

Shot Boundary Detection: Fundamental Concepts and Survey

1st Benoughidene Abdel halim
Department of computer science
University of Batna 2
Batna, Algeria
benouhalim@gmail.com

2nd Titouna Faiza
Department of computer science
University of Batna 2
Batna, Algeria
ffitouna@yahoo.fr

Abstract—A great part of the Big Data surge in our digital environments is in the form of video information. Hence automatic management of this massive growth in video content seems to be significantly necessary. At present researches topic on automatic video analyses includes video abstraction or summarization, video classification, video annotation and content based video retrieval. In all these applications one needs to identify shot boundary detection. Video shot boundary detection (SBD) is the process of segmenting a video sequence into smaller temporal units called shots. SBD is the primary step for any further video analyses. This paper presents the fundamental theory of the video shot boundary, and a brief overview on shot boundary detection approaches and their development. The advantages and disadvantages of each approach are comprehensively explored and challenges are presented. In addition to that, we focused on the machine learning technologies such as deep learning approaches for SBD could be directed as new directions for the future.

Index Terms—Shot Boundary Detection(SBD), Cut Transition (CT), Gradual Transition (GT), Temporal Video Segmentation, Video Content Analysis, Content Based Video Indexing and Retrieval (CBVIR), Feature Extraction, Machine Learning, Deep Learning, Convolutional Neural Networks (CNN), Multimedia Big Data.

I. INTRODUCTION

With the rapid development of computer networks and multimedia technology, the amount of multimedia data available every day is enormous and is increasing at a high rate, as well as the ease of access and availability of multimedia sources, which leads to big data revolution Multimedia.

Video is the most consumed data type on the Internet such as YouTube, Vimeo or Dailymotion, Yahoo Video, social networking sites like Facebook, Twitter, Instagram, etc. The explosive growth in video content leads to the problem of content management. However, people spent their time uploading and browsing huge videos to determine whether these videos were relevant or not, this is an difficult and stressful task for humans [1]. In such a scenario, it is necessary to have automated video analysis applications to represent information stored in large multimedia data. Such techniques are grouped into a single concept of Content-Based Video Indexing and Retrieval (CBVIR) systems. These applications include browsing of video folders, news event analyses, intelligent management of videos, video surveillance [2], key frame extraction [3], and

video event partitioning [4]. In addition, the video summary is the best and most effective solution for converting large, amorphous videos into structured, concise, clear and meaningful information. The main task of summarizing a video is to segment the original video into shots and extract key frames from the shots, which will be the most representative and concise of the entire video [5].

Video shot boundary detection (SBD) is also called shot segmentation, is the first process in video summarization, and its output significantly affects the subsequent processes. The main idea of video shot boundary is extracting the feature of video frames, and then detecting the shot type according to the difference the feature. There are two kinds of video shot boundary detection: Cut Transition (CT) and Gradual Transition (GT) [6]. In general, the performance of the shot boundary detection algorithm depends on its ability to detect transitions (shot boundaries) in the video sequence. Whereas, the accuracy of detection of shot boundary detection generally depends on the extracted features and their effectiveness in representing the visual content of video frames and the computational cost of the algorithm, which needs to be reduced [7]. Practically, there are some effects that appear in a video shot such as: flash lights or light variations, object/camera motion, camera operation (such as zooming, panning, and tilting), and similar background. Currently, there is no complete solution to these problems or most of them in the same algorithm. In other words, a favorable and effective method of detecting transitions between shots is still not available despite the increased attention devoted to shot boundary detection in the last two decades. This unavailability is due to randomness and raw video data size. Hence, a robust, efficient, automated shot boundary detection method is a necessary requirement [8].

Most of the existing reviews are not covering the recent advancements directions in the field of shot boundary detection as deep learning. This paper mainly focusing on review and analyze different kinds of shot boundary detection algorithms that are implemented in the uncompressed domain following their accuracy rate, computational load, feature extraction technique, advantages, and disadvantages. Future research directions are also discussed.

II. BASIC CONCEPTS OF SHOT BOUNDARY DETECTION

Partitioning a video sequence into shots is the first step toward video summarization. A video shot is defined as a series of interrelated consecutive frames taken contiguously by a single camera and representing a continuous action in time and space. As such, a shot boundary is the transition between two shots. This section presents the main concepts for shot boundary detection in videos [9].

- 1) **Video Definition :** A video is a collection of image frames arranged in a time-sequenced manner. As video consist of number of frames depend upon size of video. These frames occupy large space in memory. Frame rate is about 20 to 30 frames per second [10].
- 2) **Video Hierarchically :** A video can be broken down in scene, shot and frames. Scene is a logical grouping of shots into a semantic unit. A shot is a sequence of frames captured by a single camera in a single continuous action. The frames within a shot (intra-shot frames) contain similar information and visual features with temporal variations. A frame is the smallest unit that constitutes a shot [10]. (see Fig 1)

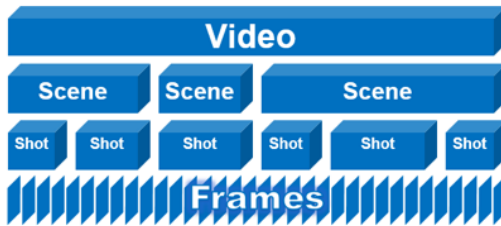


Fig. 1: Video Hierarchically.

- 3) **Shot transition types :** The transition between one shot and the following can be cut or gradual. The cut shot occurs when two successive shots are concatenated directly without any editing (special effects). This type of transition is also known as a abrupt or hard transition. The cut is considered a sudden change from one shot to another. By contrast, gradual shot occurs when two shots are combined by utilizing special effects throughout the production course. Gradual shot may span two or more frames that are visually interdependent and contain truncated information [11]. According to the different editing effects, there are several different kinds of gradual shot types, such as fade in/fade out , dissolve, wipe [12]. (see Fig 2)

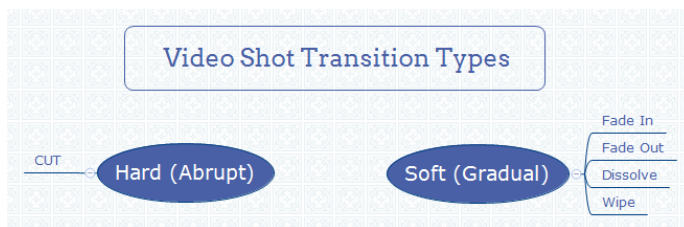


Fig. 2: Video shot transition types.

- a) **A Cut :** Is a sudden change from a video shot to another one [6]. (see Fig 3)



Fig. 3: (a) Cut Transition.

- b) **A fade out :** Occurs when the shot gradually turns into a single monochrome frame, usually dark [6]. (see Fig 4)



Fig. 4: (b) Fade out.

- c) **A fade in :** Takes place when the scene gradually appears on screen.[6]. (see Fig 5)

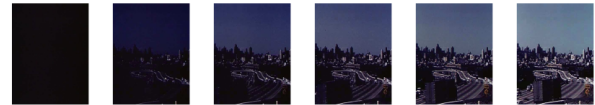


Fig. 5: (c) Fade in.

- d) **A dissolve :** Happens when a shot gradually replaces another one. One disappears as the following appears, and for a few seconds, they overlap, and both are visible. In the process of dissolve, two adjacent shots are temporally as well as spatially associated [6]. (see Fig 6)



Fig. 6: (d) Dissolve.

- e) **The wipe :** Is more dynamic and is considered as the most difficult to model and to detect. It happens when a shot pushes the other one off the screen. In this case, two adjacent shots are spatially separated at any time, but not temporally separated. Its difficulty lies in the number of types of wipe transitions that exists. Indeed, when a shot is moving from the screen (i/e leaving place to the other incoming shot), the movement can be either horizontal (i.e. from bottom to top or vice versa), vertical (e.g. from left to right), oblique (i.e. from a corner to the opposite one), starting from the center, going towards the center or others, etc [6]. (see Fig 7)



Fig. 7: (e) Various types of wipe transitions.

- 4) **The feature extraction** : Is the process to represent raw image in a reduced form to facilitate decision making such as pattern detection, classification or recognition. The features extracted from the video frames may be low-level, mid-level or high level features [13].
 - a) **Low-level features** : The low-level features are minor details of the image, like lines or dots, that does not take into consideration the visual or semantic. The low-level features consist of RGB values/histograms, intensity values, mean, variance, entropy of the pixel values etc [10].
 - b) **Mid-level features** : The mid-level features are intermediate between the low-level features and high level semantics. The mid-level features consist of feature point detectors and descriptors. Although, the feature points may be used for object identification in an image, these are not appropriate for high level semantic description of the content depicted in an image [10].
 - c) **High-level features** : High-level features are built to detect objects and larger shapes in the image, trajectory of paths followed by objects, motion vectors etc. These may be used for high level description of the content in an image [10].

Because of the importance of SBD, many researchers have presented algorithms to boost the accuracy of SBD for Cut Transition (CT) and Gradual Transition (GT). We introduce a survey on various SBD approaches below.

III. SHOT BOUNDARY DETECTION METHODS

Nowadays, many researchers are doing work to develop more reliable and accurate algorithms that can results into more precise shot boundaries. There are several common methods that deal with CT and/or GT:

A. Pixel-Based Methods

In this method, intensity of pixels is evaluated by taking two consecutive video frames and comparing pixel by pixel or the percentage of pixels that has been changed in two successive frames is compared. When the intensity of pixels is more than threshold, then it is referred to shot change [6].

The main drawback of such approaches (i.e intensity pixels), whatever the metric used, is sensitive to fast object and camera movement, camera panning or zooming. And limitations in this method is setting threshold manually.

B. Histogram-Based Methods

The most popular metric for cut transition detection is the difference between histograms of two consecutive frames. Histogram describes the distribution of gray, color, shape and texture without taking into account their position, so we can estimate the similarity between two images through the histogram similarity. This method first extracts the histograms of the video frames, and then calculates the distance between the histograms. When the distance is more than threshold, then it is referred to shot change. There are several kinds of methods to calculate the histogram distance, such as Manhattan distance, Euclidean distance and chi-square distance. Several variants of histogram-based have been proposed in the literature. Lu et al. In [14] employed Singular Value Decomposition (SVD), with Hue Saturation Value (HSV) histogram, to propose a low computational complexity SBD scheme. The candidate segment selection using adaptive threshold is implemented, The color histograms are extracted in HSV (Hue-Saturation-Value) space from all frames in each candidate segment, forming a frame feature matrix. The SVD is then performed on the frame feature matrices of all candidate segments to reduce the feature dimension. Bendraou Youssef et al. In [15] formulated a new approach for detecting both hard (CT) and gradual (GT) transitions. They proposed approach processes the video segment by segment, is composed of two main parts: static segment verification (A candidate segment that have not a transition) and shot transition identification (A candidate segment that may contain a transition CT or GT). Features are extracted from the Concatenated Block Based Histograms (CBBH). For each non static segment all frames in each this segment, forming a frame feature matrix. The economy SVD is then performed on feature matrix. An adaptive double thresholding process was employed for detecting the hard cuts. For gradual transitions detection, the folding in technique, known as SVD-updating, is used for the first time in video shot boundary detection. Hong Shao et al. In [16] Hue Saturation Value (HSV) color histogram and Histogram of Gradient (HOG) features are exploited to detect cut transition. HSV color histogram is used to detect the difference between two adjacent frames. While HOG feature is adopted for secondary detection to improve the algorithm performance.

The study confirmed that histogram difference is less sensitive to object motion than the pair-wise comparison, since it ignores the spatial changes in a frame. However, histograms

may also produce missed shots when two frames with similar histograms share a different content.

C. Edge-Based Methods

Another choice for characterizing an image is its edge information. An edge is the boundary between an object and the background, and indicates the boundary between overlapping objects. In edge-based approaches, transition is declared when the locations of the edges of the current frame exhibit a large difference with the edges of the previous frame that have disappeared. For example, Heng et al. In [17] proposed a method based on an edges. They presented the concept of an objects edge by considering the pixels close to the edge. A matching of the edges of an object between two consecutive frames was performed. Then, a transition was declared by utilizing the ratio of the objects edge that was permanent over time and the total number of edges. Zheng et al. In [18] an approach based on a Robert edge detector for detecting fade-in and fade-out transitions was proposed. First, the authors identified the frame edges by comparing gradients with a fixed threshold. Second, they determined the total number of edges that appeared. When a frame without edges occurred, fade in or fade out was declared.

The advantage of this feature is that it is sufficiently invariant to illumination changes and several types of motion, and is related to the human visual perception of a scene. Its main disadvantage is computational cost, and noise sensitivity.

D. Motion-Based Methods

Motion is a key feature in videos and forms an integral part of it. Because shots with camera motion can be incorrectly classified as gradual transitions, detecting zooms and pans increases the accuracy of a shot boundary detection algorithm. Bruno et al. In [19] proposed a linear motion prediction method based on wavelet coefficients, which were computed directly from two successive frames.

For an accurate motion estimation, each block should be matched with all blocks of the next frame, which lead to a large and unreasonable computational cost.

E. Deep Learning-Based Methods

Recently, employing deep learning algorithms in the field of computer vision received much attention from academics. Convolutional Neural Networks (CNN) is one of the most important deep learning algorithms due to its significant abilities to extract high level features from images and video frames [20].

Tong et al. In [21] used The CNN model to extract high-level interpretable features from the frames. It is capable of detecting both CT and GT boundaries. An adaptive threshold process was employed as a preprocessing stage to select candidate segments. Taken one frame as input, the output of the network is a probability distribution among 1000 classes. The five classes with the highest probabilities are selected as the high-level features of the frame and called as the TAGs of the frame for simplicity. However, in some cases when changes

in GT are small and the background is similar, semantics do not change at all thus they cannot achieve high detection of accuracy.

Jingwei Xu et al. In [22] use convolutional neural networks (CNNs) to extract typical features of frames. They adopted a candidate segment selection method to locate the positions of shot boundaries coarsely using adaptive thresholds and eliminate most non-boundary frames. Cut and gradual transitions can be obtained by using a novel pattern-matching method based on a new similarity strategy which is partially inspired by [14].

Hassanien et al. In [23] presented a shot boundary detection method on huge video data set based on spatial-temporal CNN. The Technique is named DeepSBD network that takes a segments of fixed length as input and classify it into 3 categories (cut, gradual, no transition), its output is fed through SVM classifier. This gives the first labeling estimate. Consecutive segments with the same labeling are merged and the result is passed to a post-processing step. The step reduce false alarms of gradual transitions through a histogram-driven temporal differential measurement. However, the C3D ConvNet is more complex than 2D ConvNet, which requires much computation resources and the lengths of gradual transitions are varying but DeepSBD is not designed for multi-scale detection.

Michael Gygli et al. In [24] proposed to learn shot detection, from pixels to final shot boundaries. A fully convolutional neural network has been used for shot boundary detection task. For training this model, They consider the all shot boundaries are generated. Thus, they created a dataset with one million frames and automatically generated transitions such as cuts, dissolves and fades. They considered this work as a binary classification problem to correctly predict if a frame is part of the same shot as the previous frame or not. Their method obtains state-of-the-art results on the RAI data set, while running at an unprecedented speed of more than 120x real-time. Currently, their model makes three main errors, (i) missing long dissolves, which it was not trained with, (ii) partial scene changes and (iii) fast scenes with motion blur.

Shitao Tang et al. In [25] presented a new cascade framework, a fast and accurate approach for shot boundary detection. The first stage applied adaptive thresholding to initially filter the whole video and selects the candidate segments for acceleration. In the second stage, they used a well designed 2D ConvNet learning the similarity function between two images to locate the cut transitions. The third stage utilized a novel C3D ConvNet model to locate positions of gradual transitions.

Lifang Wu et al. In [26] presented a two stage method for shot boundary detection (TSSBD) which distinguishes cut shot by fusing color histogram (HSV) and deep features (CNN) where divide the complete video into segments containing gradual transitions, and over these video segments, gradual shot change detection is implemented using 3D-convolutional neural network, which classifies clips into specific gradual shot change types with a majority voting strategy, gap filling conducts to effectively distinguish shot types of frames and locate shot boundaries.

Rui Liang et al. In [27] proposed a new video shot boundary detection method based on CNN feature. The method extracts the features using the AlexNet and ResNet-152 model for each frame, and calculate cosine similarity to describe the similarity of a pair of frames. For cut boundary detection, they used the similarity of local frames to get more accuracy, and proposed dual-threshold sliding window for gradual transition detection.

Lifang Wu et al. In [28] proposed a method for shot boundary detection with spatial-temporal convolutional neural networks based gradual shot detection and histogram base shot filtering. The cut shots are extracted from the whole video with histogram base shot filtering. Then, C3D deep model is constructed to extract features of frames and distinguish shot types of dissolve, swipe, fade in and fade out, and normal. For untrimmed videos, a frame level merging strategy is constructed to help locate the boundary of shots from neighboring frames.

However, those methods only using the CNN for feature extraction and then using traditional classifiers to detect the scene change. Recently, with the development and popularity of deep learning, many efficient networks for various of applications have been proposed. For example, the deep learning model Res-Net based networks can obtain very high accuracy in image classification and object detection for many large scale image data sets. Therefore, it can be adopted to solve the issue of shot change detection. The downside of this method is revolve around the need for large annotated data-sets. However, that the real data can contain cuts between shots of the same scene which rarely occur in the synthetic data sets due to the nature how they are generated.

F. Others approaches

Thounaojam et al. In [29] proposed a shot detection approach based on genetic algorithm (GA) and fuzzy logic. Fuzzy system is used to classify the video frames into different types of transitions (cut and gradual). Color Histogram Difference is used for feature extraction and for finding the differences between two consecutive frames in a video. GA is used as optimizer to find the optimal range of values of the fuzzy membership functions. The result shows that the combination of this feature is efficient and the accuracy increases with increase in iterations/generations of GA.

Jialei Bi et al. In [30] proposed a novel cut detection method based on information theory using SVM. They first compute the dissimilarity using information theory and construct a discriminative feature vector based on mutual information. Then a support vector machine is trained to classify the frames as cut or none-cut frames without using a traditional global or adaptive threshold.

Junaid Baber et al. In [31] the proposed method, shot boundaries are extracted from videos using frame entropy and SURF descriptors. Cut boundaries were detected by difference of entropy of the gray scale intensity in adjacent frames. And fade boundaries were detected indiscriminately based on temporal changes in the entropy of the pixel intensity

across each images. Then the false detection can be eliminated effectively by using local descriptors SURF.

Sawitchaya Tippaya et al. In [32] proposed a multi-modal visual features based SBD framework. They adopted a candidate segment selection that performs without the threshold calculation. The discontinuity signal is calculated based on the SURF matching score and RGB histogram cosine distance value.

Finally, In TABLE I demonstrates a comparison among different SBD algorithms based on features employed, frame skipping, data-set used, accuracy (precision, recall and F1 score measures). From the table, it can be observed that the algorithms used frame skipping technique have low computational cost with an acceptable accuracy as in [14]. Although some algorithms utilize frame skipping, they show a moderate computational cost because of the computation complexity of the features used such as SURF in [32]. Obviously, CNN-based SBD algorithms that show a high computational cost such as [27, 28, 29, 32, 36] gain a remarkable accuracy compared to other algorithms.

IV. SHOT BOUNDARY DETECTION EVALUATION METRICS

There are two prospective metrics that need to be used to evaluate the performance of SBD algorithms. These two aspects are the accuracy and the computational complexity. Usually improving one aspect would be on the cost of the other one. Also, for the evaluation to be truly representative and reliable for comparing various techniques, it must be done in similar conditions and with very similar data sets. In this section, we discuss the common metrics (recall, precision, and F1-score) of measuring the accuracy and the computational complexity [33].

- 1) **Precision** : It is the ratio of detection of correct experimental to the detection of correct and false.

$$precision = \frac{N_c}{N_c + N_f} \quad (1)$$

- 2) **Recall** : It is the ratio of detection of correct experimental to the detection of correct and missed.

$$recall = \frac{N_c}{N_c + N_m} \quad (2)$$

- 3) **F1 score** : It combines precision and recall to achieve one score. It is varies in the range [0, 1] where a score of 1 indicates the best efficacy of a system.

$$F1 = \frac{2 \times recall \times precision}{recall + precision} \quad (3)$$

Where, N_c is number of transitions correctly reported, N_m is number of transitions missed to be reported, and N_f is number of falsely reported transitions.

V. OPEN CHALLENGES

Although a large amount of work has been done in shot boundary detection, many issues are still open and deserve further research. We can conclude from this state of art that a good video shot detection method highly depends on

TABLE I: Comparison of different state-of-the-art SBD algorithm

Ref	Methods	Dataset	CT			GT		
			P	R	F1	P	R	F1
[14]	SVD and HSV Histogram-Based	TRECVID 2001	0.91	0.85	0.88	0.83	0.81	0.81
[15]	Histogram-Based	TRECVID 2001	0.97	0.95	0.96	0.87	0.93	0.90
[21]	Deep Learning-Based (CNN)	TRECVID 2001	0.99	0.87	0.92	0.87	0.83	0.87
[22]	Deep Learning-Based (CNN)	TRECVID 2001	1.00	0.98	0.99	0.99	0.95	0.97
[23]	Deep Learning-Based (CNN)	UCF101-SBD	0.98	1.00	0.99	0.99	0.99	0.99
[25]	Deep Learning-Based (CNN)	TRECVID 2007	0.98	1.00	0.99	0.84	0.84	0.84
[27]	Deep Learning-Based (CNN)	Other	0.95	0.97	0.96	0.86	0.91	0.87
[29]	Genetic Algorithm (GA) and Fuzzy Logic based	TRECVID 2001	0.88	0.92	0.90	0.86	0.78	0.82
[30]	Theory Information and SVM based	TRECVID 2002	0.98	0.97	0.98	-	-	-
[32]	SURF matching score and RGB histogram based	Golf Video	1.00	0.98	0.99	0.89	0.81	0.85

features, similarity measure and thresholds used. We found that the major challenges to detection techniques are by illumination changes, object and camera motion. For example color histograms are robust to small camera motion, but they are not able to differentiate the shots within the same scene, and they are sensitive to large camera motions. Edge features are more invariant to illumination changes and motion than color histograms, and motion features can effectively handle the influence of object and camera motion. If we just use a kind of feature to detect the shot boundary, the result may not be satisfactory, but if we use many kinds of features, the speed will be slow. And the major challenge is the problem of determining an automatic threshold based on the characteristics of the video. The difficulty is how to choose the optimal threshold. However, the efforts to replace thresholding by machine learning have begun only recently. The importation of these ideas may be novel drives to the advance of SBD.

VI. CONCLUSION AND FUTURE SCOPE

Video shot boundary detection is the first step of video processing, it is also the most important step. There have been a lot of studies about shot boundary at present. In this work, a comprehensive survey of SBD algorithms (or shot boundary detection algorithms) was performed. Video definitions, transition types, and hierarchies were demonstrated. The different techniques are discussed to detect a shot boundary depending upon the contents and the change in contents of video. Despite the extensive research on concrete SBD techniques, SBD still have some problems that are relevant in practice for different video scenarios which need to be studied. These challenges are represented by: Sudden illuminance changes, dim lighting frames, comparable background frames, object and camera motion, and change in small regions. Solving these challenges will surely improve the performance of SBD algorithms. Finally, the machine learning approaches have been popular and received much attention in the field of computer vision applications. However, in the field of SBD, the efforts to replace thresholding by machine learning have begun only recently. But the amount of research carried out in the domain of SBD using machine learning is quite less. Exploring the benefit of the new machine learning technologies such as

deep learning approaches for SBD could be directed as new directions for the future.

Usually, in the sequential case, the comparison of the frames and shot boundary detection sounds simple, but it can take centuries to processes multimedia big data. Performance in a lengthy video data remains an open area of research. Our future work is to focus on deep learning approaches for SBD by used technologies of analyses multimedia big data.

REFERENCES

- [1] Deepika Bajaj and Shanu Sharma. Video depiction of key frames- a review. In *Proceedings of the Sixth International Conference on Computer and Communication Technology 2015, ICCCT '15*, pages 183–187, New York, NY, USA, 2015. ACM.
- [2] Weiming Hu, Nianhua Xie, Li Li, Xianglin Zeng, and S. Maybank. A survey on visual content-based video indexing and retrieval. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 41(6):797–819, nov 2011.
- [3] Tiecheng Liu and John R. Kender. Computational approaches to temporal sampling of video sequences. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 3(2):7–es, may 2007.
- [4] Remi Trichet, Ramakant Nevatia, and Brian Burns. Video event classification with temporal partitioning. In *2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE, aug 2015.
- [5] Shayok Chakraborty, Omesh Tickoo, and Ravi Iyer. Adaptive keyframe selection for video summarization. In *2015 IEEE Winter Conference on Applications of Computer Vision*. IEEE, jan 2015.
- [6] Youssef Bendraou. *Video shot boundary detection and key-frame extraction using mathematical models*. Theses, Université du Littoral Côte d’Opale, November 2017.
- [7] Jaydeb Mondal, Malay Kumar Kundu, Sudeb Das, and Manish Chowdhury. Video shot boundary detection using multiscale geometric analysis of nsct and least squares support vector machine. *Multimedia Tools and Applications*, 77(7):8139–8161, apr 2017.

- [8] Gautam Pal, Dwijen Rudrapaul, Suvojit Acharjee, Ruben Ray, Sayan Chakraborty, and Nilanjan Dey. Video shot boundary detection: A review. In *Advances in Intelligent Systems and Computing*, pages 119–127. Springer International Publishing, 2015.
- [9] A. Hanjalic. Shot-boundary detection: unraveled and resolved? *IEEE Transactions on Circuits and Systems for Video Technology*, 12(2):90–105, 2002.
- [10] Hrishikesh Bhaumik, Siddhartha Bhattacharyya, and Susanta Chakraborty. Content coverage and redundancy removal in video summarization. In *Intelligent Analysis of Multimedia Information*, pages 352–374. IGI Global.
- [11] Guangyu Gao and Huadong Ma. To accelerate shot boundary detection by reducing detection region and scope. *Multimedia Tools and Applications*, 71(3):1749–1770, Springer Science and Business Media, dec 2012.
- [12] Zhonglan Wu and Pin Xu. Shot boundary detection in video retrieval. In *2013 IEEE 4th International Conference on Electronics Information and Emergency Communication*. IEEE, nov 2013.
- [13] Heba Ahmed Elnemr, Nourhan Mohamed Zayed, and Mahmoud Abdelmoneim Fakhreldein. Feature extraction techniques. In *Handbook of Research on Emerging Perspectives in Intelligent Pattern Recognition, Analysis, and Image Processing*, pages 264–294. IGI Global, 2016.
- [14] Zhe-Ming Lu and Yong Shi. Fast video shot boundary detection based on SVD and pattern matching. *IEEE Transactions on Image Processing*, 22(12):5136–5145, dec 2013.
- [15] Bendraou Youssef, Essannouni Fedwa, Aboutajdine Driss, and Salam Ahmed. Shot boundary detection via adaptive low rank and svd-updating. *Computer Vision and Image Understanding*, 161:20–28, aug 2017.
- [16] Hong Shao, Yang Qu, and Wencheng Cui. Shot boundary detection algorithm based on HSV histogram and HOG feature. In *Proceedings of the 2015 International Conference on Advanced Engineering Materials and Technology*, pages 951–957. Atlantis Press, 2015.
- [17] Wei Jyh Heng and King N. Ngan. An object-based shot boundary detection using edge tracing and tracking. *Journal of Visual Communication and Image Representation*, 12(3):217–239, sep 2001.
- [18] Jie Zheng, Fengmei Zou, and Mandel Shi. An efficient algorithm for video shot boundary detection. In *Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, 2004*. IEEE.
- [19] E. Bruno and D. Pellerin. Video shot detection based on linear prediction of motion. In *Proceedings. IEEE International Conference on Multimedia and Expo 2002*, volume 1, pages 289–292. IEEE, 2002.
- [20] Eralda Nishani and Betim Cico. Computer vision approaches based on deep learning and neural networks: Deep neural networks for video analysis of human pose estimation. In *2017 6th Mediterranean Conference on Embedded Computing (MECO)*. IEEE, jun 2017.
- [21] Wenjing Tong, Li Song, Xiaokang Yang, Hui Qu, and Rong Xie. Cnn-based shot boundary detection and video annotation. In *2015 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*. IEEE, jun 2015.
- [22] Jingwei Xu, Li Song, and Rong Xie. Shot boundary detection using convolutional neural networks. In *2016 Visual Communications and Image Processing (VCIP)*. IEEE, nov 2016.
- [23] Ahmed Hassanien, Mohamed A. Elgharib, Ahmed Selim, Mohamed Hefeeda, and Wojciech Matusik. Large-scale, fast and accurate shot boundary detection through spatio-temporal convolutional neural networks. *CoRR*, abs/1705.03281, 2017.
- [24] Michael Gygli. Ridiculously fast shot boundary detection with fully convolutional neural networks. In *2018 International Conference on Content-Based Multimedia Indexing (CBMI)*. IEEE, sep 2018.
- [25] Shitao Tang, Litong Feng, Zhanghui Kuang, Yimin Chen, and Wei Zhang. Fast video shot transition localization with deep structured models. In *Computer Vision – ACCV 2018*, pages 577–592, Cham, 2019. Springer International Publishing.
- [26] Lifang Wu, Shuai Zhang, Meng Jian, Zhe Lu, and Dong Wang. Two stage shot boundary detection via feature fusion and spatial-temporal convolutional neural networks. *IEEE Access*, 7:77268–77276, 2019.
- [27] Rui Liang, Qingxin Zhu, Honglei Wei, and Shujiao Liao. A video shot boundary detection approach based on CNN feature. In *2017 IEEE International Symposium on Multimedia (ISM)*. IEEE, dec 2017.
- [28] Lifang Wu, Shuai Zhang, Meng Jian, Zhijia Zhao, and Dong Wang. Shot boundary detection with spatial-temporal convolutional neural networks. In *Pattern Recognition and Computer Vision*, pages 479–491. Springer International Publishing, 2018.
- [29] Dalton Meitei Thounaojam, Thongam Khelchandra, Kh. Manglem Singh, and Sudipta Roy. A genetic algorithm and fuzzy logic approach for video shot boundary detection. *Computational Intelligence and Neuroscience*, 2016:1–11, 2016.
- [30] Jialei Bi, Xianglong Liu, and Bo Lang. A novel shot boundary detection based on information theory using SVM. In *2011 4th International Congress on Image and Signal Processing*. IEEE, oct 2011.
- [31] Junaid Baber, Nitin Afzulpurkar, Matthew N. Dailey, and Maheen Bakhtyar. Shot boundary detection from videos using entropy and local descriptor. In *2011 17th International Conference on Digital Signal Processing (DSP)*. IEEE, jul 2011.
- [32] Sawitchaya Tippaya, Suchada Sitjongsatoporn, Tele Tan, Masood Mehmood Khan, and Kosin Chamnongthai. Multi-modal visual features-based video shot boundary detection. *IEEE Access*, 5:12563–12575, 2017.
- [33] Amr Ahmed. Video representation and processing for multimedia data mining. In *Semantic Mining Technologies for Multimedia Databases*. IGI Global, 2009.