Multimedia Tools and Applications (2021) 80:641–664
https://doi.org/10.1007/s11042-020-09697-6

Video shot boundary detection using block based cumulative approach

B. S. Rashmi[1,2] • H. S. Nagendraswamy[1]

Received: 10 September 2019 / Revised: 18 August 2020 / Accepted: 21 August 2020 /
Published online: 4 September 2020

Abstract

Video data is becoming an indispensable part of today's Big Data due to evolution of social web and mobile technology. Content based video analysis has become crucial for video management. Shot boundary detection is one of the most essential task in video content analysis. In view of this, an efficient shot boundary detection approach to detect abrupt and gradual transition in videos is proposed in this work. The approach extracts block based Mean Cumulative Sum Histogram (MCSH) from each edge gradient fuzzified frame as a combination of local and global feature. The relative standard deviation (RSD) statistical measure is applied on the obtained MCSH to detect abrupt and gradual shots in the video. Efficacy of the proposed method is measured by conducting experiments on TRECVID 2001, TRECVID 2007 and VideoSeg datasets. The proposed method shows relatively a good performance when compared to some of the state-of-the-art shot boundary detection approaches.

Keywords Block · Cumulative · Fuzzy set · Gradient · Histogram · Relative Standard Deviation · Threshold · Shot boundary detection

1 INTRODUCTION

In the recent years Internet and social media platforms are ubiquitous and plethora of video information is generated in every single minute. With the proliferation of 5G technology and the advancement in smart phones, mobile users and Internet of Things (IoT) are predicted to increase mobile video data traffic. Development in video acquisition technologies has led to the creation of massive video repositories on storage platforms. The users may prefer to query videos based on the content instead of sequentially accessing the video data, which demands

✉ B. S. Rashmi
rashmibsrsh@compsci.uni–mysore.ac.in

[1] DoS in Computer Science, University of Mysore, Mysore 570006, India

[2] Department of Information Technology, Karnataka State Open University, Mysore 570006, India

sophisticated technology for representing, indexing and retrieving multimedia data. Video management in manual mode is arduous and hence it is crucial to develop efficient algorithms to store, index and retrieve the videos. This domain of research is referred as Content Based Video Retrieval (CBVR) system. CBVR seem to be inherent extension of Content Based Image Retrieval (CBIR). CBVR system is the task of providing relevant video shots/clips as per the user query. The approaches and paradigms for CBVR must promote to align computer vision in line with human perceptions [8]. The term "content" stands for image features such as color, shape, texture etc. and the term "retrieval" refers to the techniques that fetch results in relation and accordance with user perception. Thus, CBVR can be imposed as the search for videos that matches the query given by the user.

CBVR technology has been successfully used in several applications such as crime prevention, biometrics, gesture recognition, biodiversity information systems, medicine, digital libraries, historical research etc. The widespread applications of videos have increased the demand for automated tools and management for efficient indexing, browsing and retrieval of video data [13]. Since video retrieval is not effective using conventional query-by-text retrieval technique, CBVR system is considered as one of the best practical solutions for better retrieval quality [46]. The rich video structure has got tremendous scope in the area of video retrieval to enhance the performance of conventional search engines [21]. It is vital to develop appropriate measures to effectively and efficiently manage the multimedia information in a meaningful manner [7].

The research community working on CBVR has identified several challenges in the design of effective and efficient CBVR system. Shot Boundary Detection (SBD) is the crucial step in the design of CBVR system and aims at segmenting a video into a number of structural elements (scenes, shots or frames) [7]. Hence, it is termed as video temporal segmentation and has been focused for video summarization and indexing process. Some of the challenges of SBD include efficient feature descriptors and threshold free algorithms to achieve high detection rate for identification of any type of shot transition. The logical representation of video with respect to hierarchical structure is shown in Fig. 1. The shot boundaries are indicated relying upon the interruptions made between camera operations. A video stream is anticipated as a set of distinct scenes. A shot is a sequence of successive frames grabbed from a single camera. A representative frame termed as keyframe can be identified from every shot which constitutes summary of the video and is further used for retrieval purpose.

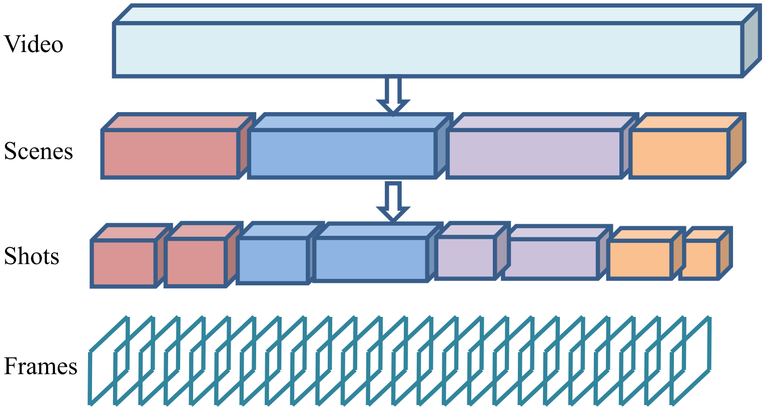


Fig. 1 Hierarchy of video units

A frame represent a single image in a video. Consecutive frames within a shot are highly redundant in appearance and behaviour. According to the video editing effects, the shot boundaries are classified as either abrupt (hard) or gradual (soft) transition based on the inherent behaviour, properties and length of the videos [48]. Figure 2 presents the categories of video shot transitions. Determining gradual transition is complex when compared with abrupt transition due to camera and object motion [8].

Abrupt shot detection The rapid change in visual content between adjacent frames in a video causes abrupt transition. Such transitions portray substantial visual discontinuity between frames and termed as hard cut.

Gradual shot detection The slow and continuous change in visual content across multiple frames causes gradual transition. Such transitions exhibits progressive change over varied group of frames and termed as smooth transition. Gradual transition detection is tedious due to continuous nature of the editing effects. The editing effects considered as gradual transition are fade-in, fade-out, dissolve etc.

The primary step in SBD involves feature extraction and representation of frames. Subsequently, similarity/dissimilarity measures are computed to locate the transition between frames. Based on significant changes, shot boundaries are declared. Shot transition occurs when there is a drastic change in visual contents between the frames. In the literature, the SBD techniques are broadly classified into two categories based on the feature extraction domain viz., compressed and uncompressed. Features extracted from compressed domain make SBD algorithms fast, as no decoding process for video frames are required. However, researchers pay more attention to uncompressed domain because of its richness in visual information of video frames as discussed in [3]. The SBD algorithms proposed in this work are all based on uncompressed domain.

Researchers in the field of image and video analytics have revealed that, the methods based on soft computing techniques [5, 33, 34, 47] have shown better performance when compared to conventional methods [9]. The two dimensional discretization has caused inherent uncertainty in digital frames [29] and hence, there exists some amount of uncertainty even in simplest feature extraction approach [29]. The ambiguities in digital frames occur due to position of object and pixel intensity. Amongst the interesting features, edges interpret the boundary of the objects with variations in pixel intensities. Fuzzy set theory has been a vital choice in handling ambiguities and processing of edges [29].

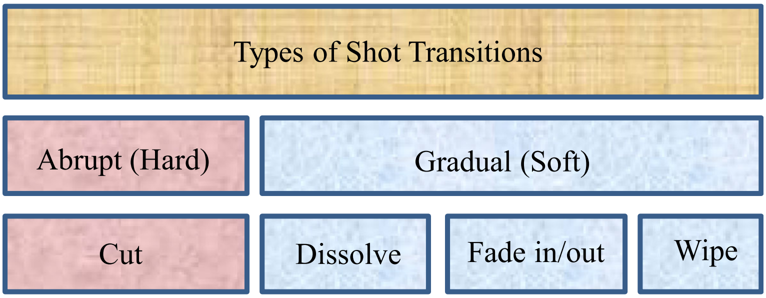


Fig. 2 Categories of video shot transition

In the proposed work, the task of shot boundary detection is achieved by employing edge information and incorporating fuzzy logic [57]. The edge detection process is likely to consist of several sub processes or phases [52]. The mathematical framework introduced by Bezdek et al., [6] consisting of several phases viz., conditioning, feature extraction, blending and scaling for edge detection is used by many methods in the literature. The concept of fuzzy logic can be applied in all the phases or at a specific phase depending on the nature and complexity of the video to produce better shot detection results [29]. In classical edge detection, binarization step causes information to be lost in digital frames [35]. The use of fuzzy sets for intermediate representation of edges may capture the important information [29, 51], which can help to describe the content of the video frame better and handle ambiguities in digital frames. This has motivated the authors to carry out the proposed SBD work.

The focus of this work is to identify abrupt and gradual transition in the videos. An attempt is made to enhance edge detection capability and address uncertainty in digital frames. Initially, the grayscale frames are transformed into gradient frames using Sobel detector. The Sobel gradient distribution of pixels are subjected to fuzzification process using triangular membership function (MF). The proposed method employs block based cumulative sum approach on each 3 × 3 block pixels of fuzzified gradient frame. This local feature discriminates the spatial distribution and is robust to noise and illumination variation. Further, the mean of cumulative sum is computed and is used to produce MCSH histogram of every video frame. Thus, the video is represented in terms of MCSHs and each MCSH describes the video frame information globally. Threshold devising strategy is accomplished by applying RSD statistical measure on the obtained MCSH histograms of every frame of a video for shot transition detection. Efficacy of the proposed method in terms of precision, recall and F1-score has been demonstrated by conducting extensive experiments on the TRECVID and VideoSeg datasets.

Rest of the paper is organized as follows: Section 2 gives a detailed description of the related work. Section 3 presents the proposed methodology covering the details about feature extraction and shot boundary detection. Experimental analysis and results are discussed in section 4 followed by conclusion in section 5.

2 LITERATURE REVIEW

With the advancement in CBVR technology, there is a great demand for robust and reliable SBD algorithms [7]. A comprehensive survey on recent developments of SBD has been reported by Abdulhussain et al., [3]. Numerous challenges of shot boundary detection and extensive review of several techniques are presented by Hu et al., [19] and Yuan et al., [56]. Some of the important work related to SBD, which address abrupt and gradual transitions are discussed in the following paragraphs. Among the several approaches proposed in the literature to address the problems of SBD, following are some of the interesting works, which have explored histograms, edge based, block based features and soft computing techniques.

Histogram is a global feature and does not capture spatial details of the pixels. Hence, it is robust to camera or object motion than pixel based methods. Mas and Fernandez [32] have depicted the effectiveness of color histogram descriptor using color space and quantization method by discriminating the least significant bits of each RGB component. City block distance between color histograms were measured and compared against threshold to detect shot cuts. Ji et al., [22] used the concept of accumulative histogram difference and support points for detection of dissolve transition. Lu and Shi [30] proposed singular value decomposition and candidate segment selection method for SBD. A frame feature matrix is formed by extracting color histogram in hue saturation value to identify cut and gradual transition. Li et al., [27] presented three-stage approach based on Multilevel Difference of colour histograms for detecting cut and gradual boundaries. Detection of shot boundaries are attempted by Hannane et al., [16] by extracting SIFT and edge-SIFT keypoints from each frame. Adaptive threshold is applied on the computed distance values of SIFT-PDH between the frames and shot boundaries are identified. Prasertsakul et al., [36] presents a novel technique for classifying several camera operations in videos using 2D histogram. 2D motion vector (MV) fields are generated by applying an existing block based MV estimation method in polar coordinates. MVs in each frame that share the similar magnitude and orientation features are utilized to classify the camera operations by representing the 2D histogram.

Edge is an important local feature to represent discontinuity in pixel intensity. Pixels belonging to same object exhibit continuity in pixel intensity and vice versa. Significant changes in the edge pixels between consecutive frames, indicate a shot change. Since spatial information is not considered, missed shot boundaries may occur [48]. This technique is applied to detect both abrupt and gradual transitions. Heng and Ngan [18] presented shot boundary detection using object based edge detection. The authors proposed time stamping transferring mechanism, which utilizes information across multiple frames. Moving objects across the gradual transition frames instead of adjacent frames are tracked by the concept of edge object tracking. Zheng et al., [60] proposed heuristic algorithm for detection of fade in and fade out transition. This work utilizes Robert edge detector and transition is detected by identifying the separation from object motion by employing predefined adaptive threshold. Adjeroh et al., [4] introduced adaptive edge oriented framework using multilevel features based on shot variability to address the problem of identifying abrupt transition. Three levels of adaptation are considered by the authors: at the feature extraction stage using locally-adaptive edge maps, at the video sequence level, and at the individual shot level. Adaptive parameters for multilevel edge based approach are formulated to determine adaptive thresholds for detection of shot boundaries. Priya and Domnic [38] used edge strength as feature vector that are extracted by projecting block of frames over vector space. The sum of absolute difference between the features of the blocks of the corresponding frames are evaluated. Shot transitions are categorized by using similarity difference values.

Block based approach acts as intermediary between local and global feature based approaches. Since the spatial resolution is reduced by using blocks instead of pixels, this method is less sensitive to object and camera motion. Shahraray [43] proposed block based technique by dividing the frame into 12 non overlapping blocks. Non linear order statistics is used to find the best match between respective neighbourhoods of the previous frame. Sustained low level increase in match values are identified to detect shot cuts. Lee et al. [25] performed block differences using HSV color space. The mean values of Hue and Saturation for two successive blocks are computed and shots were detected. Lian [28] has proposed pixel, histogram and motion based frame difference to resist flash and light detection to avoid false positives to address shot boundary detection. Jiang et al., [23] have proposed both pixel and histogram based method for detection process using uneven blocked color histogram and uneven pixel value difference in the moving windows. Rashmi and Nagendraswamy [39] have proposed shot cut method using edge information and constructing histogram by assigning binary weights to each sliding window of 2×2 block/mask of a video in overlapping and non overlapping mode. To enhance discriminative capability among spatial distribution, Rashmi and Nagendraswamy [40] have proposed Midrange LBP texture descriptor where midrange threshold value is applied for each pixel across 3×3 block of image matrix to produce histogram and adaptive threshold is used to detect shot boundaries. Wu et al., [54] proposes Unsupervised Deep Video Hashing where balanced code learning and hash function learning are integrated and optimized for video retrieval. Feature clustering and binarization are used to preserve neighbourhood structure. Smart rotations is used for generating effective hash codes. Wu and Xu [55] proposed bottom-up and top-down attention model to perform color image saliency detection in news video. Multi-scale local and global motion conspicuity maps are computed on eye-tracking datasets. Shen et al., [44] proposed video event detection using subspace selection technique. Unified transformation matrix is used for projecting different modalities for individual recognition tasks. Zhang et al., [59] proposed flash model and cut model using local window based method to detect false transitions. Zhang et al., [58] have proposed a shot boundary detection technique based on block-wise principle component analysis by dividing the video into several segments. Shot eigen spaces are established on the training segments and the candidate segments are projected onto the corresponding shot eigen space to extract the feature vectors. Analysis and pattern matching are performed to identify abrupt and gradual shot transitions in the video. Cirne et al., [11] proposed video summarization method using color co-occurrence matrices as frame representation. Feature extraction has been performed at multiple scales. Normalized sum of squared differences are computed between the frames for detecting the shots. Abdulhussain et al., [2] proposed Orthogonal Polynomial (OP) algorithm for detection of hard transitions. The OP domain are computed using Krawtchouk-Tchebichef polynomial. The shots are identified using Support Vector Machines.

In the recent years, many researchers have emphasized their work on soft computing techniques to handle uncertainties in images for addressing SBD. Fuzzification of frame-to-frame-property difference values using Rayleigh distribution and fuzzy rules are framed by Jadon et al., [20] for detection of abrupt and gradual changes. Lee et al. [26] used an ART2 neural network for video scene change detection. Küçüktunç et al., [24] presents color histogram based shot boundary detection algorithm to detect both cuts and gradual transitions with the fuzzy linking method on L*a*b* color space. A set of fuzzy rules are evaluated and fuzzy rule based cut detection approach is suggested by Dadashi and Kanan [12]. Thounaojam et al., [50] used normalized RGB color histogram difference as feature extraction method and finding the difference between consecutive frames is studied for shot detection. The authors have utilized fuzzy logic system optimized by Genetic Algorithm to find optimal range of values of fuzzy membership functions. Hassanien et al., [17] presented SBD technique based on spatio-temporal Convolutional Neural Networks (CNN). The authors have applied deep neural techniques on the large SBD data set of 3.5 millions of frames of sharp and gradual transitions. Image compositing models are used to generate the transitions. Rashmi and Nagendraswamy [41] applied correlation coefficient between consecutive fuzzified frames using fuzzy sets and IFS techniques for the purpose of abrupt shot detection. Artificial neural networks (ANN) represents an important paradigm in the soft computing domain. Gygli et al., [15] developed Ridiculously Fast Shot Boundary Detection with Fully Convolutional Neural Networks. Large temporal context is used by Convolutional Neural Network (CNN) with unprecedented speed for detection process.

3 PROPOSED METHODOLOGY

The framework of the proposed methodology for Shot Boundary Detection process is presented in Fig. 3. The main challenge is to develop a simple approach to eliminate false shot detections that occur due to illumination, camera operation, object motion and noise.

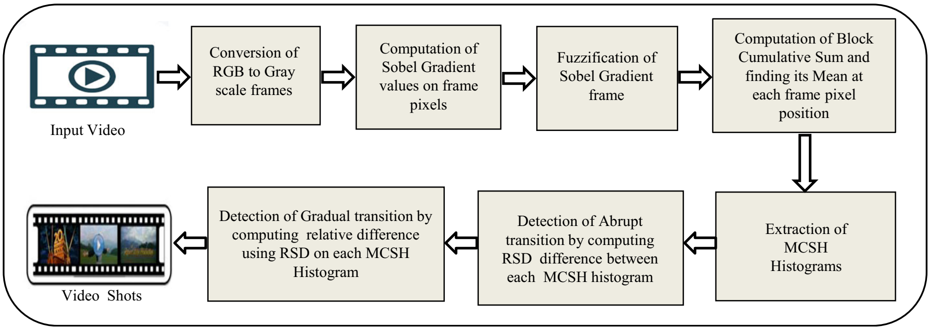

Input Video
Conversion of RGB to Gray scale frames
Computation of Sobel Gradient values on frame pixels
Fuzzification of Sobel Gradient frame
Computation of Block Cumulative Sum and finding its Mean at each frame pixel position
Extraction of MCSH Histograms
Detection of Abrupt transition by computing RSD difference between each MCSH histogram
Detection of Gradual transition by computing relative difference using RSD on each MCSH Histogram
Video Shots

Fig. 3 Framework of the Proposed Methodology for Shot Boundary Detection

Therefore, a combination of local and global feature is considered to address the aforementioned problems by constructing MCSH histograms. The proposed model addresses the detection of both abrupt and gradual transitions present in the videos using MCSH histograms.

3.1 Feature Extraction and Representation

Extraction of meaningful feature and efficient representation of video frames plays a very important role in effective detection of shots in videos. The following subsections presents the detail description of the proposed feature extraction and representation of video frames.

3.1.1 Sobel Gradient Frames

Initially, all the RGB frames of the video are converted to gray scale frames. There exists wide variety of edge detection techniques to measure intensity changes [10, 37, 45] and it is found that Sobel edge detector outperform other edge detectors in terms of accuracy and computational efficiency [1]. In this work, Sobel detector [45] is used to convolve the grayscale frame pixels with their respective convolution mask to obtain gradient frame from grayscale frame. For every pixel in the grayscale frame, the vertical and horizontal components of the gradient is obtained by applying convolution with two 3-by-3 convolution masks as formulated in Fig. 4:

The magnitude of the gradient gives the measure of rate of change in intensity at the pixel location *x,y* and is computed as:

$$\mathbf{G_x} = \begin{array}{|c|c|c|}\hline -1 & 0 & +1 \\ \hline -2 & 0 & +2 \\ \hline -1 & 0 & +1 \\ \hline \end{array} \qquad \mathbf{G_y} = \begin{array}{|c|c|c|}\hline +1 & +2 & +1 \\ \hline 0 & 0 & 0 \\ \hline -1 & -2 & -1 \\ \hline \end{array}$$

Fig. 4 Sobel 3 × 3 convolution masks

$$G_{(x,y)} = \sqrt{G_x^2 + G_y^2} \quad (1)$$

3.1.2 Fuzzification of Sobel Gradient Frames

In order to capture the vagueness and uncertainty present in the data, the crisp data has to be converted to fuzzy data using the process of fuzzification. Membership functions (MFs) are used to carry out fuzzification. MF is a curve that defines how each point in input space is mapped to membership value between 0 and 1 [57]. Different MFs can be used to fuzzify the data which has to be determined empirically by studying the functions for a specific application. The most commonly used MFs in literature are trapezoidal, triangular and Gaussian as they produce good results. The responsibility of choosing the shape of the MF lies with the user and the application. The triangular/trapezoidal MF is used if the system needs significant dynamic variation within short period of time and a Gaussian MF is used if high control accuracy is selected [31].

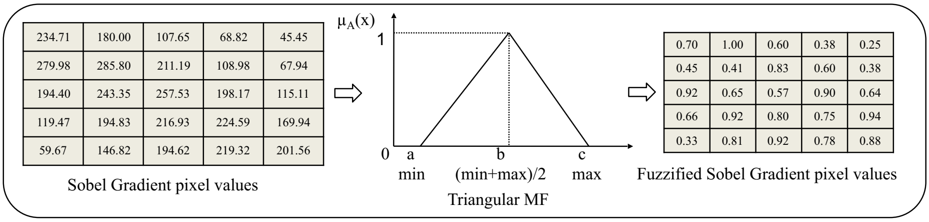
In the proposed approach, the Sobel gradient frame is subjected to fuzzification using triangular MF and the parameters are formulated as follows:

$$\mu_A\left(G_{ij}\right) = \begin{cases} \dfrac{G_{ij}-a}{b-a} & if \;\; a \leq G_{ij} \leq b \\ \dfrac{c-G_{ij}}{c-b} & if \;\; b \leq G_{ij} \leq c \\ 0 & otherwise \end{cases} \quad (2)$$

Where G_{ij} represent the gradient frame. The parameters a, b and c specifies the boundaries with the criteria $(a < b < c)$ and determines the x coordinates of the boundaries of fuzzy triangular MF. In the proposed approach, the left and right boundary values are evaluated as minimum and maximum pixel value of the respective gradient frame and core value is set to midrange value. The numerical illustration of the fuzzified Sobel gradient pixel values are depicted in Fig. 5.

3.1.3 Block Computation and Histogram Construction

The SBD process in the proposed work is based on establishing MCSH histograms for every frame of a video. The illustration for the process of extraction of MCSH histogram for a sample frame #881 of anni006 video sequence is presented in Fig. 6. Initially, the video frame undergoes transformation mechanism from gray scale to fuzzified gradient form as discussed in sections 3.1.1 and 3.1.2.



234.71	180.00	107.65	68.82	45.45
279.98	285.80	211.19	108.98	67.94
194.40	243.35	257.53	198.17	115.11
119.47	194.83	216.93	224.59	169.94
59.67	146.82	194.62	219.32	201.56

Sobel Gradient pixel values

0.70	1.00	0.60	0.38	0.25
0.45	0.41	0.83	0.60	0.38
0.92	0.65	0.57	0.90	0.64
0.66	0.92	0.80	0.75	0.94
0.33	0.81	0.92	0.78	0.88

Fuzzified Sobel Gradient pixel values

Fig. 5 Illustration of obtaining Fuzzified Sobel Gradient Values

For illustration purpose, a 3 × 3 block at each pixel position of the fuzzified frame is considered for evaluation of cumulative sum in overlapping mode. Each block slides over the frame at every pixel position from left to right and top to bottom position. For each block, cumulative sum is evaluated as follows:

$$CS(i) = \textstyle\sum_{k=1}^{i} B(k) \tag{3}$$

Where *B(k)* is 3 × 3 block pixel values and *CS(i)* is the corresponding cumulative sum values considering the block movement in overlapping mode. Further, the mean of cumulative sum values for each block will be computed as follows:

$$\mu = \frac{\sum\limits_{i=1}^{n} CS(i)}{n} \tag{4}$$

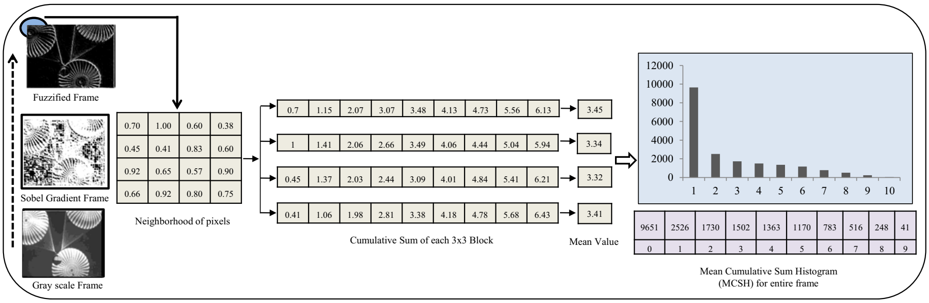
Where μ is the mean value for cumulative sum values for each 3 × 3 block in overlapping mode and $n = 9$. Thus, the evaluated mean value for each 3 × 3 block at each pixel position is used to construct histogram for each fuzzified frame and represented as feature vector as illustrated in Fig. 6. The representation of MCSH histograms of two different frames #881 and #1236 of anni006 video sequence is depicted in Fig. 7 which exhibits distinct bin values.

3.2 Shot Boundary Detection

The task of shot boundary detection is carried out on some of the video sequences of TRECVID and VideoSeg datasets. The dataset is challenging, since it includes large variation of shot breaks. Both abrupt and gradual shot transitions are detected with the aid of RSD measure applied on each MCSH histogram. Also, detection mechanism for elimination of unwanted frames is proposed. The following subsections present a detailed description about the proposed shot detection process.

3.2.1 Detection and Elimination of Unwanted Frames

TRECVID dataset contains non transition frames other than abrupt and gradual transitions. The frames present in non transition group are unwanted frames caused due to flash/light variations, object or camera motion. Even blank/black frames are considered as unwanted frames which will act as abrupt transition [49]. In order to reduce false detections, it is necessary to eliminate unwanted frames prior to abrupt and gradual transition detection. This



0.70	1.00	0.60	0.38
0.45	0.41	0.83	0.60
0.92	0.65	0.57	0.90
0.66	0.92	0.80	0.75

0.7	1.15	2.07	3.07	3.48	4.13	4.73	5.56	6.13	3.45
1	1.41	2.06	2.66	3.49	4.06	4.44	5.04	5.94	3.34
0.45	1.37	2.03	2.44	3.09	4.01	4.84	5.41	6.21	3.32
0.41	1.06	1.98	2.81	3.38	4.18	4.78	5.68	6.43	3.41

9651	2526	1730	1502	1363	1170	783	516	248	41
0	1	2	3	4	5	6	7	8	9

Fig. 6 Illustration of Block Cumulative Sum and Histogram construction on a sample frame

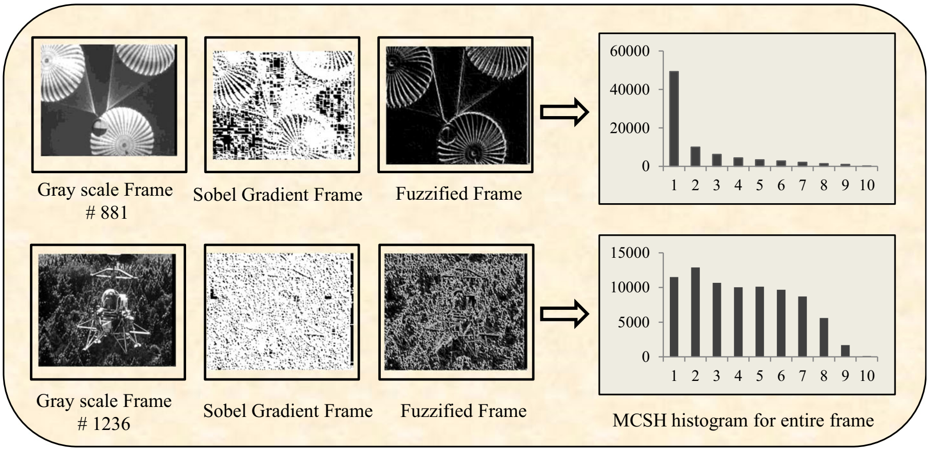


Fig. 7 Representation of MCSH Histogram for two different frames of anni006 video sequence

task is accomplished by applying RSD statistical measure on the obtained MCSH histograms of every frame of a video.

Let P_j contain MCSH histogram values of a frame where $\{j = 1,2,...,n\}$, then RSD_i is the coefficient of variation value corresponding to the i^{th} MCSH histogram and is computed as follows:

$$RSD_i = \frac{\sigma}{\mu} \tag{5}$$

Where, $\mu = \frac{\sum_{j=1}^{n} P_j}{n}$ and $\sigma = \sqrt{\frac{\sum\limits_{j=1}^{n} \left(P_j - \mu\right)^2}{n}}$

A threshold is empirically set for each video to determine transition and non transition frames using RSDs evaluated for all MCSH histograms of a video as follows:

$$T_{NT} = \mu + \alpha\sigma \tag{6}$$

Where T_{NT} is the threshold value, α is constant value, μ is mean value and σ is standard deviation value of all computed RSDs of the entire video. The constant value is chosen by observing the RSD graph and a sample illustration for BG_37309 video is depicted in Fig. 8. During experimentation, it is found that the RSD value of the frames above the threshold T_{NT} are considered as unwanted frames and are excluded from shot detection process. This segregation mechanism ensures reduction in false detection.

It is to be noted that the blank frames will be removed only during abrupt shot detection process. During gradual shot detection blank frames are included in the sequence, as they act as integral part of fade-in and fade-out editing effects. The RSD values as computed above for all the corresponding videos are further utilized by abrupt and gradual detection algorithms and are detailed in the following sub sections.

3.2.2 Abrupt Shot Transition Detection

The significant difference between the frames depends upon the salient content present within the frames. Ford et al., [14] have analysed that, the histogram metric yields best results when

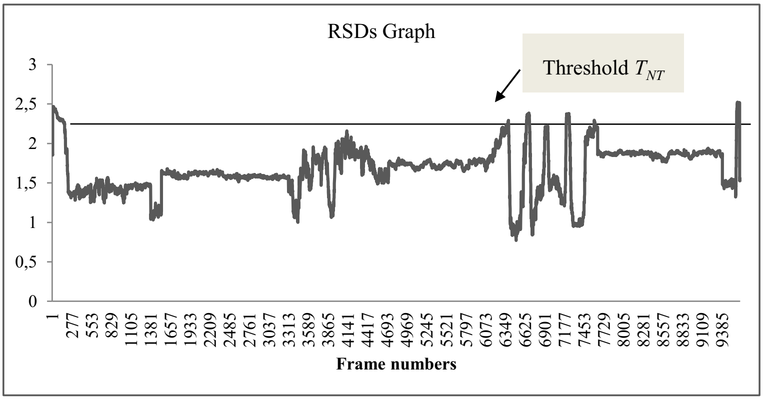


Fig. 8 Illustration of RSD showing non transition frames for BG_37309 video

computed for blocks in case of abrupt transition. In the proposed work, the difference between the RSD values computed for each MCSH histogram as formulated in Eq. 5 of section 3.2.1 is used to detect abrupt shots. Let RSD_k and RSD_{k+1} be the RSD values of the two consecutive MCSH histograms of a video. The comparison between RSD_k and RSD_{k+1} denoted by distance D_{RSD} is computed by finding the difference as follows:

$$D_{RSD} = RSD_k - RSD_{k+1} \tag{7}$$

Thus, the difference value is computed for all other consecutive frames of the entire video. The pictorial representation of distance values thus evaluated is depicted in Fig. 9 for D6 video of TRECVID 2001 dataset.

It can be clearly observed from Fig. 9 that, the distance comparison of RSD values between two consecutive frames belonging to same shot will produce low peaks and prominent peaks for camera break shots. The change in camera breaks is signified by peak variations in the distance of RSD values. Therefore, threshold mechanism has to be devised for identifying prominent peaks. Let μ be the mean, σ be the standard deviation of D_{RSD} values, α is chosen as a constant and a threshold T_{AT} is computed as follows:

$$T_{AT} = \mu + \alpha\sigma \tag{8}$$

The distance values above the set threshold value T_{AT} is considered as prominent peaks indicating camera break operation.

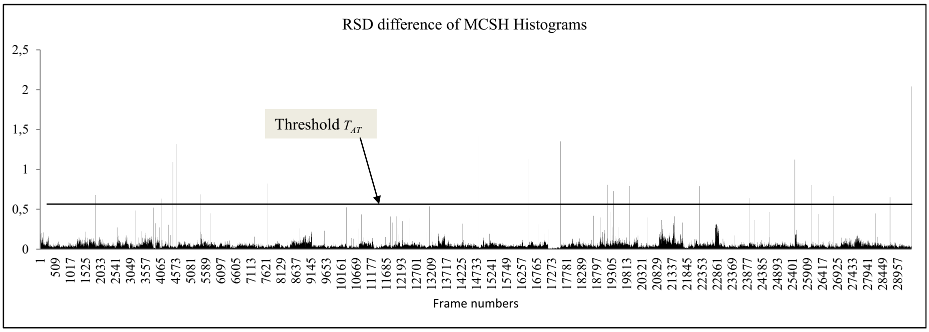


Fig. 9 Distribution of difference between RSD values of D6 (nad58) video

3.2.3 Gradual Shot Transition Detection

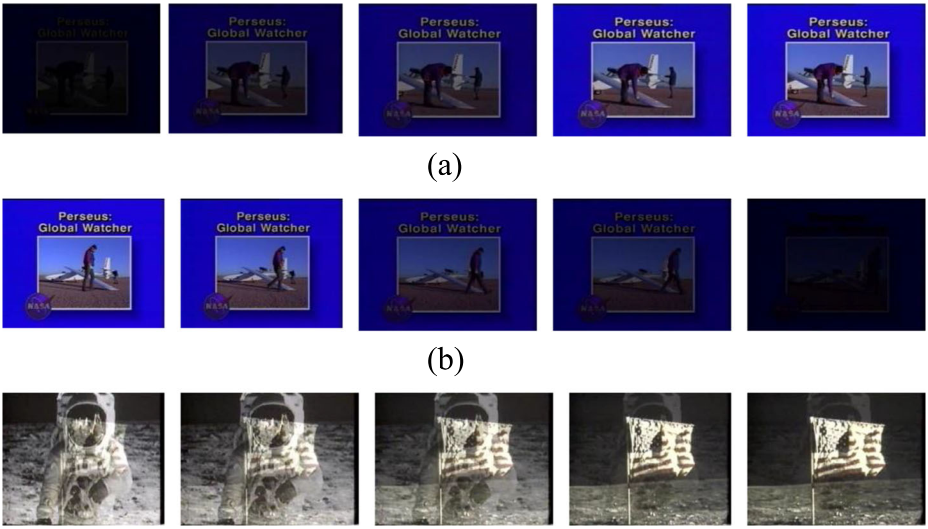
Determining gradual transition is complex when compared to abrupt transition due to camera and object motion [8]. The complexity arises due to varying frame features spread across number of frames. Some of the statistical parameters aid in presenting a distinct pattern for gradual transition [8]. It has been observed that, the patterns representing fade-in, fade-out and dissolve transitions can be categorized with its specific patterns.

Fade-in transition is superimposed combination of blank frames and initial frames of the shot. In this pattern, the blank frames decreases and frames of the appearing shot gets prominent. Fade-out is reverse of fade-in transition. The visual content of the frames of the current shot lose its intensity and gradually turn into black frame. Dissolve transition lasts for few frames when a shot overlaps with succeeding shot. During the overlap process the intensity of the current shot decreases gradually and the intensity of the appearing shot increases linearly. This represents a good combination of fade-out and fade-in transition. It is found that the feature value of last frame in fade-out and first frame in fade-in will be nearing to zero. Usually, dissolve transition is a combination of fade-in and fade-out excluding the occurrence of blank frames as depicted in Fig. 10.

Before applying gradual transition algorithm, the abrupt shots and non transition frames identified using threshold mechanism as described in section 3.2.1 are excluded from sequence of frames. In order to choose the frames for gradual detection process, a threshold has been set using RSD values of MCSH histograms as follows:

$$T_{GT} = \mu + \sigma \tag{9}$$

Where μ be the mean, σ is the standard deviation of computed *RSD* values of entire video. The frames belonging to range T_{GT} and T_{NT} are considered by gradual transition detection algorithm as shown in Fig. 11. This criteria helps in curtailing false detections and aid in improving efficiency of the algorithm.



(a)

(b)

(c)

Fig. 10 Illustration of (**a**) Fade-in (**b**) Fade-out and (**c**) Dissolve transitions

In order to identify the overlapping information across multiple/group frames, an appropriate technique should be used to identify and recognize the patterns. In the proposed approach, RSD measure applied on each MCSH histogram of the corresponding group is utilized by gradual shot detection algorithm. In the subsequent step, mean of all RSD values related to every frame of the corresponding group is computed as follows.

$$M_{RSD} = \frac{1}{n}\sum_{i=1}^{n} RSD \tag{10}$$

Where M_{RSD} is the mean of all *RSD's* in the frame group. Further, the difference between *RSD* and M_{RSD} of each frame is computed and its square value is found. Finally, in order to find the frame feature, the relative difference is computed as formulated in the following equation:

$$F_i = \frac{(RSD_i - M_{RSD})^2}{M_{RSD}} \tag{11}$$

where F_i is the feature value computed for each frame in the sequence. While conducting experiments, feature values are computed considering group of frames and group size is chosen empirically at each instance. Since gradual transitions occur over multiple sequences of frames, it is essential to observe patterns over multiple frames. After plotting F_i values for each frame in the group, the increase or decrease pattern is examined that represents various types of gradual transitions (dissolve, fade-in and fade-out excluding wipe transition). Based on the pictorial representation of F_i values plotted, the pattern can be characterized to be fade-out, fade-in or dissolve gradual transition. This step is repeated for all the remaining group of frames in the sequence. The results of gradual transition detection presented in Fig. 12 (a) shows increasing pattern representing fade-in and Fig. 12 (b) shows decreasing pattern representing fade-out. Figure 12 (c) shows dissolve pattern.

4 EXPERIMENTAL RESULTS AND DISCUSSION

Dataset The experimental analysis of the proposed method has been performed on TRECVID and VideoSeg benchmark dataset and has been assessed with common metrics and compared

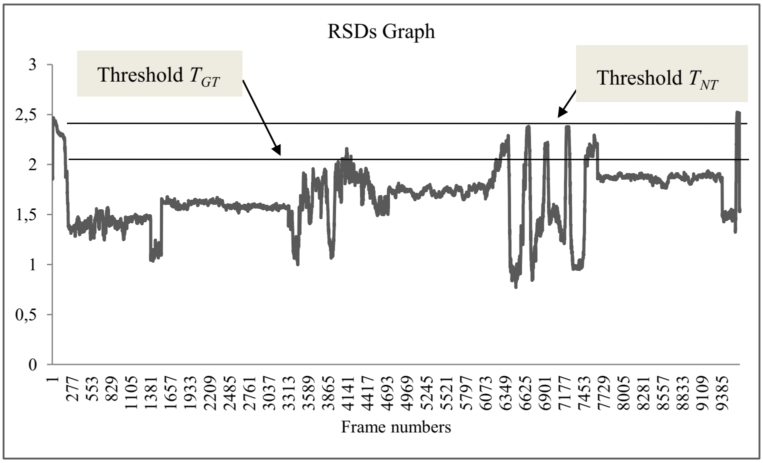
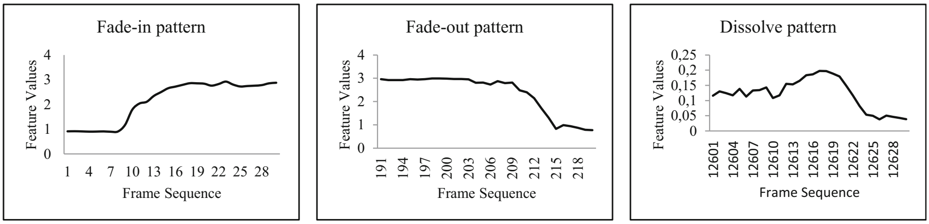


Fig. 11 Criteria set for gradual transition detection



(a) (b) (c)

Fig. 12 Illustration of (**a**) Fade-in (**b**) Fade-out and (**c**) Dissolve patterns

with the baselines. The description of the benchmark datasets is described along with the ground truth information related to camera effects of every video in the following tables. The potentiality of the proposed method is analyzed using video sequences taken from US National Institute of Standards (NIST) TRECVID 2001 and 2007 dataset. TRECVID 2001 dataset described in Table 1 can be downloaded from the *Open Video Project* whereas TRECVID 2007 data described in Table 2 has to be obtained from *Netherlands Institute for Sound and Vision*. Also. the VideoSeg benchmark dataset [53] containing 10 different videos with varied quality and resolution is used for experimental analysis and summarized in Table 3.

The dataset considered for experimentation is of varied length, genre and challenging scenarios which comprises of video editing effects along with camera/object motion and illumination variation. The presence of camera/object motion and sudden light variation causes ambiguous shot boundaries.

Discussion The performance of the proposed method is evaluated using quantitative evaluation metrics such as Recall, Precision and F1-score which is formulated as follows:

$$Recall = \frac{N_C}{N_C + N_M} \tag{12}$$

$$Precision = \frac{N_C}{N_C + N_F} \tag{13}$$

Table 1 Description of TRECVID 2001 dataset

Video Name	Video Title	No. of Frames	Cuts	Gradual	Total
anni005	NASA_25th_Anniversary_Show_Segment_5	11,364	38	27	65
anni006	NASA_25th_Anniversary_Show_Segment_6	16,586	41	31	72
anni009	NASA_25th_Anniversary_Show_Segment_9	12,307	38	65	103
anni010	NASA_25th_Anniversary_Show_Segment_10	31,389	98	55	153
nad31	Spaceworks - Episode 6	52,405	187	55	242
nad33	Spaceworks - Episode 8	49,768	189	26	215
nad53	A&S_Reports_Tape_#4_Report_#260	25,783	83	75	158
nad57	A&S_Reports_Tape_#4_Report_#264	12,781	45	31	76
nad58	A&S_Reports_Tape_#5_Report_#265	13,648	40	45	85
bor03	Challenge at Glen Canyon	48,451	231	11	242
bor08	The Great Web of Water	50,569	380	151	531
Total		325,051	1370	572	1942

Table 2 Description of TRECVID 2007 dataset

Video Name	No. of Frames	Cuts	Gradual	Total
BG_3027	49,813	127	1	128
BG_3097	44,987	91	0	91
BG_3314	35,800	42	0	42
BG_16336	2462	20	0	20
BG_28476	23,236	176	2	178
BG_36136	29,426	88	12	100
BG_37309	9639	11	8	19
BG_37770	15,836	8	29	37
BG_2408	35,890	101	20	121
BG_9401	50,004	259	30	289
BG_11362	16,414	89	3	92
BG_14213	83,113	104	4	108
BG_34901	34,387	106	61	167
BG_35050	36,997	224	16	240
BG_35187	29,023	98	4	102
BG_36028	44,989	135	23	158
BG_36182	29,608	87	0	87
BG_36506	15,208	96	13	109
BG_36537	50,002	77	6	83
BG_36628	56,563	192	10	202
BG_37359	28,906	164	6	170
BG_37417	23,002	76	12	88
BG_37822	21,958	119	10	129
BG_37879	29,017	95	4	99
BG_38150	52,648	215	4	219
Total	848,928	2800	278	3078

$$F1\text{-}score = \frac{2{*}Recall{*}Precision}{Recall + Precision} \tag{14}$$

Where N_C is the number of correct detections, N_M is the number of missed detections and N_F is the number of false detections. F1-score is defined as the harmonic mean of recall and precision which reflects on recall and precision rates. An algorithm having highest F1-score is regarded as an efficient algorithm. Experiments were carried out using MATLAB on Intel Core i5 processor, running at 2.20 GHz with 8 GB RAM. The algorithm complexity of the

Table 3 Description of VIDEOSEG dataset

Video Name	Video Title	Duration (MM:SS)	Frame Size	#Frames	Cuts
A	Cartoon	00:21	144 × 192	649	7
B	Action	00:36	144× 32	957	8
C	Horror	00:53	288 × 384	1618	54
D	Drama	01:45	272× 336	2630	34
E	Science Fiction	00:17	288 × 384	535	30
F	Commercial	00:07	112× 160	235	0
G	Commercial	00:16	288 × 384	499	18
H	Comedy/Drama	03:25	240 × 352	5132	38
I	News/Documentary	00:15	288 × 384	478	4
J	Trailer/Action	00:36	180 × 240	871	87
Total				13,604	280

Table 4 Performance comparison for abrupt shot transition with Thounaojam et al., (2016, 2017) on TRECVID 2001 dataset

Video	Proposed Method			Thounaojam et al., [50]			Thounaojam et al., [49]		
	R	P	F1	R	P	F1	R	P	F1
D2 (anni006)	0.947	0.923	0.935	0.952	0.889	0.919	0.952	0.889	0.919
D3 (anni009)	1.000	0.897	0.946	0.846	0.805	0.825	0.923	0.720	0.809
D4 (anni010)	0.969	0.888	0.927	0.878	0.935	0.906	0.949	0.869	0.907
D6 (nad58)	1.000	0.930	0.964	1.000	0.889	0.941	1.000	0.930	0.964
Average	0.979	0.910	0.943	0.919	0.880	0.898	0.956	0.852	0.900

proposed method depends on the resolution of the video frame pertaining to the dataset. In general the time complexity of the proposed method is of the order of θ(n^2). The time taken for feature extraction per frame is measured in terms of milliseconds (ms) and has been recorded as 0.84 ms for TRECVID 2001, 1.32 ms for TRECVID 2007 and 4.2 ms for VideoSeg dataset.

The efficiency of the proposed system relies on specific threshold values set empirically for each category of transition. The strength of MCSH histograms signifies the overall performance of the proposed approach. The contribution of RSD statistical measure aid in effective detection of abrupt and gradual transitions. Removal of unwanted frames are performed by setting the threshold T_{NT} as described in section 3.2.1. The constant α value is chosen in the range 0.1 to 1 to curtail/reduce false detections.

4.1 Results on Abrupt Transition Detection

During abrupt transition detection, an appropriate threshold value is set to identify dominant peaks based on the observations made by viewing RSD difference graph obtained for each video. Threshold T_{AT} as described in section 3.2.2 is experimented to figure out F1-score. For TRECVID 2001 dataset, different values of α is chosen in the range 0.1 to 3 by observing RSD difference graph (example shown in Fig. 9) and experimental results are recorded. Whereas, the constant value α is chosen in the range 0.1 to 1 for TRECVID 2007 dataset and range 0.1 to 2.5 for VideoSeg dataset.

The achievement of the proposed method based on the analogy of the results obtained with regard to other state-of-the-art approaches are reported in Tables 4, 6 and 8 for TRECVID

Table 5 Performance comparison for abrupt shot transition Sasithradevi et al.,(2020) on TRECVID 2001 dataset

Video Sequence	Proposed Method			Sasithradevi et al.,[42]		
	R	P	F1	R	P	F1
anni005	100	98	99.0	100	97.4	98.7
anni009	100	99.1	99.5	100	97.4	98.7
nad31	99.0	99.0	99.0	98.9	99.5	99.2
nad33	99.4	98.7	99.0	99.4	98.4	98.9
nad53	100	99	99.5	100	98.8	99.4
nad57	100	99.5	99.7	100	100	100
bor03	98.5	99.5	99.0	97.0	98.7	97.9
bor08	96.5	99.2	97.8	93.1	97.2	95.1
Average	99.2	99.0	99.1	98.6	98.4	98.5

Table 6 Performance comparison for abrupt shot transition with Thounaojam et al., (2016) on TRECVID 2007 dataset

Video Name	Proposed Method			Thounaojam et al., [49]		
	R	P	F1	R	P	F1
BG_3027	0.945	0.923	0.934	0.945	0.902	0.923
BG_3097	0.890	0.976	0.931	0.868	0.987	0.924
BG_3314	0.810	0.919	0.861	0.786	0.943	0.857
BG_16336	0.950	1.000	0.974	0.950	1.000	0.974
BG_28476	0.983	0.966	0.975	0.977	0.955	0.966
BG_36136	0.977	0.989	0.983	0.977	0.977	0.977
BG_37309	1.000	0.846	0.917	1.000	0.846	0.917
BG_37770	1.000	0.889	0.941	1.000	0.889	0.941
Average	0.944	0.938	0.939	0.928	0.937	0.935

2001, TRECVID 2007 and VideoSeg dataset respectively. An additional comparison has been made with the recent state-of-the-art algorithm proposed by Sasithradevi et al., [42] for some of the video sequences of TRECVID 2001 dataset and TRECVID 2007 dataset and the obtained results are recorded in Tables 5 and 7 respectively.

The results shown in Tables 4,5,6,7 and 8 and signifies improved efficiency of the proposed algorithm with the state-of-the-art methods and the performance is depicted graphically in Figs. 13(a) to 13(e). By comparative analysis, it is noticeable from the graphs depicted that, the proposed method outperform other SBD approaches. However, the proposed method segments the video considering combination of local and global feature of the frame. Significant improvement has been noticed based on the capability of the proposed method in detecting meager number of missed and false transitions.

Table 7 Performance comparison for abrupt shot transition with Sasithradevi et al.,(2020) on TRECVID 2007 dataset

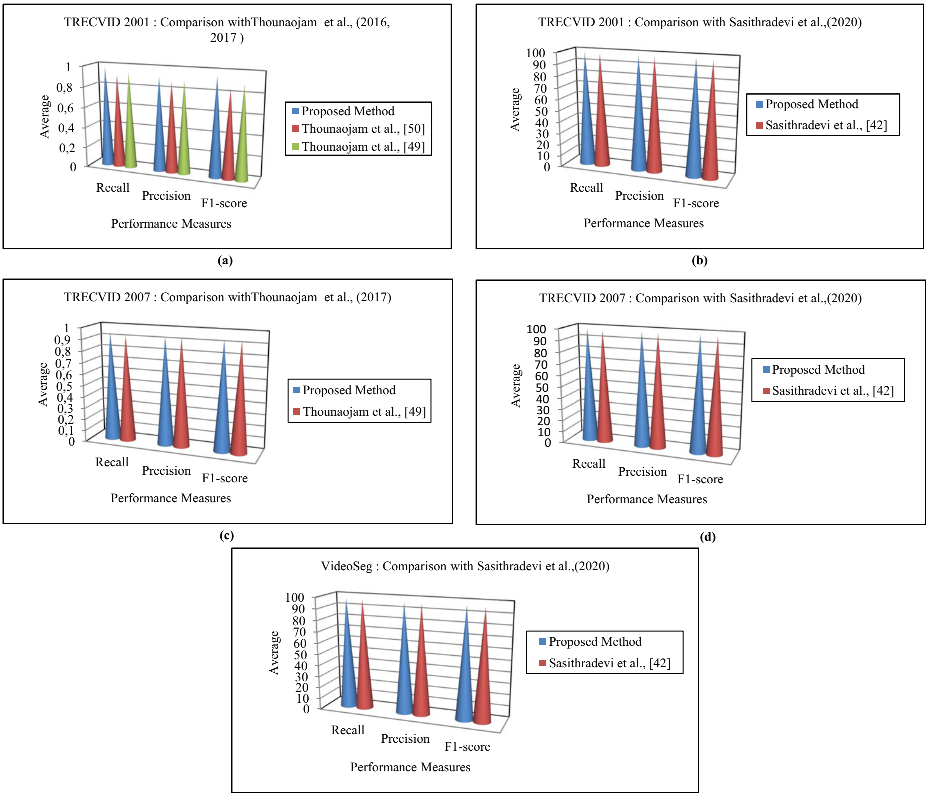
Video Name	Proposed Method			Sasithradevi et al., [42]		
	R	P	F1	R	P	F1
BG_2408	99.40	99.50	99.45	99.01	100	99.50
BG_9401	99.50	99.00	99.25	98.46	99.61	99.03
BG_11362	98.00	100	98.99	97.75	100	98.86
BG_14213	98.00	98.70	98.35	97.09	98.04	97.56
BG_34901	100	98.50	99.24	100.00	97.25	98.60
BG_35050	99.50	99.00	99.25	99.11	99.11	99.11
BG_35187	99.00	99.00	99.00	98.98	97.98	98.48
BG_36028	98.50	99.50	99.00	98.52	99.25	98.88
BG_36182	98.50	97.50	98.00	97.70	95.51	96.59
BG_36506	98.20	98.80	98.50	96.88	98.94	97.89
BG_36537	98.70	99.00	98.85	97.40	98.68	98.04
BG_36628	100	100	100	100	100	100
BG_37359	98.70	99.00	98.85	97.56	98.77	98.16
BG_37417	98.80	100	99.40	97.37	100	98.67
BG_37822	97.00	98.50	97.74	95.80	98.28	97.02
BG_37879	98.60	98.10	98.35	97.89	96.88	97.38
BG_38150	98.60	99.20	98.90	97.67	98.59	97.13
Average	98.76	99.02	98.89	98.07	98.64	98.29

Table 8 Performance comparison for abrupt shot transition on VideoSeg dataset

Video Name	Proposed Method			Sasithradevi et al., [42]		
	R	P	F1	R	P	F1
A	100	99.00	99.50	100	100	100
B	99.00	90.50	94.60	100	88.89	94.12
C	98.00	97.60	97.80	98.11	96.30	97.20
D	98.40	96.20	97.30	97.06	94.29	95.65
E	95.50	94.50	95.00	96.67	93.55	95.08
F	99.00	99.50	99.20	100	100	100
G	97.20	100.00	98.60	94.44	100.00	97.14
H	98.20	94.70	96.40	97.37	92.50	94.87
I	100	99.00	99.50	100	100	100
J	96.80	99.50	98.10	95.40	98.81	97.08
Average	98.21	97.05	97.60	97.91	96.43	97.11

4.2 Results on Gradual Transition Detection

Criteria set for gradual transition detection is established by using two local adaptive threshold values based on observation made on RSD difference graph (example shown in Fig. 11).


TRECVID 2001 : Comparison withThounaojam et al., (2016, 2017)

(a)

TRECVID 2001 : Comparison with Sasithradevi et al.,(2020)

(b)

TRECVID 2007 : Comparison withThounaojam et al., (2017)

(c)

TRECVID 2007 : Comparison with Sasithradevi et al.,(2020)

(d)

VideoSeg : Comparison with Sasithradevi et al.,(2020)

(e)

Fig. 13 Comparative results of abrupt shot transition detection

Table 9 Performance comparison for gradual shot transition with Lu and Shi (2013) and Thounaojam et al., (2016, 2017) on TRECVID 2001 dataset

Video	Proposed Method			Lu and Shi [30]			Thounaojam et al., [50]			Thounaojam et al., [49]		
	R	P	F1	R	P	F1	R	P	F1	R	P	F1
D2	0.871	0.844	0.857	0.935	0.725	0.817	0.806	0.833	0.819	0.870	0.794	0.830
D3	0.844	0.871	0.857	0.734	0.940	0.824	0.764	0.942	0.844	0.812	0.867	0.838
D4	0.855	0.723	0.783	0727	0.741	0.734	0.727	0.816	0.769	0.836	0.676	0.747
D6	0.889	0.909	0.899	0.844	0.927	0.884	0.844	0.864	0.854	0.867	0.907	0.887
Average	0.865	0.837	0.849	0.810	0.833	0.814	0.785	0.864	0.822	0.846	0.811	0.825

Threshold T_{NT} has already been discussed in previous section. Threshold T_{GT} has been set based on mean and standard deviation of computed *RSD* values of MCSH histograms of entire video. The frames falling in the range T_{NT} and T_{GT} are considered by gradual transition detection mechanism. Thus, formation of this sequence of frames excluding non transition frames reduces processing time and false transition during detection phase.

Since gradual transitions share a common behaviour, in the proposed method the patterns have been generated using relative frame feature difference computed for group frames using Eq. 11 of section 3.2.3. Thounaojam et al., [49] have observed that, the length of gradual transition ranges from 6 to 32 frames in the group for TRECVID videos.

Presuming this, the authors in the proposed method have made an empirical study for making observations of the patterns by grouping the frames in terms of 5, 10, 20, 25 and 30. This is achieved by plotting the relative frame difference values for the indicated group specifically. After making observation and thorough analysis, a frame group of 30 has yielded the expected behavioural pattern for ascertaining fade-in, fade-out and dissolve transitions (excluding wipe transition). The empirical setup and analysis has been performed to find detection rates to determine F1-score.

The analogy of the results with state-of-the-art approaches using benchmark datasets are detailed in Tables 9, 10 and 11 for TRECVID 2001 and TRECVID 2007 dataset. An additional comparison has been performed with the recent state-of-the-art algorithms proposed

Table 10 Performance comparison for gradual shot transition with Sasithradevi et al.,(2020) on TRECVID 2001 dataset

Video Sequence	Proposed Method			Sasithradevi et al., [42]		
	R	P	F1	R	P	F1
anni005	100	94.50	97.17	100	93.10	96.40
anni009	83.00	86.20	84.57	81.50	85.50	83.50
nad31	98.50	98.80	98.65	98.20	98.20	98.20
nad33	81.00	85.90	83.38	80.80	84.00	82.40
nad53	99.10	98.30	98.70	98.70	96.10	97.50
nad57	95.70	92.60	94.12	93.50	90.60	92.10
bor03	92.00	93.40	92.69	90.90	90.90	90.90
bor08	92.90	95.30	94.08	90.10	93.80	91.90
Average	92.78	93.13	92.92	91.71	91.53	91.61

Table 11 Performance comparison for gradual shot transition Thounaojam et al., (2016) on TRECVID 2007 dataset

Video Name	Proposed Method			Thounaojam et al., [49]		
	R	P	F1	R	P	F1
BG_3027	1.000	0.500	0.667	1.000	0.500	0.662
BG_28476	1.000	0.500	0.667	1.000	0.400	0.571
BG_36136	0.667	0.800	0.727	0.667	0.727	0.696
BG_37309	0.750	0.667	0.706	0.750	0.667	0.706
BG_37770	0.935	0.844	0.887	0.931	0.818	0.871
Average	0.870	0.662	0.731	0.869	0.622	0.702

by Sasithradevi et al., [42] for some of the video sequences of TRECVID 2001 and 2007 dataset in Tables 10 and 12 respectively After experimental study and comprehensive analysis, the proposed method exhibits significant progress when compared with state-of-the-art approaches as depicted graphically in the Figs. 14(a) to 14(d). The proposed method has taken care of non transition frames that depict camera/object motion and thus yielding significant progress in the achieved results.

As a summary, the empirical study on the dataset emphasize that the proposed method performs consistently well in complex environment, preserving good trade-off between recall and precision. During feature extraction from the frames, the spatial resolution is reduced by using blocks instead of pixels. Hence, this method is less sensitive to object and camera motion. However, the proposed method is threshold dependent and sensitive to complex camera and light variation affecting the overall performance of the algorithm. The algorithm

Table 12 Performance comparison for gradual shot transition with Sasithradevi et al.,(2020) on TRECVID 2007 dataset

Video Name	Proposed Method			Sasithradevi et al., [42]		
	R	P	F1	R	P	F1
BG_2408	94.50	100	97.17	95.00	100	97.44
BG_9401	94.00	99.00	96.44	93.33	100	96.55
BG_11362	97.00	98.00	97.50	100.00	100	100.00
BG_14213	77.20	99.00	86.75	75.00	100	85.71
BG_34901	98.50	98.10	98.30	98.36	98.36	98.36
BG_35050	96.70	97.30	97.00	93.75	100	96.77
BG_35187	99.50	94.00	96.67	100	100	100
BG_36028	99.00	93.40	96.12	100	92.00	95.83
BG_36182	–	–	–	–	–	–
BG_36506	96.30	96.10	96.20	92.31	100	96.00
BG_36537	84.00	99.00	90.89	83.33	100	90.91
BG_36628	98.00	92.70	95.28	100	90.91	95.24
BG_37359	85.70	99.00	91.87	83.33	100	90.91
BG_37417	92.00	93.20	92.60	91.67	91.67	91.67
BG_37822	99.00	92.70	95.75	100	90.91	95.24
BG_37879	99.50	99.80	99.50	100	100	100
BG_38150	98.00	99.10	98.55	100	100	100
Average	94.31	96.90	95.41	88.59	91.99	90.04

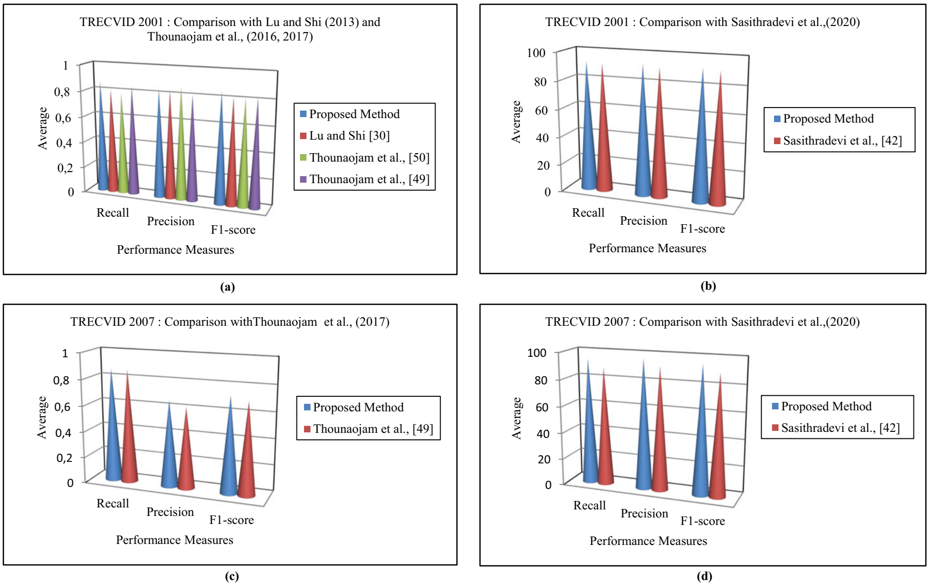


Fig. 14 Comparative results of gradual shot transition detection

efficiency limits with the identification of camera zooming and panning effects. The proposed method is computationally expensive than global histogram techniques. False detections may be encountered when frames of two different shots have similar histograms due to similar color values.

5 CONCLUSION

In this work, a simple and effective method to detect shot boundaries in videos is proposed. The method exploits the concept of fuzzy sets, Sobel gradient, block based MCSH histogram and RSD statistical measure to address the task of abrupt and gradual transition detection in videos. The algorithm applies RSD measure on each MCSH histograms with threshold mechanism to determine transitions. The experimental observations signifies that the discriminating strength of MCSH histograms using benchmark datasets have produced good results. Abrupt transition is identified by finding the difference between RSD measure of each MCSH histogram. Patterns for gradual transition are observed by plotting the relative difference of RSD values obtained from MCSH histogram in each group of frames. Experiments were performed on some of the benchmark datasets viz. TRECVID 2001, TRECVID 2007 and VideoSeg datasets. The efficacy of the proposed method is on par with some of the state-of-the-art SBD methods. As part of the future work, efforts will be made to reduce computational complexity of the algorithm. Also, there is a need to explore other feature descriptors to analyze the visual contents of the video frame. Advanced fuzzy logic can also be explored to better address uncertainty problem prevalent in most video frames.

Acknowledgments Sound and Vision video is copyrighted. The Sound and Vision video used in this work is provided solely for research purposes through the TREC Video Information Retrieval Evaluation Project Collection.

References

1. Abdesselam A (2013) Improving local binary patterns techniques by using edge information. Lecture Notes on Software Engineering 1(4):360
2. Abdulhussain SH, Mahmmod BM, Saripan MI, Al-Haddad SAR, Jassim WA (2019) Shot boundary detection based on orthogonal polynomial. Multimed Tools Appl 78(14):20361–20382
3. Abdulhussain SH, Ramli AR, Saripan MI, Mahmmod BM, Al-Haddad SAR, Jassim WA (2018) Methods and Challenges in Shot Boundary Detection: A Review. Entropy 20(4):214
4. Adjeroh D, Lee MC, Banda N, Kandaswamy U (2009) Adaptive edge-oriented shot boundary detection. EURASIP Journal on Image and Video Processing 2009(1):859371
5. Alshennawy AA, Aly AA (2009) Edge detection in digital images using fuzzy logic technique. World Acad Sci Eng Technol 51:178–186
6. Bezdek JC, Chandrasekhar R, Attikouzel Y (1998) A geometric approach to edge detection. IEEE Trans Fuzzy Syst 6(1):52–75
7. Bhaumik H, Bhattacharyya S, Nath MD, Chakraborty S (2016) Hybrid soft computing approaches to content based video retrieval: A brief review. Appl Soft Comput 46:1008–1029
8. Bhaumik H, Chakraborty M, Bhattacharyya S, Chakraborty S (2017). Detection of Gradual Transition in Videos: Approaches and Applications. In Intelligent Analysis of Multimedia Information (pp. 282-318). IGI Global
9. Camarena JG, Gregori V, Morillas S, Sapena A (2010) Two-step fuzzy logic-based method for impulse noise detection in colour images. Pattern Recogn Lett 31(13):1842–1849
10. Canny J (1986) A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell 6:679–698
11. Cirne MVM, Pedrini H (2018) VISCOM: A robust video summarization approach using color co-occurrence matrices. Multimed Tools Appl 77(1):857–875
12. Dadashi R, Kanan HR (2013) AVCD-FRA: A novel solution to automatic video cut detection using fuzzy-rule-based approach. Comput Vis Image Underst 117(7):807–817
13. Dimitrova N, Zhang HJ, Shahraray B, Sezan I, Huang T, Zakhor A (2002) Applications of video-content analysis and retrieval. IEEE multimedia 3:42–55
14. Ford RM, Robson C, Temple D, Gerlach M (2000) Metrics for shot boundary detection in digital video sequences. Multimedia Systems 8(1):37–46
15. Gygli M (2018). Ridiculously fast shot boundary detection with fully convolutional neural networks. In 2018 International Conference on Content-Based Multimedia Indexing (CBMI) (pp. 1–4). IEEE
16. Hannane R, Elboushaki A, Afdel K, Naghabhushan P, Javed M (2016) An efficient method for video shot boundary detection and keyframe extraction using SIFT-point distribution histogram. International Journal of Multimedia Information Retrieval 5(2):89–104
17. Hassanien A, Elgharib M, Selim A, Bae S H, Hefeeda M, Matusik W (2017). Large-scale, fast and accurate shot boundary detection through spatio-temporal convolutional neural networks. arXiv preprint arXiv: 1705.03281
18. Heng WJ, Ngan KN (2001) An object-based shot boundary detection using edge tracing and tracking. J Vis Commun Image Represent 12(3):217–239
19. Hu W, Xie N, Li L, Zeng X, Maybank S (2011) A survey on visual content-based video indexing and retrieval. IEEE Trans Syst Man Cybern Part C Appl Rev 41(6):797–819
20. Jadon RS, Chaudhury S, Biswas KK (2001) A fuzzy theoretic approach for video segmentation using syntactic features. Pattern Recogn Lett 22(13):1359–1369
21. Jain AK, Vailaya A, Wei X (1999) Query by video clip. Multimedia Systems 7(5):369–384
22. Ji Q G, Feng J W, Zhao J, Lu Z M (2010). Effective dissolve detection based on accumulating histogram difference and the support point. In 2010 First International Conference on Pervasive Computing, Signal Processing and Applications (pp. 273-276). IEEE
23. Jiang X, Sun T, Liu J, Chao J, Zhang W (2013) An adaptive video shot segmentation scheme based on dual-detection model. Neurocomputing 116:102–111
24. Küçüktunç O, Güdükbay U, Ulusoy Ö (2010) Fuzzy color histogram-based video segmentation. Comput Vis Image Underst 114(1):125–134
25. Lee MS, Yang YM, Lee SW (2001) Automatic video parsing using shot boundary detection and camera operation analysis. Pattern Recogn 34(3):711–719
26. Lee MH, Yoo HW, Jang DS (2006) Video scene change detection using neural network: Improved ART2. Expert Syst Appl 31(1):13–25
27. Li Z, Liu X, Zhang S (2016). Shot Boundary Detection based on Multilevel Difference of Colour Histograms. In 2016 First International Conference on Multimedia and Image Processing (ICMIP) (pp. 15-22). IEEE

28. Lian S (2011) Automatic video temporal segmentation based on multiple features. Soft Comput 15(3):469–482
29. Lopez-Molina C, De Baets B, Bustince H (2011) Generating fuzzy edge images from gradient magnitudes. Comput Vis Image Underst 115(11):1571–1580
30. Lu ZM, Shi Y (2013) Fast video shot boundary detection based on SVD and pattern matching. IEEE Trans Image Processing 22(12):5136–5145
31. Mahmoud M S(2017). Fuzzy Control, Estimation and Diagnosis: Single and Interconnected Systems. Springer.
32. Mas J, Fernandez G (2003). Video shot boundary detection based on color histogram. Notebook Papers TRECVID2003, Gaithersburg, Maryland, NIST, 15.
33. Melin P, Mendoza O, Castillo O (2010) An improved method for edge detection based on interval type-2 fuzzy logic. Expert Syst Appl 37(12):8527–8535
34. Pal SK, King RA (1983) On edge detection of X-ray images using fuzzy sets. IEEE Trans Pattern Anal Mach Intell 1:69–77
35. Perez-Ornelas F, Mendoza O, Melin P, Castro JR, Rodriguez-Diaz A, Castillo O (2015) Fuzzy index to evaluate edge detection in digital images. PLoS One 10(6):e0131161
36. Prasertsakul P, Kondo T, Iida, H, Phatrapornnant (2020) Camera operation estimation from video shot using 2D motion vector histogram. Multimed Tools Appl 1–24
37. Prewitt J M S (1970). Object enhancement and extraction picture processing and psychopictorics
38. Priya GL, Domnic S (2012) Edge strength extraction using orthogonal vectors for shot boundary detection. Procedia Technology 6:247–254
39. Rashmi B S, Nagendraswamy H S (2016). Abrupt Shot Detection in Video using Weighted Edge Information. In Proceedings of the International Conference on Informatics and Analytics (p. 69). ACM.
40. Rashmi B S, Nagendraswamy H S (2016). Video shot boundary detection using midrange local binary pattern. In Advances in Computing, Communications and Informatics (ICACCI), 2016 International Conference on (pp. 201-206). IEEE
41. Rashmi BS, Nagendraswamy HS (2018) Effective Video Shot Boundary Detection and Keyframe Selection using Soft Computing Techniques. International Journal of Computer Vision and Image Processing (IJCVIP) 8(2):27–48
42. Sasithradevi A, Roomi SMM (2020) A new pyramidal opponent color-shape model based video shot boundary detection. J Vis Commun Image Represent 67:102754
43. Shahraray B (1995) Scene change detection and content-based sampling of video sequences. In Digital Video Compression: Algorithms and Technologies 1995 (Vol. 2419, pp. 2-14). International Society for Optics and Photonics
44. Shen J, Tao D, Li X (2008) Modality mixture projections for semantic video event detection. IEEE Transactions on Circuits and Systems for Video Technology 18(11):1587–1596
45. Sobel I, Feldman G (1968). A 3x3 isotropic gradient operator for image processing. a talk at the Stanford Artificial Project in 271-272
46. Stanchev P, Green D Jr, Dimitrov B (2003) High level color similarity retrieval. International Journal of Information Theories and Applications 10(3):363–369
47. Tab F A, Shahryari O K (2009). Fuzzy edge detection based on pixel's gradient and standard deviation values In Computer Science and Information Technology, 2009. IMCSIT'09. International Multiconference on (pp. 7-10). IEEE
48. Tao D (Ed.) (2009). Semantic mining technologies for multimedia databases. IGI Global.
49. Thounaojam DM, Bhadouria VS, Roy S, Singh KM (2017) Shot boundary detection using perceptual and semantic information. International Journal of Multimedia Information Retrieval 6(2):167–174
50. Thounaojam DM, Khelchandra T, Singh KM, Roy S (2016) A genetic algorithm and fuzzy logic approach for video shot boundary detection. Computational intelligence and neuroscience 2016:14
51. Tizhoosh H R (2002). Fast fuzzy edge detection. In Fuzzy Information Processing Society, 2002. Proceedings. NAFIPS. 2002 Annual Meeting of the North American IEEE 239-242
52. Torre V, Poggio TA (1986) On edge detection. IEEE Trans Pattern Anal Mach Intell 2:147–163
53. VideoSeg n.d.. http://www.site.uottawa.ca/~laganier/videoseg/
54. Wu G, Liu L, Guo Y, Ding G, Han J, Shen J, Shao L (2017). Unsupervised deep video hashing with balanced rotation. IJCAI
55. Wu B, Xu L (2014) Integrating bottom-up and top-down visual stimulus for saliency detection in news video. Multimed Tools Appl 73(3):1053–1075
56. Yuan J, Wang H, Xiao L, Zheng W, Li J, Lin F, Zhang B (2007) A formal study of shot boundary detection. IEEE transactions on circuits and systems for video technology 17(2):168–186
57. Zadeh LA (1965) Information and control. Fuzzy sets 8(3):338–353

58. Zhang D, Lei W, Zhang W, Chen X (2019) Shot boundary detection based on block-wise principal component analysis. Journal of Electronic Imaging 28(2):023029
59. Zhang D, Qi W, Zhang H J (2001). A new shot boundary detection algorithm. In Pacific-Rim Conference on Multimedia (pp. 63-70). Springer, Berlin, Heidelberg
60. Zheng J, Zou F, Shi M (2004). An efficient algorithm for video shot boundary detection. In Intelligent Multimedia, Video and Speech Processing, 2004. Proceedings of 2004 International Symposium on IEEE 266-269

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.