agartala, India
[3] Dept. of ECE, NIT Agartala, India
[4] Dept. of IT, GCELT Kolkata, India
[5] Dept. of CSE, BCET, Durgapur
[6] Dept. of ETCE, Jadavpur University Kolkata, India
{goutamtit2003,dwijen.rudrapal,acharjeesuvo,ruben.ray,
sayan.cb,neelanjan.dey}@gmail.com

Abstract. Video image processing is a technique to handle the video data in an effective and efficient way. It is one of the most popular aspects in the video and image based technologies such as surveillance. Shot change boundary detection is also one of the major research areas in video signal processing. Previous works have developed various algorithms in this domain. In this paper, a brief literature survey is presented that establishes an overview of the works that has been done previously. In this paper we have discussed few algorithms that were proposed previously which also includes histogram based, DCT based and motion vector based algorithms as well as their advantages and their limitations.

Keywords: Histograms, Transition, Shots, Thresholding, Shot boundary detection.

1 Introduction

Shot boundary detection (SBD) is necessary for automatic video indexing and browsing. It can be applicable for numerous applications like indexing in video database, video compression etc. The basic unit of any video is frame. The structure of a video is shown in Fig. 1. The frame sequences are indexed by frame number. After breaking the video, the obtained frames have identical size. Generally in every one second 25-30 frames are taken. A video shot is a sequence of interrelated consecutive frames taken by a single camera at a stretch. In general, shots are combined to produce a video. Scene may be consist of a single or multiple shots which describe a story unit within a video. Shot boundary detection is based on the identification of visual dissimilarity due to the transitions. The mismatch between two frames generally found while shot change. This dissimilarity appears in different form which is categorized into two types: abrupt (as hard cut) and gradual (dissolve, fade in, fadeout, wipe). An abrupt shot change can be seen in single frame. Fade refers to the slow change of brightness in video that often results in a solid black frame. If the

S.C. Satapathy et al. (eds.), *Emerging ICT for Bridging the Future – Volume 2,*
Advances in Intelligent Systems and Computing 338, DOI: 10.1007/978-3-319-13731-5_14

first shot's pixel are replaced with pixels of second shot in a sequential pattern (for eg. in a pattern from left edge of the video frames to the right), then wipe occurs. A dissolve refers to the process of first shot's images getting dimmed and second shot's images getting brighten, that results into frames within the transition illustrating one image is getting superimposed on the other images.

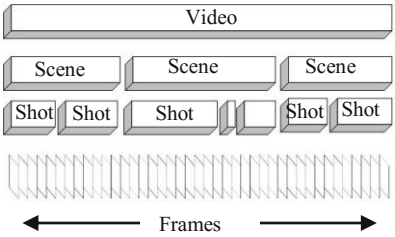

Fig. 1. Structure of a Video

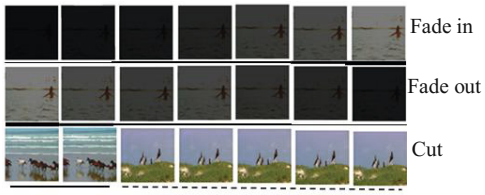

Fig. 2. Gradual and Sudden shot change

1.1 Literature Review

This section focuses on the previous works carried out in the domain of shot boundary detection. A discussion proposed by Boreczky et al. [1] in the year 1996, described different video shot boundary detection techniques and presented a study on the performances of various shot boundary detection algorithms. In the year 1999 Gauch et al. [2] introduced a new technique for Shot Change Detection. This work stated, if two similar shots were joined with a gradual cross fade, then the visual changes might become much smaller than expectation. Shot detection in the VISION (video indexing for searching over networks) system is done by combining three parameter : (1) the average brightness of each video frame, (2) the change in pixel values between two successive frame and (3) the change in color distribution between two consecutive frame. These three quantities are compared with dynamic thresholds to identify the shot change boundaries. In 2001, Heng et al. [3] proposed a shot detection technique based on objects. In this work, finding information with the help of a time stamp transferring mechanism from multiple frames was proposed. Gradual transitions were efficiently handled by this mechanism. This algorithm showed many advantages on the traditional algorithms. In the same year, a simple method of histogram comparison was proposed by Lee et al. [4]. This method ignored spatial information in the frame and was not suitable during highly abrupt luminance changes. Later, Liaoff et al. [5] proposed a smart approach for detecting dissolve in a video by using binomial distribution model. This method determined the threshold needed for discriminating the dissolve due to motions. Liu et al. [6] proposed an algorithm with constant false-alarm ratio (CFAR) for video segmentation. For video cut detection, a theoretical threshold determination strategy using the non-parametric CFAR was developed. It was also capable of finding a controllable precision. Liang et al. [7] used number of edge-based for detecting sudden shot boundaries to avoid the influence of frequent flashlights in many videos. Fuzzy logic was used by Fang et al. [8] to integrate hybrid features that can detect shot boundaries correctly. Shot boundary with gradual shot

cuts and sudden shot cuts were detected by different process. Different features were used for fuzzy logic approach based temporal segmentation of videos. Cao et al. [9] proposed a classifier based approach to find the wipes and digital video effects. Six parameters that are causes the formation of feature vector were evaluated for every frame in a temporal window. A supervised SVM classifier based on feature vectors classified the frames in sudden shot change, gradually shot change and no scene change categories. An automatic shot detection technique proposed by Huang, et al. [10] performed very poorly due to high false detection rates caused by camera or object motion. This work tried to overcome the issues in shot detection mechanisms, using local key point matching of video frames that detected both sudden and gradual change efficiently. Changes with long time (i.e fade in, fade out or dissolve) are very hard to locate using low level features. Although, objects compared between two consecutive frames can help to easily identify shot [11] changes. On one hand effect of camera motion and object motion can be easily avoided by detecting objects inside the frames. At the same time it can efficiently detect both abrupt transitions and gradual transitions. This provides a single approach for different shot boundary detection. ASCD can be considered as an automatic operation in real time application. Weighting variance of frame difference and histogram variation extensively used by ASCD. Adaptive threshold can be calculated from the following frame of a last shot change to the previous one of a current frame. It also uses histogram variation of successive frames to set an automatic weighting factor. An automatic SCD algorithm using mean and variance based have a higher detection rate than pixel based methods. K-means clustering was used by Xu et al. [12] in the shot detection algorithm. It first extracted the color feature and then obtained the dissimilarity of video frames using the features. Video frames were grouped by performing graph-theoretical algorithm in Xu et al.'s [13] proposed work. Block based motion was used in order to detect the change in shot which was described by Park et al. [15]. The modified displaced frame difference (DFD) and the block wise motion similarity are combined for the detection of change in shot. Mishra et al. [16] proposed an algorithm that firstly mined structure features from every video frame [17] using dual tree complex wavelet transform and then spatial domain structure similarity was computed between consecutive frames. Later a comparative study was performed [18] between block matching SBD algorithm and dual tree complex wavelet transform based SBD algorithm, in terms of various parameters like false rate, hit rate, miss rate tested on a set of different video sequence. Lu et al. [17] presented a Video Shot Boundary Detection technique that was based on segment selection and singular value decomposition (SVD) with Pattern Matching. In this work shot boundaries' position and gradual transitions' length were calculated using adaptive thresholds and most of the non-boundary frames were discarded at the same time.

2 Methodology

2.1 Features Based Techniques

Features are calculated either from the whole frame or from a portion of it, which is also known as region of interest (ROI). Features include the following parameters: (I)

Luminance and color: The simplest feature which is used to recognize a ROI is its average grayscale luminance. It is capable of illumination changes [19, 20], (II) Histogram**:** Another feature for ROI is grayscale or color histogram [21]. It is easier to calculate and often insensitive to rotational, translational and zooming motion of the camera, (III) Image edges: This feature is based on edge information of a ROI. Edges can be used to combine objects or to extract ROI statistics. These are independent of illumination changes, camera motion and correspond to the human visual perception. Although, it has high computational cost, high noise sensitivity and high dimensionality, (IV) Transform coefficients (DFT, DCT, and wavelet): These features are used to describe the texture of a ROI.

Merits: Easy to calculate and fast processing
Demerits: These kinds of features are variant to camera zoom.

2.2 Spatial Domain Feature

The region from where, individual features taken, plays a great role in the performance of shot change detection. Features like luminance are extracted from every pixel of a frame and used for shot detection. In next approach each frame is divided into equal-size blocks and a set of features is extracted per block.

Merits: 1) This approach is invariant to small camera and object motion [20].
2) Feature extraction from arbitrarily shaped and sized regions exploits the homogeneous regions, with better detection of discontinuities.
Demerits: 1) This kind of feature is high computational complexity and instability due to the complexity of the algorithms involved.
2) It also has poor performance while measuring the difference [21] among two similar shots.

2.3 Temporal Domain of Continuity Metric

Another choice of shot boundary detection algorithms in which temporal window of frames is used to do shot change detection also referred as temporal domain of continuity metric. Dissimilarity between two successive frames can be measured by looking for discontinuity metric's higher value between two adjacent frames [22]. This approach may fail if different parts' activity varies significantly. To resolve the above problems the dissimilarity can be detected by using the features of all existing frames within that temporal window [21]. This can be done by calculating dynamic threshold and comparing with frame-by-frame discontinuity metric or by measuring discontinuity metric directly on the window. In another method one or more statistical features is calculated for whole shot and compare with the next frame for consistency [21, 22]. If there is existence of variation within shots, statistical features then calculation for an entire shot may not be effective. In different approach the complete video is taken as consideration to measure its characteristics for detecting shot change [28].

Demerits: 1) The system fails if the video has variation within and between the shots. 2) If there is existence of variation within shots, statistical features then calculation for an entire shot may not be effective.

2.4 Shot Change Detection Technique

(I) Thresholding: Calculated feature values are compared with a fixed threshold [21, 22], (II) Adaptive Thresholding: In this type of thresholding the above mentioned problem is solved by taking threshold value which can vary based on average discontinuity within a temporal domain [19], (III)Probabilistic Detection: Shot changes detection can be done by modeling the pattern of specific types of shot transitions and then changing the shot estimation assuming their specific probability distributions [20,23], (IV) Trained Classifier: This technique formulates the shot change detection as a classification problem, with two classes: "Shot change" and "no shot change" [24].

Merits: This is more efficient than previous approaches.
Demerits**:** This type of techniques comes with a higher computational complexity.

3 Various Shot Difference Measurement Approach

Previous works were done to detect automatic cut within video, but recent research are mainly focused on detecting gradual transitions. Thus, the techniques used recently for these types of works use pixel differences, statistical differences, histogram comparisons, edge differences, compression differences, and motion vectors.

3.1 Pixel Comparison

Pixel difference between two successive video frames or the percentage of pixels that has been changed in two successive frames is compared. This approach is sensitive to fast object and camera movement, camera panning or zooming,

3.2 Statistical Based Difference

Frames are divided into small regions and statistical feature of each pixels within these regions are calculated of each successive frames. Kasturi and Jain [25] calculated standard deviation as well as the mean of the gray levels in various regions of the images. This approach is noise tolerant, but slow due to complex statistical computation.

3.3 Likelihood Ratio

It minimizes the problem of false detection due to camera movements. Without comparing the pixels, likelihood ratio compares statistical features known as likelihood ratio of the corresponding regions or blocks in two successive frames. If the likelihood ratio is larger than the threshold, then it is assumed that region is changed.

3.4 Histogram-Based Difference

Color histogram of each frame is calculated and compared among each other to detect shot boundaries. If the bin-wise difference among the two histograms becomes larger than the preset threshold, then a shot boundary is detected. In order to detect shot boundaries, Ueda *et al.* [26] applied the color histogram change rate.

3.5 Region Based Histogram Differences

Boreczky and Rowe [28] spitted each image into 16 blocks consists of 4x4 pattern**.** For each image, a 64-bin gray-scale histogram is measured for each region. A Euclidean distance is further calculated to find the difference between the region histograms of two successive images. If the distance is more than a threshold, the region count for that image is incremented. If the value of count is larger than some preset threshold, a shot boundary is detected.

3.6 Edge Based Difference

In this approach the edges of consecutive aligned frames are detected first and then the edge pixels are paired with nearby edge pixels in the other image to find out if any new edges have entered the image or whether some old edges have disappeared. Zabih *et al.* [29] compared color histograms, chromatic scaling and their own technique based on edge detection.

3.7 Motion Vectors

Ueda *et al.* [30] and Zhang *et al.* [31] used motion vectors [32,32,34] on block matching algorithm in order to measure if the shot was a pan or a zoom.

3.8 Adaptive Thresholding

Choosing the proper threshold value is an important criterion in both color histogram comparison and edge change tracking algorithms. An adaptive threshold can be a better option to enhance the shot change detection precision. It uses the local thresholds of the feature or similarity function to be compared, which may be histogram similarity and equivalent contextual region (ECR), respectively.

4 Results and Discussion

Different features and different methods were used to solve the different challenges in the shot boundary detection algorithm. Sometimes multiple features are combined for selection. In this section, a comparative study of popularity of different features and techniques is provided. Though these algorithms are efficient for structured videos but still shot boundary detection is a huge challenge till now for videos with fast changes

between shots. This study will help us to find the most popular features and techniques as well as the less explored features and techniques in SBD. The less explored ones can be explored further to solve the challenges faced in SBD in our future work.

Table 1. Utilization popularity of different features in Shot Boundary Detection algorithms

Luminance and Color	Histogram Analysis	Edge Information	Transformation Coefficients (DCT, DFT etc.)	Statistical Measurement	Motion Analysis	Object Detection
4	28	3	7	16	13	2
8%	56%	6%	14%	32%	26%	4%

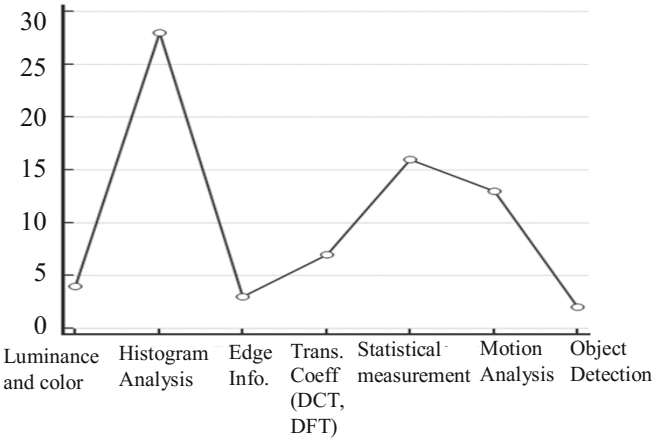

Fig. 3. Utilization popularity of different features in Shot Boundary Detection algorithms

Table 1 and Fig. 3 show the utilization popularity of different features in shot boundary detection algorithm. Fifty research papers were studied from year 1996 to 2014 to calculate the utilization popularity. Hence, the percentage in Table 1 suggests the percentage of usage popularity of every feature in the procedures. The most popular feature to find the change in shot is histogram analysis. Histogram can be of gray images or color images that are used mainly in histogram analysis. Different statistical measurement like variance and mean of intensity and color are also used frequently in different algorithms. Sometimes statistical measurements are combined with histogram and other features for detection of shot boundary. Motion analysis between two consecutive frames is another most popular technique in shot boundary detection.

5 Conclusion

Shot change detection is a very challenging task. For unstructured video, the shot change detection techniques performed were very poor. Only for structured video,

these techniques produced better and comprehensive results with a very steady shot change rate. Besides these, the results of the discussed techniques in this work were inefficient for videos with very fast shot change. Hence, there is a scope of further improvement in these techniques. The less explored features like object detection and object analysis can be explored more to improve the techniques.

References

1. Boreczky, J.S., Rowe, L.A.: Comparison of video shot boundary detection Techniques. Journal of Electronic Imaging 5(2), 122–128 (1996)
2. Gauch, J.M., Gauch, S., Bouix, S., Zhu, X.: Real time video scene detection and classification. Information Processing and Management 35, 381–400 (1999)
3. 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, 217–239 (2001)
4. Lee, M.-S., Yang, Y.-M., Lee, S.-W.: Automatic video parsing using shot boundary detection and camera operation analysis. Pattern Recognition 34, 711–719 (2001)
5. Liao, H.Y.M., Su, C.W., Tyan, H.R., Chen, L.H.: A motion-tolerant dissolve detection algorithm. In: IEEE 2nd Pacific-Rim Conference on Multimedia, vol. 2195, pp. 1106–1112 (2002)
6. Liu, T.-Y., Lo, K.-T., Zhang, X.-D., Fengc, J.: A new cut detection algorithm with constant false-alarm ratio for video segmentation. J. Vis. Commun. Image R. 15, 132–144 (2004)
7. Liang, L., Liu, Y., Lu, H., Xue, X., Tan, Y.-P.: A Enhanced Shot Boundary Detection Using Video Text Information. IEEE Transactions on Consumer Electronics 51(2), 580–588 (2005)
8. Fang, H., Jiang, J., Feng, Y.: A fuzzy logic approach for detection of video shot boundaries. Pattern Recognition 39, 2092–2100 (2006)
9. Cao, J., Cai, A.: A robust shot transition detection method based on support vector machine in compressed domain. Pattern Recognition Letters 28, 1534–1540 (2007)
10. Huang, C.-R., Lee, H.-P., Chen, C.-S.: Shot Change Detection via Local Key point Matching. IEEE Transactions on Multimedia 10(6), 1097–1108 (2008)
11. Kim, W.-H., Moon, K.-S., Kim, J.-N.: An Automatic Shot Change Detection Algorithm Using Weighting Variance and Histogram Variation. In: ICACT, pp. 1282–1285 (2009)
12. Xu, L., Xu, W.: A Novel Shot Detection Algorithm Based on Clustering. In: 2010 2nd International Conference on Education Technology and Computer (ICETC), pp. 1570–1572 (2010)
13. Xu, W., Xu, L.: A Novel Shot Detection Algorithm Based on Graph Theory. In: 2010 2nd International Conference on Computer Engineering and Technology (ICCET), pp. 3628–3630 (2010)
14. Donate, A., Liu, X.: Shot Boundary Detection in Videos Using Robust Three-Dimensional Tracking. IEEE Transactions on Consumer Electronics, 64–69 (2010)
15. Park, M.-H., Park, R.-H., Lee, S.-W.: Efficient Shot Boundary Detection for Action Movies Using Blockwise Motion-Based Features. In: Bebis, G., Boyle, R., Koracin, D., Parvin, B. (eds.) ISVC 2005. LNCS, vol. 3804, pp. 478–485. Springer, Heidelberg (2005)
16. Mishra, R., Singhai, S.K., Sharma, M.: Video shot boundary detection using dual-tree complex wavelet transform. In: 2013 IEEE 3rd International Advance Computing Conference (IACC), pp. 1201–1206 (2013)

17. Lu, Z.M., Shi, Y.: Fast Video Shot Boundary Detection Based on SVD and Pattern Matching. IEEE Transactions on Image Processing 22(12), 5136–5145 (2013)
18. Mishra, R., Singhai, S.K., Sharma, M.: Comparative study of block matching algorithm and dual tree complex wavelet transform for shot detection in videos. In: 2014 International Conference on Electronic System, Signal Processing and Computing Technologies (ICESC), pp. 450–455 (2014)
19. 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)
20. Lelescu, D., Schonfeld, D.: Statistical sequential analysis for real-time video scene change detection on compressed multimedia bit stream. IEEE Trans. on Multimedia 5(1), 106–117 (2003)
21. 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)
22. 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)
23. Hanjalicb, L.: Shot-boundary detection: Unraveled and resolved. IEEE Transaction Circuits and Systems for Video Technology 12(2), 90–105 (2002)
24. Lienhart, R.: Reliable dissolve detection. In: Storage and Retrieval for Media Databases. Proc. of SPIE, vol. 4315, pp. 219–230 (2001)
25. Kasturi, R., Jain, R.: Dynamic vision. In: Kasturi, R., Jain, R. (eds.) Computer Vision: Principles, pp. 469–480. IEEE Computer Society (1991)
26. Ueda, H., Miyatake, T., Yoshizawa, S.: IMPACT: an interactive natural-motion-picture dedicated multimedia authoring system. In: Proc. CHI 1991, pp. 343–350 (1991)
27. Nagasaka, A., Tanaka, Y.: Automatic video indexing and fullvideo search for object appearances. In: Visual Database Systems II, pp. 113–127 (1992)
28. 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)
29. Zabih, R., Miller, J., Mai, K.: A feature-based algorithm for detectingand classifying scene breaks. In: Proc. ACM Multimedia 1995, pp. 189–200 (1995)
30. Ueda, H., Miyatake, T., Yoshizawa, S.: IMPACT: an interactive natural-motion-picture dedicated multimedia authoring system. In: Proc. CHI 1991, pp. 343–350 (1991)
31. Zhang, H.J., Kankanhalli, A., Smoliar, S.W.: Automatic partitioning of full-motion video. Multimedia Systems, 10–28 (1993)
32. Acharjee, S., Chaudhuri, S.S.: A New Fast Motion Vector Estimation Algorithm for Video Compression. In: ICIEV 2012, pp. 1216–1219 (2012)
33. Acharjee, S., Dey, N., Biswas, D., Chaudhuri, S.S.: A Novel Block Matching Algorithmic Approach with Smaller Block Size for Motion Vector Estimation in Video Compression. In: ISDA 2012, pp. 668–672 (2012)
34. Acharjee, S., Dey, N., Biswas, D., Chaudhuri, S.S.: An Efficient Motion Estimation Algorithm using Division Mechanism of Low and High Motion Zone. In: i-Mac4s 2013, pp. 169–172 (2013)