Al-Rafidain Journal of Computer Sciences and Mathematics (RJCSM)

www.csmj.uomosul.edu.iq

The Performance of Different Video Segmentation or Video Shot Boundary Detection Techniques: A Survey

Rafal Ali Sameer

Collage of Science, Computer Department, University of University of Baghdad, University of Baghdad, Iraq
Email: rafal.a1@sc.uobaghdad.edu.iq

Article information

Article history:
Received 22 January ,2025
Revised 19 May ,2025
Accepted 31 May ,2025
Published 26 June ,2025

Keywords:
Video Segmentation,
Video Shot Boundary Detection Techniques,
Video Transition Effects.

Correspondence:
Rafal Ali Sameer
Email:
rafal.a1@sc.uobaghdad.edu.iq

Abstract

Video is a way to convey information on the internet which is more popular than the text. Video processing expanded with the using of videos in various fields and highly developing of communication technologies. Video shot boundary detection technique is one of the major research areas and first step for video processing like browsing, video retrieval, and indexing. The process of identifying the video shots is called video segmentation or video shot boundary detection. Any video created by merging number of shots between these shots the visual effects will be added to create smooth transition between shots. There are different types of visual effects added between shots of video transitions. The main visual effects are cut or gradual transitions while the other transitions derived from these two types. The aim of shot boundary detection techniques to detect the shot boundaries and transitions between successive shots. There are various proposed techniques for different types of videos in this domain. Almost the shot segmentation techniques based on the type of the video transition or visual effects. This paper presents a wide range of researches with various techniques used for shot boundary detection and their performance including their advantages, disadvantages, limitations, and future works.

DOI: 10.33899/csmj.2025.156866.1165, ©Authors, 2025, College of Computer Science and Mathematics, University of Mosul, Iraq.
This is an open access article under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0).

1. Introduction

Video become an important subject in various fields including personal and professional applications such as education, medicine, security applications, databases, entertainment and even low-bandwidth wireless applications, therefore the use of the video becoming increasingly popular. A video may be a movie, a news clip, a documentary, or a traffic surveillance clip [1]. Video is an advanced multimedia that has been enabled by the availability of the internet in the communication field [2]. The techniques of video processing are based on the feature of its content (objects and semantic) [1]. Video consists of audio synchronized with a sequence of frames while the video structure can be defined by a set of stories, scenes, shots, and frames. The shot of video is a group of connected sequential frames (frame is an image synchronized in time). Shots can be considered as the smallest indexing unit where no changes in scene content can be perceived and higher level concepts are often constructed by combining and analyzing inter and intra shot relationships. Video Shot Boundary Detection (VSBD) or video temporal segmentation is the fundamental process towards video processing, summarization, indexing, browsing, and retrieval. A successful video temporal segmentation can lead into a better shot identification [2][3][4]. Video has visual information varying with time (or moving visual information) [5]. The visual information computed from video frames is used by many techniques to segment the video and even to detect and recognize objects and events. Some of the most popular visual features are: color, shape, movement, and texture that extracted and descripted by global or local feature extraction and description technique [4]. In general, the features defined by performing some calculations on the image, these calculations identify many of the informative data for the image's objects. The term of feature detection and description is the process of defining the interested points in an image that gives the image's contents description [6][7]. In the last decade, many researchers contributed the shot boundary detection (SBD) for different types of videos. Recently, a great number of SBD strategies have been presented, and all of them can be categorized based on a few key ideas that underlie the various detection schemes [8][9]. The aim of this work is giving a comprehensive overview about the types of video transitions, the techniques for SBD, and the most recent works for SBD with their performance.

The rest of the paper organized as follows, section two presents the works that have similar objective to this work. Section three and section four present theoretical concepts of video segmentation or video shot boundary detection and visual effects of video respectively. Section five presents a survey on the techniques of SBD, their outcomes, and their performance. Section six presents a conclusion of this work

2. Related works

Various reviews for VSBD techniques have been existed before this review. The previous reviews have the same objective but with different ideas and content, which are as follow:

In 1996, John S. Boreczky and Lawrence A. Rowe, present a comparison for several classification approaches and SBD including discrete cosine transform, block matching, histograms, and motion vector techniques. Based on an extensive range of video sequences with a decent mix of transition types, the effectiveness and simplicity of choosing acceptable thresholds for various algorithms are assessed [10]. In 2001, Rainer Lienhart, presents a survey emphasizes various detection methods for the three most popular effects of video transition: dissolves, cuts, and fades. The most widely used approaches are thoroughly tested and show its detail, in the other hand the other methods are listed only. In the literature whenever they are stated, they provide reliable performance. The guidelines also provided for video processing [9]. In 2006, Costas Cotsaces, et.al, present a review focuses on SBD and condensed video representation (also called summarization and abstraction). An overview of the fundamental issues in each task is provided, and recent work on the subject is described and is critically reviewed [11]. In 2013, Salim A. Chavan and Sudhir G. Akojwar, discuss various algorithms used for gradual shot change detection. Performance measures such as precision and recall, for testing different algorithms are also considered [12]. In 2013, Saranya K and Kethsy Prabavathy, presented a review on recent developments in visual automatic abrupt SBD technique which detects cut transitions. The state of the art of existing approaches in each major issue has been described with the focus on the following tasks: video analysis including Gabor filtering, fuzzy color histogram, robust video detection, comparison of VSBD techniques and CBIR system [13]. In 2014, Mohini Deokar and Ruhi Kabra, provide a SBD classification method that takes into account techniques for gradual shot transitions [14]. In 2014, Nikita Sao and Ravi Mishra, present a brief literature survey that depicts the SBD techniques done till date. The newly developed method seems to produce good result and also give an idea about its capacity to detect many other transitions [15]. In 2014, Dalton Meitei Thounaojam, et.al, present a short survey on MPEG-compressed videos segmentation [16]. In 2015, Gautam Pal, et.al, established an overview on a few previously proposed algorithms. The paper also includes histogram based, DCT based and motion vector based algorithms as well as their advantages and their limitations [17]. In 2016, Mikhal Rakshaskar and Salim Chavan, present a survey of various novel algorithm for detecting fade-in and fade-out used by renowned personals with different methods. This survey also emphasizes on different core concepts underlying the different detection schemes for the most used video transition effect: fades [18]. In 2017, Israa Hadi and Hikmat Z. Neima, present an extend study of SBD techniques. This study classifies SBD techniques based on the level they work on and most recent methods are investigated as well [19]. In 2018, Sadiq H. Abdulhussain, et.al, present an overview of a large collection of SBD techniques and their advancement. Each strategy's benefits and drawbacks are thoroughly investigated. The created algorithms are described, as well as the difficulties and suggestions [20]. In 2019, Noraida Haji Ali and Fadilah Harun, focus on employing SBD methods to find hacking attempts and video modifications. This study developed a new model for detecting video structures utilizing SBD and the histograms feature, which is important for creating potent tools for identifying video originality and locating tampering attempts. Comparison made between pixel differences, statistical differences, histogram differences showed that histogram techniques were a good deal between accuracy and speed to handle video structure. Sometimes other techniques also combined with histogram like statistical measurements features for SBD. Another often used SBD technique is motion analysis between two successive frames [21]. In 2020, Pragati Ashok Kevadkar and Jadhav dattatray present a review for different approaches of VSBD [22]. In 2020, Benoughidene Abdel halim and Titouna Faiza, present the fundamental theory, a brief overview, and the development of the video SBD approaches. Each approach's benefits and drawbacks are thoroughly examined, and problems are put forth. The machine learning technologies for SBD also focused like methods to deep learning that might be used as new paths in the future [23]. In 2020, Shrikant P. Chavate, presents a review of different approaches for SBD [24]. In 2020, M. Raja Suguna and A. Kalaivani, present a thorough investigation of shot boundary segmentation approaches using deep learning and traditional machine learning methods. This work reviews a number of proposed SBD algorithms employing various methods, including motion vector, histograms, edges, pixel differences, deep learning, and statistics techniques. Every technique has its own benefits and drawbacks. Camera movements, changes in camera settings like magnification, changes in flash settings, and no changed background are the current problems that have an impact on the correctness of the SBD algorithm. In order to overcome these difficulties, deep learning algorithms that can handle vast amounts of photos can be researched [25]. In 2020, C. Victoria Priscilla1 and D. Rajeshwari, reveal the investigation of SBD extraction feature and key-frame feature extraction techniques. When compared to other methods, the experimental results of SBD over histogram produced the greatest results for key frame extraction. The advantages and disadvantages of key frame extraction reveal its inability to represent a vast video library compared to more conventional techniques. The most recent strategy for overcoming this reality is to identify the video sequence using deep learning techniques, which can quickly process millions of videos. State-of-the-art precision is one of the benefits of employing the deep learning model. In a very large dataset, further research on the deep learning method for the extraction of key frames and SBD can produce more valuable results [26]. In 2021, Shrikant Chavate, et.al present a survey of different methodologies applied to perform SBD along with the performance criteria. This contains various preprocessing methods, methods for extracting features, methods for computing similarity, etc. The results of various strategies are framed in terms of comparison, calculation speed, accuracy, F1-score, recall, and precision [8]. In 2022, Swati Hadke and Ravi Mishra, present a review of various SBD methods along with the performance criteria [27].

3. Video Segmentation or Video Shot Boundary Detection (VSBD)

A video is a spatial-intensity pattern that varies with time. A video is represented by a time sequence of still images or frames. The frames are the basic unit of any video which displayed at a constant rate. The frame rate is measured in terms of the number of frames displayed per second or Hertz (Hz). In general extracting frames from video will produce 25-30 frames per second. Increasing number of frames will create smooth move for illusion. The size of the video file grows as the number of frames increases. In addition to visual information, videos can also include text and audio data that can be synced to specific frame orders. These details establish a video as a sophisticated data container [1][4][5]. Typically, the video structure is a set of scenes, shots, and frames see **Figure 1**. Scene describes the video story and may include a single shot or several shots which similar by some features and typically appear around the same time. The shot represents a spatio-temporal frame sequence which presents continued action captured by one camera where camera and object motion allowed and the keyframes can be used to represent the visual information. While a shot represents a physical video unit, a scene represents a semantic video unit. Keyframe is the frame that represents a salient visual content of a shot. SBD depends on the detection of transition-related visual dissimilarity. The discrepancy between two frames is typically discovered during shot change, this difference defines itself in several ways that can be divided into two categories: abrupt (such as a hard cut) and gradual (such as a fade-in, fade-out, dissolve, or wipe) [3][4].

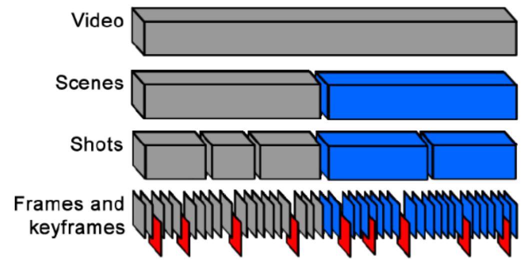

Figure 1. The structure of video [4].

Segmentation of shots is typically classified into compressed domain and non-compressed domain shot segmentation. A SBD algorithms depended on a compressed domain have less information of features but don't require for decompression process of video. The fundamental idea behind SBD based on non-compressed domain is to directly extract shape, texture, color, and other fundamental data from the video frame [28]. The main methods for SBD can be divided to [11]:

3.1 Pixel Comparison

Pixel Comparison is a very primal method used for detecting frame similarities or frame changes. Using this technique, the variations in pixel intensity between neighboring frames are counted to determine differences between the frames when intensity of pixels exceeds the pre-determined threshold value then shot change is declared, see **Figure 2.**The pixel comparison method is very sensitive to zooming, quick object and camera movements [14][15][17][25][29]. The idea of this procedure applied in [30][31]:

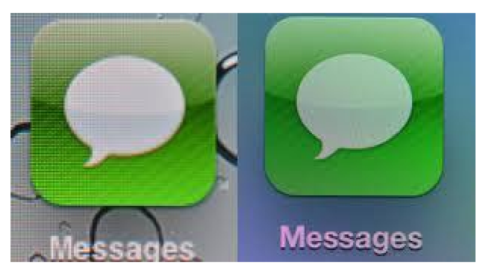

Figure 2. Pixel difference [15].

3.2 Statistical Comparison

While the pixel comparison method for transition detection is not very robust because the camera effects result in false transition detection, an expansion method have been produced. Techniques for statistical comparison Block the frames, and then compute the standard deviation and mean for each block. In general, compare the standard deviation and mean for the blocks in succeeding frames, when particular blocks number exceeds a predetermined value of threshold then defined the shot change. In general, this method outstanding the pixel comparison method despite the fact that more complex versions can be costly in terms of computation. Noise is not tolerated by this strategy but it moves slowly due to statistical computation

[14][15][17][25][29]. See **Figure 3**.

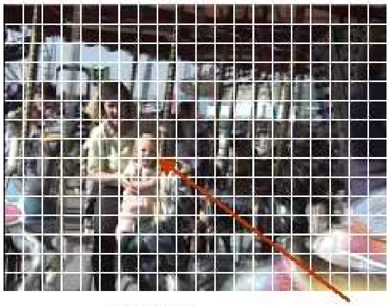

Figure 3. Statistical difference [15].

3.3 Histogram Comparison

Perhaps the most popular technique for identifying shot boundaries is using histograms. The image's histogram defines the pixel intensities frequency graphically, where the horizontal-axis represents the range values of pixel intensity and the vertical-axis represents the values of pixel intensity frequency. The value of the pixel intensity frequency represented by bins in a histogram. A basic histogram technique creates gray or color histogram for the two consecutive video frames, and calculates the variance between bins. The shot boundary detection defined when the variance is more than predefined threshold. In general, histograms are much faster than analogous statistical methods, more effective than basic pixel comparisons, easier to calculate, and invariant to rotation, translation, and zooming. This technique is typically used in conjunction with a gray histogram to detect shot transitions, although it has been proven to be susceptible to variation in flash light, noise, object motion, and camera motion [14][15][17][22][25][29]. The idea of this procedure applied in [28][32][33][34]. See **Figure 4**.

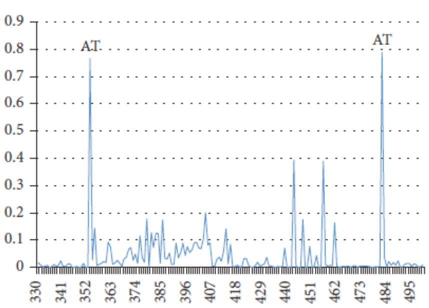

Figure 4. Histogram difference [35].

3.4 Transform-Based Difference (Compression Features)

These techniques represent compression difference computation using different transformation methods, such as Discrete Cosine Transformation (DCT), Discrete Fourier Transform (DFT), and Discrete Wavelet Transform (DWT) which are used to extract the texture features from the frame. In order to identify shot boundaries, transform methods rely on the characteristics of the encoding standard. The transform methods are fast and easy to calculate because these methods not necessary to fully decode the video stream resulting in a significant reduction in the amount of data to be processed but the only disadvantage of these features are variant to zooming [15][22]. The idea of this procedure implemented in [36][37][38].

3.5 Edge-Based Difference (Edge detection and tracking)

In order to determine whether any new edges have entered the image or if some old edges have vanished, the edges of successive aligned frames are first recognized, the edge pixels are then paired with nearby edge pixels in the other image. This requires creating an edge image of each frame pair using a filtering technique, such as Sobel filtering, as in **Figure 5**. The edge photos are compared and assessed after that. The edge can be defined more precisely by replacing Sobel with Canny (more robust filter), especially in scenes that so dark or bright. While not outperforming the straightforward histogram-based method, edge-based solutions are computationally more expensive. The false detection in both gradual and cut video transition can be eliminated by this method especially with the existence of flash light. If there is a noticeable difference between the edge's position in the current frame and its position in the previous frame, a shot transition has occurred [14][15][17][25][29]. The idea of this procedure applied in [31][39].

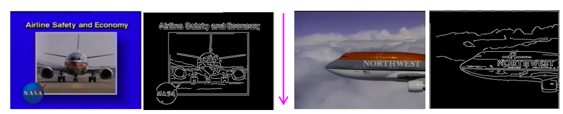

Figure 5. Transition detection based on edge variation [40].

3.6 Block Matching (or Motion Vector) method

This method divides each video frame into a predetermined number of blocks, from which motion vectors are then extracted. This method compares each and every block in the following video frames to the block in the current video frame. Any camera movement or shot transition will be picked up if the block variances are greater than the specified threshold value. Gradual video transition can be thought of as the movement of the camera within a shot. With the incorrect choice of motion vectors, shot detection accuracy will affected [14][25][29][41]. The idea of Motion vector implemented in [42].

3.7 Deep Learning Methods (DL)

Because deep learning approaches automatically extract features from images and videos, their implementation in computer vision received much interest. Convolutional neural networks can analyse a huge number of photos more quickly and extract information from the images more quickly [25].

With video processing the deep learning methods can be grouped as Deep Neural Network (DNN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Hybrid-method. CNN in video process the motion locally by using the architecture with many resolution including low and high image resolution. DNN is an improved approach for Neural Network (NN) but with more complexity. RNN is used in general for data with time series while video content is spatio-temporal therefore many RNN approaches for video processing are improved recently. The hybrid-method for video processing combines many DL approaches [43]. The idea of CNN implemented in [44][45][46][47].

4 Visual Effects in Videos

Shooting and edition operations are the two main processes for video production. The process of generating various shots that construct a video defined as the shooting operation, while the process of creating the final video structure (joining shots together) by adding different visual effects and providing smooth transitions between the shots is defined as the edition operation. There are two visual effects that can occur between shots which are: cuts (discontinuous or abrupt) shot transitions and gradual (continuous) shot transitions (fades or dissolves) [4].

4.1 Cut (Sharp, Discontinuous, or Abrupt)

Cut shot transitions is an instantaneous change of visual content between consecutive shots. This transition doesn't need any complicated impact and easy to identify [3][4][11].

4.2 Gradual (Continuous)

Gradual shot transitions: the two neighbor shots will have spread the borders by making new intermediate frame; therefore, this transition will be defined by a beginning and a termination frame. Fade and dissolve transitions are types of continuous transition.

4.2.1 Fade

Fade transition has two types, Fade-in and Fade-out. Fade-in is occurring when an image displayed from the black image (a shot gradually appears from a constant image, and the fade-in effects defined by showing the shot gradually), while fade-out appears when images faded to a black screen (the shot progressively disappears from a fixed image, and the fade-out effects is described by shot darkening gradually till the end frame reach black entirely [4][11].

4.2.2 Dissolve

Dissolve transition points to process of dimming the first image and brightening the second image in the shot progressively, making one frame overlies the other frames without null interval. Dissolve is the asynchronous occurrence of fade-in and fades-out. These two effects are layered for a fixed period, fade-out of the present shot and fade-in the next one. The last few frames of the disappearing shot temporally overlap with the first few frames of the appearing shot. During the overlap, the intensity of the disappearing shot decreases from normal to zero (fade-out), while that of the appearing shot increases from zero to normal (fade-in) [3][4][11][22]. Because it is possible to clearly see the visual discontinuity, abrupt transitions are simple to identify. However, smooth transitions can be hard to detect because there is little change in the visual content of the frames during such transitions. When shot boundaries are determined by identifying dissolve-type shot transitions, the video is over-segmented and create many smaller video sequences (almost belong to the same shot) are mistakenly identified as separate shots [48]. See **Figure 6**. to distinguish between the effects of video transitions.

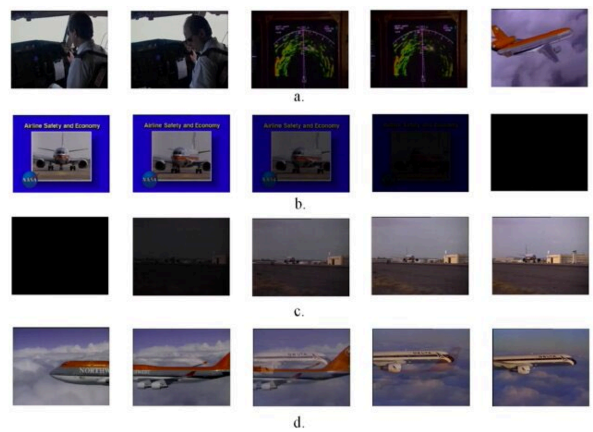

Figure 6. Video transitions (a) cut (b) fade-out (c) fade-in (d) dissolve [40].

5 The Recent Video Segmentation or Video Shot Boundary Detection Techniques

There are various techniques used for different types of SBD. Some of a new VSBD techniques (starting from year 2020) will be described in this section with their performance and some of them have future works:

Technique (1): A novel VSBD method suggested based on the feature fusion and clustering approach (FFCT). The video sequence's interval frames are chosen in this approach, rendered as grayscale images, and scaled via sampling. The SURF and fingerprint features are taken from the noncompressed and compressed domains with the frames, and the taken features are then merged. The merged features are then clustered using the K-means approach. Linear Discriminant Analysis (LDA) is utilized for mapping the classes in order to achieve class cohesiveness and class-level looseness. The features in every class are then chosen by density calculation and matched to realize the coarse detection and fine detection of VSBD once the feature classes' correlation between frames has been computed. The FFCT method takes into account both the global and local aspects of the video's interval frames, resulting in a thorough and precise features extraction and fusion. **Performance:** The experimental results show that the suggested method has high degree of precision and effectiveness by comparing to the most recent representative methods particularly for gradual and long shot transition detection. The verification of both the actual application and the public standard datasets (TRECVID 2001 and 2007) are used to evaluate the performances. The effectiveness of the three methods (HLFPN-KM, FFCT, and SSA-SURF) evaluated by the number of boundary detection and the execution time of the SBD. In contrast to the other two algorithms, the FFCT algorithm's time complexity is significantly lower. Additionally, despite the fact that the HLFPN-KM algorithm is more efficient, the time consumption and total SBD's time for all sequences are both decreased, demonstrating that the FFCT algorithm has a lower time complexity than the HLFPN-KM technique. The proposed method produces F_1 measure about 92.99 and 94.99 for TRECVID2001 and TRECVID2007 respectively which outperform PSOGSA, ORB-SSIM, HLFPN-KM. For future work, it is possible to identify unstructured data matching and retrieval (homogenous/heterogeneous) through a large scale analyzing the content of video [49].

Technique (2): This study uses the Midhinge Local Binary Pattern (MHLBP) to extract texture information from the frames of video and identify abrupt video transitions. MHLBP's discrimination abilities surpass those of midrange LBP and mean LBP, as well as basic LBP. By using midhinge statistics on every mask of a video frame, the MHLBP histogram is created. Euclidean distance is used for the distance measurements between the histogram features of the consecutive frames. The measured distances are put through an adaptive threshold technique to detect abrupt video transition. **Performance:** The suggested algorithm evaluated by using recall, precision, and F-measure on video dataset TRECVID 2001 where the obtained outcomes are satisfactory in terms of effectiveness and precision. The suggested technique produced an F-measure of 89%, while the (original, mean, and midrange) LBP variations yielded 78%, 80%, and 84% of F-measure, respectively. The proposed method outperforms the other existing methods according to measurements done on TRECVID 2001 dataset. For future work, the performance of the LBP feature is improved by extracting the MHLBP histogram from every video frame [50].

Technique (3): This approach uses the deep learning technique including Three-dimensional convolutional architecture and training data artificially to increase the precision of SBD. An essential video preprocessing step is the detection of typical shot transition which is done by using a network TransNetV2. A list of a video's shots is put together as follows: When the expected transition's possibility falls below a threshold, a shot begins at the first frame, and when it surpasses the threshold, the shot terminates at the first frame. Because the network is taught to predict one frame only for whatever transition, even with long dissolves, the acceptance threshold was set less than (0.1) rather than the usual (0.5) in all trials because it worked quite well for the majority of the models. **Performance:** On the RAI dataset, the network, which has multiple dilated three dimension convolutional processes per layer, produces state-of-the-art outcomes. Although TransNet V2's deep network for shot transition detection achieves cutting-edge performance, perfect accuracy is still impossible. The proposed work evaluated and compared to other recent SBD approaches. The experimental results show that the suggested algorithm surpasses comparable works on a variety of public benchmarks. Numerous problems with earlier shot detectors are solved, and the training procedure and training data collection are covered in detail. The proposed BERT model has low performance for the dataset in TRECVID 2016 and TRECVID 2018 but it has better results in TRECVID 2017 [51].

Technique (4): This paper presented an approach for VSBD by using CNN that is depended on feature extraction. There are two steps to accomplish this work, initially extract features by using HVS parameters depended on the difference of mean log and implement the histogram distribution function. The CNN will detect the shots that depended on the probability function where the input to CNN is the extracted feature. **Performance:** The proposed work evaluated by using recall, precision, and F1 measures which give high values. The experimental results give good results for cartoon videos according to precision and F1 score while for sport videos the proposed approach work better for all measures. The results of the proposed approach mean the complicated videos have better efficiency according to F1 and precision measures because of sudden changes in features. F_1 measures are between [77.26 to 96.01]. For future work, increase the accuracy of gradual transition by processing motion of objects and camera in videos [46].

Technique (5): A novel approach to SBD for the cut transition based on Convolution Neural Network (CNN) presented in this paper. The proposed work done by using CNN and RAI dataset. This type of preprocessing increases the accuracy and speed for VSBD. Then shots are detected by using semantic mark that created along with shot detection. **Performance:** The efficiency of the proposed SBD method done by using different types of video in TRECVID IACC.3 dataset like news, nature, movie, sports, cartoon, and info-graphic. The proposed method detected correctly the most video transitions as shown by the experiment. The results have evaluated using precision, recall and F1-measure. An average precision is 0.966, recall is 0.864 and F1-score is 0.912. These results prove to be effectiveness and efficiency of the proposed algorithm for cut transition by comparing it with adaptive background algorithm. This method can detect shot boundaries accurately when there are small differences between shot boundaries and there are similar backgrounds for those shots in the other hand other methods fail to detect this type of shot boundary. For future work, the gradual and fade VSBD method can be improved by increasing the accuracy of the training data sets [47].

Technique (6): The proposed method detects abrupt and gradual videos transition where the detection process of shot in videos is done by extract the histogram of Mean Probability Binary Weight (MPBW) from frames that are normalized and Kirsch magnitude as combination between global and local features. Then, to detect gradual video transition the coefficient of variance statistical measure and mean deviation performed on MPBW histograms, while the abrupt video transition detected by applying the adaptive threshold and calculate the distance between histograms. **Performance:** The proposed work evaluated by using datasets TRECVID 2001, 2007. By comparing precision, recall, and F1-score of the proposed method with other state-of-the-art SBD methods the results show an improved SBD method [52].

Technique (7): A self-supervised shot contrastive learning approach (ShotCoL) to learn a shot representation for long-form videos using unlabeled video data that maximizes the similarity between nearby shots compared to randomly selected shots is presented. This observation used to consider nearby similar shots as augmented versions of each other and demonstrated that when used in a contrastive learning setting. **Performance:** A detailed comparative result for scene boundary detection apply on the MovieNet dataset while requiring only 25% of the training labels, using 9× fewer model parameters and offering 7× faster runtime. First present the results to distill the effectiveness of the learned shot representation in terms of its ability to encode the local scene-structure, and then use detailed comparative results to show its competence for the task of scene boundary detection. Finally, the results of ShotCoL for a novel application of scene boundary detection demonstrated, i.e. finding minimally disruptive ad cue-points. This augmentation scheme can encode the scene-structure more effectively than existing augmentation schemes that are primarily geared towards images and short videos. The proposed model uses an unlabeled dataset AdCuepoints for learning process and achieve 48.40 AP [53].

Technique (8): This work proposes an efficient VSBD method for abrupt transition and gradual transition detection. A block based Mean Cumulative Sum Histogram (MCSH) has been extracted by this method from every edge gradient fuzzified frame as global and local feature. For gradual and abrupt transition detection the statistical measure (Relative Standard Deviation (RSD)) is done on the extracted MCSH. **Performance:** The efficiency of this approach done by using VideoSeg, and TRECVID 2001, 2007 datasets. By comparing this method with some other SBD approaches the results show a high efficiency of the proposed work. The F_1 measures reach to 0.964 outperform Thounaojam et al. model [54].

Technique (9): This method for SBD depends on many features of videos like: histograms of color and the boundaries of object. **Performance:** The evaluation of the proposed method was done on the RAI and open BBC Planet Earth, and MSU CC datasets. F1 score metric was used to measure the accuracy and recall of the found scene changes. The best result of F1-score was 0.979 that is better than other algorithm for scene transition detection. The algorithm speed is slower than analogues, because it uses a larger number of features. The proposed model outperforms VQMT, FFmpeg, and PyScene when applying on datasets OSVSD, RAI and BBC Planet Earth [3].

Technique (10): The proposed method for VSBD based on calculating the histogram for every frame in video, and then takes the difference between the consecutive frames and the average luminance histogram for every frame. The proposed method extract features of texture then compare the differences between frames. The sequence of work will be as follows: detect the video shot mutation and the residual parts will send to next level to continue revealing if the shot changing progressively and check the type of gradual transition. The abrupt transition detected by the proposed method when there is large significant difference between shot frames back and front. The cut shot transition can be detected by giving the threshold a high value. **Performance:** This method evaluated by comparing with other algorithms: distributed machine learning, decision tree, and data mining. This method reduces false transition detection and gives high accuracy than others based on the experimental results. The experiments were done by using precision and recall on 4-videos and show that the algorithm give high accuracy for (sport) videos while (science and technology) and (film) videos have result close to each other, (cartoon) videos has the smallest accuracy. For future work, the work can be improved by using video panorama. The proposed technique outperforms data mining, machine learning, and decision tree algorithms according to the experiments [55].

Technique (11): A novel symmetry method presented to increase the accuracy of SBD. For the total length of video, the global HSV color histogram features are extracted and the threshold calculated. Every initial segment is divided into candidate segments by using the differences of HSV color feature histogram and the local threshold which is computed for every initial segment. Thereafter, each candidate segment is analyzed by computing the SURF for candidate segment boundaries, and checked for the possibility of being segmented more. The cut transition is defined when the elected segment can't be segmented more, otherwise the segment defined as gradual transition. The local threshold is continuously computed for every segment. The global feature extracted by this method once for the full-length video and the SURF feature is extracted only for the candidate segment boundary frames, which greatly reduces the algorithm's complexity. **Performance**: The performance of the proposed work is evaluated by using the dataset TRECVID 2001 and comparing the F1-score results with state-of-the-art approaches which increase the accuracy of cut transition and gradual transition detection. F_1 measures was between [96.9 to 100] for cut transition, and [87 to 96.2] for gradual transition. For future work, this method can be improved to detect fade- in, fade-out, wipe, and dissolve which are types of gradual transition [34].

Technique (12): This work proposed a method for shot transition detection (abrupt and gradual) with high accuracy like fade-in and fade-out. The proposed work combines

DTCWT-WHT for feature detection from each block of frame. There are some processes initially applied on the frames for noise removal where the intensity noise reduces by FAPG filtering. Then a classification process has done accurately by Deep Belief Network (DBN) for gradual transitions detection. With the existence of illumination, the shot transitions detected accurately by the proposed method. **Performance:** The proposed approach evaluated by using Recall, Precision, F-measure, and compared with other approaches such as RNN, CNN, DBN, and DNN by using TRECVID 2016 to 2019 datasets. A comparison of the proposed system with other approaches like walsh hadamard transform, perceptual scheme, singular value decomposition, and multi-modal visual features. The experimental results give better performance of the proposed approach than the other approaches listed previously. The proposed work detects the abrupt and gradual transitions with small false detection. The value of recall, precision, and F1- score obtained by the proposed approach for abrupt detection show the efficiency of this work. A high result values obtained for gradual transition detection by comparing with RNN, CNN, DBN, and DNN. The proposed model applied on TRECVID 2016, 2017, 2018, and 2019 datasets where F_1 measures is [92 to 93.5] for abrupt transitions and [85.46 to 88.13] for gradual transitions [56].

Technique (13): This proposed SBD method is efficient and has many invariant features like Color Layout Descriptor (CLD), Scale Invariant Feature Transform (SIFT), and edge change ratio (ECR) that assist the improvement of SBD accuracy. The false boundary detection is reduced due to the use of the previous feature detection methods. The frames are classified to shot or transition frame by using SVM classifier. **Performance:** TRECVID 2007 dataset were used to analyze the proposed method. The performance of this approach was high however when there is variance with illumination, scaling, rotation, etc. The performance of this method done by using recall, precision, F1-score, and compared with some state-of-art methods. In the SBD method the Color Layout Descriptor (CLD) and Edge Change Ratio (ECR) where illustrated by the average of precision, recall, F1-score. This approach shows a 97.6 F1-score, which is a comparatively the better result. A comparison of the experimental results with existing methods reveals that the proposed method outperforms in SBD process by reducing the false positives [35].

Technique (14): The proposed technique based on the framwork pyramidal relative entropy-based LSTM where used to detect video transition. The LSTM network is trained to detect the transitions through temporal signature and categorizes them as no transition, abrupt transition and gradual transition. Initially, the uncompressed video frame is modeled using pyramidal attributes, and then the temporal signature is generated based on forward-backward ratio of relative entropy measure. **Performance:** The performance of this work done by using recall, precision, and F1-score metrics. The average F1-score presents the performance of the proposed work and done on the datasets TRECVID 2001, TRECVID 2007, VIDEO-SEG 2004, RAI and BBC. The experimental results done by MATLAB and show that the proposed work performed with a high F1-score degree on benchmark dataset. The proposed work gives better results than most of the other techniques except the OP model, however the only the abrupt transitions only detected by the OP model. The proposed work exceeded the performance of the OP model in the evaluation metrics. The proposed model applied on TRECVID 2001 where the average of F_1 measure is 92.7 for abrupt transitions while the average of F_1 measure for the state-of-the-arts methods are 91.6, 89.6, 84.9. The average of F_1 measure is 98.8 for gradual transitions while the average of F_1 measure for the state-of-the-arts methods are 98.5, 96.9, 92.3. For future work, use suitable frame sampling, customized LSTM architecture and fusion techniques to deal with the complexity of the learning algorithm [57].

Conclusion

Video processing applications need an essential step defined by shot boundary detection (SBD) or video segmentation. There are many researches with various techniques used for SBD, but the recently proposed techniques have been listed in this work for the comparison process. The measurements used to prove the efficiency of the proposed techniques are recall, precision, and F1-measures. The values of the F1-measure will be discussed for each technique to summarize the efficiency of the listed works. All methods give good results, but Technique (14) [57] provided the best results for gradual and abrupt transitions, where the average F1 measure is 92.7 for abrupt transitions and 98.8 for gradual transitions. This paper provides a survey on types of video transitions, the recent research for SBD, their performance, and the future works. For future work, this survey can be improved by discussing one of the video transitions and listing the most recent research that relates to that transition.

Acknowledgement

I highly acknowledge the valuable comments of referees to improve this paper. A lot of thanks to my families for their support.

Conflict of Interest

There are no conflicts of interest.

References

[1] Aishy A. Object and Event Extraction for Video Processing and Representation in On-Line Video Applications (Doctoral dissertation, Canada); 2001.

[2] Mohammed H. A Method for Static Video Summarization for Reducing Power Consumption and Its Impact on Bandwidth (master's thesis in Computer Science, Middle East University, Jordan); 2016.

[3] Alexander G, Anastasia A, Dmitriy V. Shot Boundary Detection Method Based on a New Extensive Dataset and Mixed Features. 31 International Conference on Computer Graphics and Vision, Russia; 2021.

[4] Edward J Y. A New Method for Static Video Summarization Using Visual Words and Video Temporal Segmentation (master's thesis, Federal University of Ouro Preto, Brazil); 2013.

[5] Tekalp A M. Digital Video Processing. Prentice Hall, 2nd Edition; 2015.

[6] Abdul-Amir A K, Rafal A S. Image Feature Extraction and Selection. Iraqi Journal of Science, Vol. 59, No. 3B, 2018; pp. 1501-1508.

[7] Rafal A S. A General Overview on the Categories of Image Features Extraction Techniques: A Survey. Journal of Al-Qadisiyah for Computer Science and Mathematics, Vol. 15, No. 1, 2023; pp. 179–188.

[8] Shrikant C, Ravi M, Pranay Y. A Comparative Analysis of Video Shot Boundary Detection using Different Approaches. 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART), IEEE; 2021.

[9] Rainer L. Reliable Transition Detection in Videos: A Survey and Practitioner's Guide. International Journal of Image and Graphics, Vol. 1, No. 03, 2001; pp. 469-486.

[10] John S B, Lawrence A R. Comparison of Video Shot Boundary Detection Techniques. Journal of Electronic Imaging, Vol. 5, No. 2; 1996.

[11] Costas C, Nikos N, Ioannis P. Video Shot Boundary Detection and Condensed Representation: A Review. Published in: IEEE Signal Processing Magazine, Vol. 23, No. 2, 2006; pp. 28-37.

[12] Salim A C, Sudhir G A. Gradual Transition Detection Algorithms in Video Segmentation: A Survey. International Journal of Scientific & Engineering Research, Vol. 4, No. 2; 2013.

[13] Saranya K, Kethsy P A. Survey on a New Approach for Video Cut Detection Using Fuzzy Logic Control. International Journal of Engineering Research & Technology, Vol. 2, No. 12; 2013.

[14] Mohini D, Ruhi K. Video Shot Detection Techniques Brief Overview. International Journal of Engineering Research and General Science, Vol. 2, No. 6; 2014.

[15] Nikita S, Ravi M. A Survey Based on Video Shot Boundary Detection Techniques. International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, No. 4; 2014.

[16] Dalton M, Amit T, Manglem S, Sudipta R. A Survey on Video Segmentation. Intelligent Computing, Networking, and Informatics, Advances in Intelligent Systems and Computing, Springer, India; 2014.

[17] Gautam P, Dwijen R, Suvojit A, Ruben R, Sayan C, Nilanjan D. Video Shot Boundary Detection: A Review. Springer International Publishing Switzerland, Vol. 2; 2015.

[18] Mikhal R, Salim C. Comparative Study of Various Algorithms for Detection of Fades in Video Sequences. The International Journal of Engineering and Science, Vol. 5, No. 7. 2016; pp. 77-82.

[19] Israa H, Hikmat Z. Video Shot Boundary Detection: A Comprehensive Review. Journal of Engineering and Applied Sciences; 2017.

[20] Sadiq H, Abd-Rahman R, Iqbal M, Basheera M, Syed A, Wissam A. Methods and Challenges in Shot Boundary Detection: A Review. Entropy (Basel), Vol. 20, No. 4; 2018.

[21] Noraida H, Fadilah H. Video Authentication using Shot Boundary Detection Technique-An Overview. Journal of Physics: Conference Series; 2019.

[22] Pragati A, Jadhav D. A Review on Shot Boundary Detection. International Journal for Research in Engineering Application & Management, Vol. 05, No. 12; 2020.

[23] Benoughidene A, Titouna F. Shot Boundary Detection: Fundamental Concepts and Survey. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0); 2019.

[24] Shrikant P C. Survey and Comparative Study of Video Shot Boundary Detection with Different Approaches. International Journal of Engineering Research and Management (IJERM), Vol. 7, No. 6. 2020; pp. 1773-1779.

[25] Raja S M, Kalaivani A. Performance Analysis of Video Shot Segmentation Algorithm. European Journal of Molecular & Clinical Medicine, Vol. 7, No. 5; 2020.

[26] Victoria C, Rajeshwari D. Perspective Study on Content Based Video Retrieval. International Journal on Emerging Technologies, Vol. 11, No. 2; 2020.

[27] Swati H, Ravi M. Comparative Study of Different Video Shot Boundary Detection Techniques. International Journal of Trend in Scientific Research and Development, Vol. 6 No. 5; 2022.

[28] Shujing P, Shutao S, Lifang Y, Fengfeng D. Video Shot segmentation algorithm based on SURF. 4th International Conference on Mechatronics, Materials, Chemistry and Computer Engineering; 2015.

[29] Colin O. Analysis of Shot Boundary Detection Techniques on a Large Video Test Suite (master's thesis for the degree of Science, Ireland); 1999.

[30] Biswanath C, Siddhartha B, Susanta C. Video Shot Boundary Detection Using Support Vector Machine. International Conference on Computer Science and Information Engineering. 2015; pp. 45-54.

[31] Ruaa H, Ali A. Comparative Performance of Shot Change Detection Techniques in MPEG Stream. Journal of Babylon University/Pure and Applied Sciences, No. 3, Vol. 24; 2016.

[32] Youssef B. Video shot boundary detection and key-frame extraction using mathematical models (Doctoral dissertation, France); 2017.

[33] Dalton T, Thongam K, Thokchom J, Sudipta R, Khumanthem S. Colour Histogram and Modified Multi-Layer Perceptron Neural Network Based Video Shot Boundary Detection. The International Arab Journal of Information Technology, Vol. 16, No. 4, 2019; pp. 686-693.

[34] Raja S M, Kalaivani A, Anusuya S. The Detection of Video Shot Transitions Based on Primary Segments Using the Adaptive Threshold of Colour-Based Histogram Differences and Candidate Segments Using the SURF Feature Descriptor. Symmetry, Vol. 14, No. 10; 2022.

[35] Jasmin T, Rajkumar S, Muhammad R, Achyut S, Pavika S, Mohammad R. Efficient Shot Boundary Detection with Multiple Visual Representations. Hindawi, Mobile Information Systems; 2022.

[36] Jun L, Youdong D, Wei L, Yunyu S. DWT-based Shot Boundary Detection Using Support Vector Machine. International Journal of Computer Information Systems and Industrial Management Applications, Vol. 1, 2009; pp. 214-221.

[37] Ahmed T, Nada F, Amaal G. Using Daubachy Wavelet for Shot Boundary Detection. Journal of Computer Engineering, Vol. 16, No. 6; 2014.

[38] Thakre K, Rajurkar A. Shot Boundary Detection of MPEG Video using Biorthogonal Wavelet Transform. International Journal of Pure and Applied Mathematics, Vol. 118, No. 7; 2018.

[39] Rashmi B S, Nagendraswamy H S. Abrupt Shot Detection in Video using Weighted Edge Information. Publication rights licensed to ACM; 2016.

[40] Ruxandra G T. Segmentation and Structuring of Video Documents for Indexing Applications (Doctoral dissertation, Telecommunications and Electronics, Paris); 2013.

[41] Swati D B, Bijal J T. Analysis of Popular Video Shot Boundary

Detection Techniques in Uncompressed Domain. International Journal of Computer Applications, Vol. 60, No. 3; 2012.

[42] Kavitha J, Sowmyayani S, Arockia P, Shot Boundary Detection Using Octagon Square Search Pattern. World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering, Vol. 10, No. 7; 2016.

[43] Vijeta S, Manjari G, Ajai K, Deepti M. Video Processing using Deep Learning Techniques: A Systematic Literature Review. IEEE, Vol. 9, 2021; pp. 139489-139507.

[44] Lifang W, Shuai Z, Meng J, Zhe L, Dong W. Two Stage Shot Boundary Detection via Feature Fusion and Spatial-Temporal Convolutional Neural Networks. Special Section on Ai-Driven Big Data Processing: Theory, Methodology and Applications; 2019.

[45] Helm D, Kampe M. Video Shot Analysis for Digital Curation and Preservation of Historical Films. Eurographics Workshop on Graphics and Cultural Heritage; 2019.

[46] Neelam L, Yogeshkumar S, Parul S. Key Feature Extraction for Video Shot Boundary Detection using CNN. International Journal of Recent Technology and Engineering, Vol. 8, No. 5; 2020.

[47] Dattatraya J, Yogeshkumar S, Parul A. Profound Learning Approach for Shot Boundary Location. International Conference on Communication and Information Processing; 2020.

[48] Padalkar M G. Histogram Based Efficient Video Shot Detection Algorithms (master's thesis, India); 2010.

[49] Feng D, Fei M. Video Shot Boundary Detection Based on Feature Fusion and Clustering Technique. IEEE Access; 2020.

[50] Nandini H, Chethan H, Rashmi B. Abrupt Shot Change Detection using Midhinge Local Binary Pattern. IEEE Xplore; 2020.

[51] Tomáš B. Deep Learning-Based Approaches for Shot Transition Detection and Known-Item Search in Video (master's thesis, Ukrainian); 2020.

[52] Nandini H, Chethan H, Rashmi B. An Efficient Method for Video Shot Transition Detection Using Probability Binary Weight Approach. International Journal of Computer Vision and Image Processing, Vol. 11, No. 3; 2021.

[53] Shixing C, Xiaohan N, David F, Dongqing Z, Vimal B, Raffay H. Shot Contrastive Self-Supervised Learning for Scene Boundary Detection. IEEE Xplore; 2021.

[54] Rashmi B S, Nagendraswamy H S. Video shot boundary detection using block based cumulative approach. Multimedia Tools and Applications, Vol. 80, No. 4, 2021; pp. 1-24.

[55] Xu C. Sports Video Panorama Synthesis Technology Based on Edge Computing and Video Shot Boundary Detection. Hindawi Wireless Communications and Mobile Computing. 2022; pp. 1-9.

[56] Shrikant C, Ravi M. An Efficient Approach for Shot Boundary Detection in Presence of Illumination Effects using Fusion of Transforms. International Journal of Engineering Trends and Technology, Vol. 70, No. 4, 2022; pp. 418-432.

[57] Sasithradevi A, Mohamed M, Saurav G. Pyramidal-Relative Entropy Based Temporal Signature for Video Transition Detection using LSTM. Research Sequare; 2022.