Artificial Intelligence Review (2024) 57:104
https://doi.org/10.1007/s10462-024-10742-1

Video shot-boundary detection: issues, challenges and solutions

T. Kar[1] · P. Kanungo[2] · Sachi Nandan Mohanty[3] · Sven Groppe[4] · Jinghua Groppe[4]

Accepted: 24 February 2024

Abstract
The integration of high data transmission rates and the recent digital multimedia technology, paves the way to access a huge amount of video over the internet, in seconds. Additionally, uploading videos to different websites is no more confined to expert software professionals resulting in duplication of video data which led to exorbitant growth of multimedia information in cyberspace in a short span of time. This necessitates the development of efficient data management techniques including storage, searching and annotation mechanism. Automatic shot boundary detection is considered to be the first and foremost step towards such management. It is a booming area of research gaining attention in the domain of image processing, computer vision and pattern recognition. In this review paper, we present a detailed description of the methods and algorithms of shot boundary detection, reported in the last two decades. This review shows that using multiple features performs well in comparison to using only a single feature in the shot boundary detection problem although it leads to higher complexity. The major sources of disturbance in the boundary detection are the sudden illumination variation and presence of high motion in the video. An adaptive threshold outperforms a single global threshold in the boundary detection problem and the threshold requirement can be avoided through learning based strategies at the cost of larger training data and higher computation time. Moreover the present review includes a critical analysis of relative merits and demerits of existing algorithms and finally opens promising research directions in the area.

Keywords Shot boundary detection · Abrupt shot transition · Gradual shot transition · Video indexing · Video retrieval

1 Introduction

Multimedia is a combination text, graphics, image, audio, video and animation (Abdulrahaman et al. 2020), where video is considered to be the most popular information media around us (Chakraborty et al. 2022). The great advances in multimedia and information technologies facilitated easy availability of low-cost video acquisition tools, portable and large capacity storage devices and above all easy to use video editing software. This generates rich information content and huge volume of multimedia data (Hameed et al. 2021).

Extended author information available on the last page of the article

The spontaneous rise of the volume of video data in current scenario demands efficient video data management systems. This not only provides effective data storage, but also facilitate the users for quick searching and accessing the pertinent data from the huge volume of stored data. Early prototype of image searching rely on a simple text-based search paradigm, which is a time consuming process, since it exploits some identifiers which ignore the available information in the image itself (Parmar and Angelides 2015). This can be solved by Google's image based search,[1] Yahoo's Image based Search[2] or by some other similar platforms. Accessing the stored multimedia data is inefficient and inconvenient due to the unstructured characteristics of multimedia data (Cunhaa et al. 2021). The typical functionalities of the video data management system include efficient modeling, organizing and storing of large volume of video data, designing interactive tools for presenting video information, formulation of flexible queries, data embedding and manipulation, simpler, quicker and effective solution for browsing and retrieving the desired data. Traditional techniques of multimedia data retrieval (Pickering and Rüger 2003) relies on keyword based indexing, retrieval and manual annotations (Lorenzo et al. 2017). However, these are highly inefficient in terms of cost and processing time for large data size. But a more promising way to access the multimedia data is by their content. Video is the most important multimedia data and is the most dominant data type available on the net. Additionally, it devours a lion's share in terms of storage space. Moreover, human brain tends to capture most of its surrounding information visually and can process these visual data more conveniently at a much faster rate than text or any other form of information. However, it can not be readily manipulated as text. Irrespective of the video formats used for representation, it is quite inefficient and time consuming to operate a video by considering all the frames. For efficient management the large video is divided into small manageable segments for further processing. This is also known as temporal video segmentation.

The main contributions of our survey are as follows.

- All the components of the shot boundary detection (SBD) taxonomy are explained in a simplified and well-organized hierarchical style. Further, the correlation between these components are summarized for better clarity.
- The state-of-the-art methods in the field of SBD using different threshold specifications, detection of different types of shot boundaries, and use of additional and apriori information for performance improvement of the SBD are analyzed.
- The relative strengths and weaknesses of the different SBD approaches are summarized. Our work primarily turned the spotlight on the review of recent approaches as a supplement to the previously reported surveys.
- A detailed discussion on the future directions in SBD are also unlocked for the whole benefit of the research community.
- Attention has been paid to the potential of various feature space and metrics that made the continuity/discontinuity values independent of illumination variation and motion. Our contributions clearly demarcate between our survey and the existing surveys on video SBD. To the best of our knowledge, our survey is the most recent and extensive in the field.

[1] http://images.google.com/

[2] http://images.search.yahoo.com

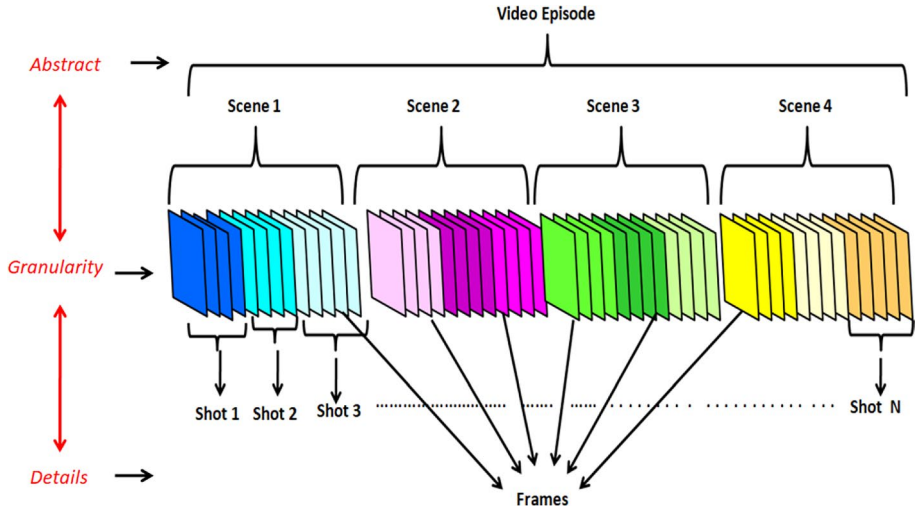

Fig. 1 Hierarchical structure of a video

The rest of the paper is arranged as follows. Related background is summarized in sect. 2. Different applications of SBD are presented in sect. 3. Section 4 and 5 deal with evaluation metrics and popular datasets used in SBD. The review of different SBD frameworks with feature and algorithm based classification is described in sect. 6. Section 7 describes various issues and challenges in SBD. Finally, sect. 8, presents a concluding remark with an overall discussion of SBD, the state-of-the-art performance and future perspectives.

2 Related background

A video is a collection of video frames typically played at frame rates of 25 or 30 fps. The easy availability of inexpensive and miniaturized digital storage devices such as high capacity pen drives, hard disks and DVDs to the common man are the primary driving force for the dramatic increase of digital video in contrast to traditional analog counterpart and thus there is a huge increase in the content of the video databases which received enormous attention in recent years. However, it is the most complex one to deal with. The gigantic volume, large variation in the nature and the complex spatio temporal attributes of the video make it difficult to manage. The sudden rise in video sources, has an intense craving to build a content based video database system that can facilitate smart and intelligent means of video searching and retrieval.

With granularity from high to low, the hierarchy of segmentation starts with a video episode, a scene, a shot and a frame. A typical video modeled as a hierarchical structure is presented in Fig. 1.

- A *shot* is a convenient and meaningful segment representing an unbroken sequence of frames, picked up during continuous rolling of a single camera indicating a continuous action in time and space.
- A *scene* is defined as a collection of adjacent shots having semantic similarity in object and usually representing a complete sequence of actions.

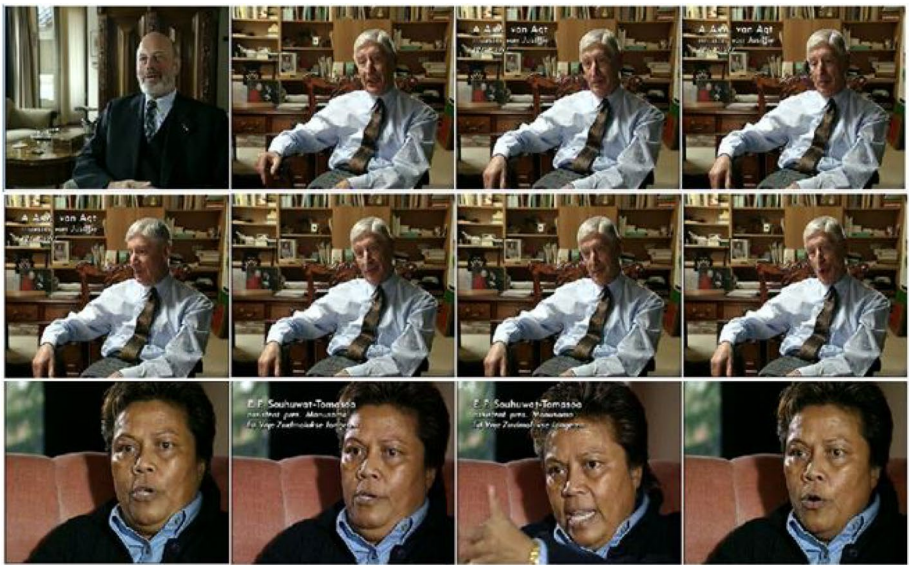

Fig. 2 Frames from the video BG38655 [8] to illustrate AST

– An *episode* can be defined as a group of scenes describing a common thread of action.

Identification of the syntactic structure of video is the first and foremost step in any content based video management system that is achieved through complete SBD. The target of SBD is to temporally partition a video into its basic units (shots) which is delivered for further analysis. The fringes of a shot, is either an abrupt shot transition(AST) or gradual shot transition (GST) which follows another shot.

2.1 Abrupt shot transition

"*Abrupt shot transition(AST)*" indicates a spontaneous change of temporal visual information, where successive shots are isolated by one frame. Figure 2 shows an example of AST, where the first frame of one shot is the immediate successor of the last frame of the previous shot.

2.2 Gradual shot transition

When the videos are re tailored by professionals to include special effects and computer graphics to make it visually more appealing, the type of transition created is known as "*gradual shot transitions(GST)*". It is an artificial effect, stretched over several frames varying between typically seven to forty frames. The frames involved during the course of transition, hold information from both the shots. Further GSTs are classified as fade, dissolve and wipe transition as illustrated in Fig. 3. Figure 3a, b, c and d illustrate a sample fade in, fade out, dissolve and wipe pattern respectively. Dissolves in movies signify a passage of time which is common in movies or other video material.

2.2.1 Fade-in transition

The type of transition where, the intensity values of next shot gradually appear from a frame with fixed intensity is known as fade in transition.

2.2.2 Fade-out transition

The type of transition where, the intensity values of a frame are progressively replaced by a darker frame, is known as fade out transition.

2.2.3 Dissolve transition

In dissolve type transition, the intensity values of the pixels in one frame progressively recede from one shot and the intensity values of the pixels in subsequent frames gradually emerge from the next shot, where the superposition of the shots are partially visible.

2.2.4 Wipe transition

A wipe type of transition is evolved due to fractional replacement of one shot by those in the next shot until the current shot is completely replaced by the next shot, such that the region visible from the current scene gradually diminishes while the region visible from the next scene progressively increases following an organized spatial pattern. Such transition involves evolution of single or multiple image boundary lines with varying shapes, direction of motion and speed of motion.

2.3 Basic SBD module

The basic idea behind any SBD approach is to identify the presence and position of discontinuities of the flow of visual content in a video. Irrespective of the scheme of detection process, the heart of the process are three folds (Yuan et al. 2007) i.e. a suitable representation of the visual content through feature extraction, evaluation of visual content similarity and finally setting up a threshold manually or automatically for final classification. However, machine learning (ML) based techniques rely on feature extraction followed by learning based classification. A basic shot bondary detection module is illustrated in Fig. 4.

2.4 Representation of visual content: the feature extraction phase

Feature selection is the most important part of SBD module that plays an important role in overall performance of the SBD system. Better the feature, better the emphasis at the point of transition. The features used to represent the frame-content should not be much influenced by changes in the content due to illumination variation and COM and must be capable enough to discriminate between the within shot frames and the transition frames. Global features are often incapable to distinguish between inter-shot and intra-shot frames while local features are sensitive to local variations within a shot. Thus, a

Fig. 3 **a** Frames from the video NAD57 [9] to illustrate Fade in effect, **b** Frames from the video NAD58 [9] ▶ to illustrate Fade out effect, **c** Frames from the video BG38655 [8] to illustrate Dissolve effect **d** Frames to illustrate wipe effect

meticulous selection of features is highly desirable to fulfill the diverging requirements of the characteristics of local and global features.

2.5 Formulating the similarity measure: the evaluation of similarity between the frames

Feature selection for representation of the visual content, is followed by formulation of a similarity measure to represent the similarity between the frames. Ideally, the similarity measure captures high values for within shot frames and low values at transition points. The various similarity measures also known as distance measures addressed in the literature are manhattan distance/ city block distance/ the sum of absolute feature difference or $L1$ Norm (Fan et al. 2017), Euclidean Distance or $L2$ Norm (Li et al. 2016), Sum of squared distance/Euclidean Norm (Bendraou 2017), Bhattacharya Distance (Dutta et al. 2016; Hassanien et al. 2017), Mahalanobis Distance (Dutta et al. 2016), Likelihood ratio(Bendraou 2017), χ^2 square Distance (Nagasaka and Tanka 1991), Histogram Intersection (Ngo et al. 2002; Janwe and Bhoyar 2013), Kullback Leibler Divergence (Li et al. 2015), Correlation Coefficient (Warhade et al. 2013; Porter et al. 2003), Cosine Distance (Lu and Shi 2013), "ECR" (Lienhart 2001; Jacobs et al. 2004), Hausdorff distance (Kim and Park 2002), Mutual Information (Cernekova et al. 2006) and Joint Entropy (Cernekova et al. 2006).

Although different similarity measures are found in the literature, but the choice of the similarity measure is of decisive importance.

2.6 Threshold evaluation

Threshold is another factor which highly impacts the performances of the algorithm. Setting considerably high threshold may result miss detection while setting a low value of threshold leads to false detection. Thus threshold has to be selected judiciously to fit long and wider range of videos. Threshold can be either an empirical or global threshold that is set manually or generated through some automatic process. The automatic threshold can be either global or adaptive (Ngo et al. 2005) in nature. Since the video content undergoes dramatic variation from shot to shot and also from video to video. Hence a single global threshold won't be a suitable choice to evaluate all the transitions in a video. In such scenarios adaptive threshold (Warhade et al. 2011) works well that is obtained considering local feature values in some window length which has to be set judiciously. Adaptive threshold (Kar and Kanungo 2018) with the ASWD (Kar and Kanungo 2018) similarity measure for video 'anni 006' is illustrated in Fig. 5. However, adaptive threshold which is evaluated based on different parameters within certain window, the difficulty arises for selection of the window length.

Significant development has been made for transition detection in videos over last two decades. Most of the existing reviews are neither extensive and nor give a comparative study along with future scope of improvement in the existing form. Inasmuch as, it is required to have a novel and extensive review and in depth discussion of the state-of-the-art algorithms in this field. In the light of the past works, our work may be a crux, that presents

(a)

(b)

(c)

(d)

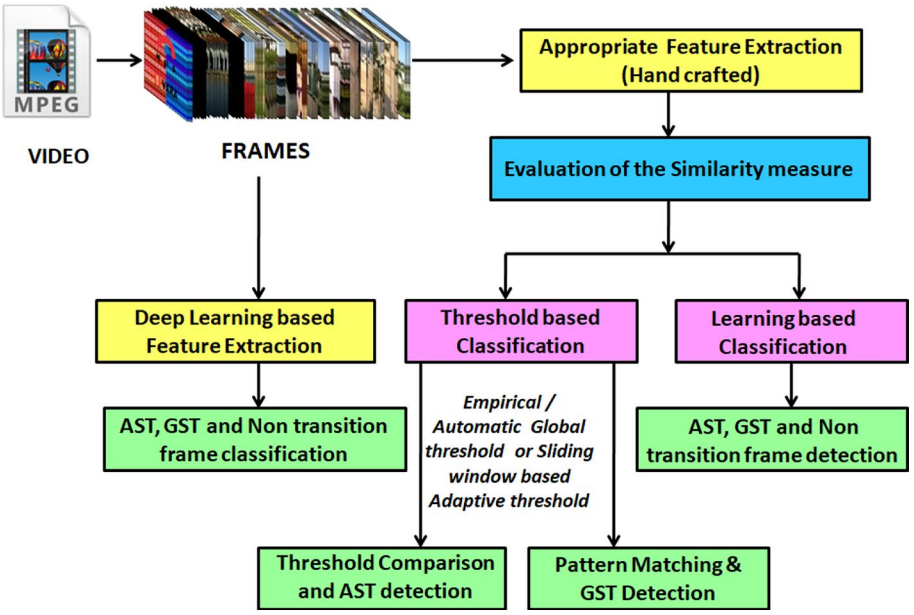

Fig. 4 A basic SBD module

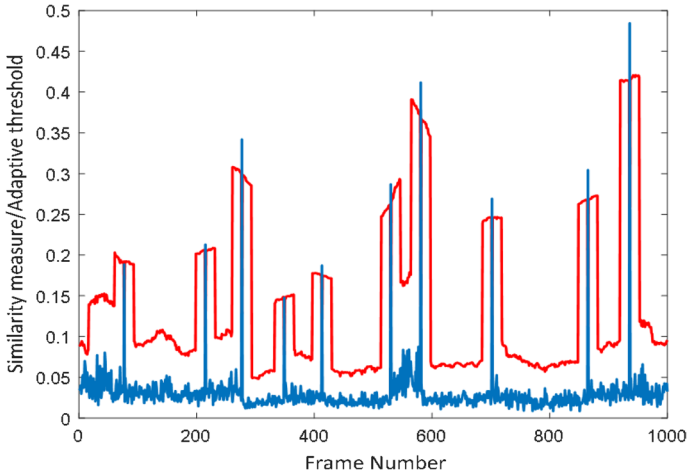

Fig. 5 Illustration of adaptive threshold (Kar and Kanungo 2018)

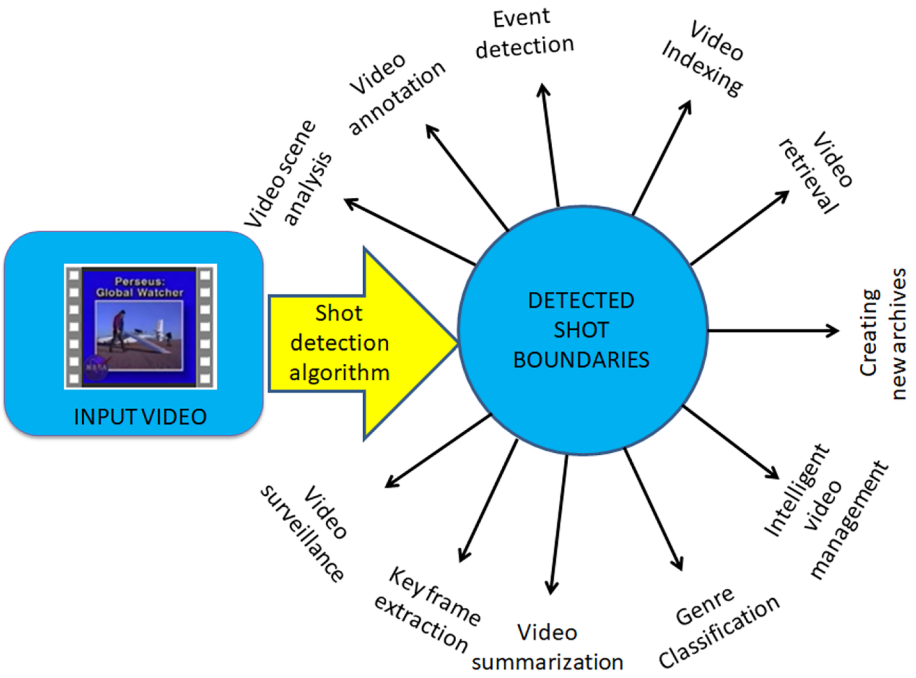

Fig. 6 Possible applications of SBD

a comprehensive review of recent breakthrough and unlocked future directions in the field of SBD.

3 Application scenarios

The primary goal of automatic detection of temporal structure in videos is to extract the segments needed for further processing for subsequent applications.

Various possible application scenarios that benefit from the extraction of temporal video units include: content-based video indexing (Snoek and Worring 2005; Liao et al. 2011), classification (Brezeale and Cook 2008) and retrieval (Iyer et al. 2016; Hameed et al. 2021) for quick browsing of video folders, keyframe extraction(Liu et al. 2003; Rasheed and Shah 2005; Gianluigi and Raimondo 2006; Sheena and Narayanan 2015; Shi et al. 2017; Bendraou 2017; Sun and Zhou 2011; Hannane et al. 2016; Bommisetty et al. 2021; Nandini et al. 2020), video summarization [43], (Ejaz et al. 2014; Vinicius and Pedrini 2017; Bajaj and Sharma 2016; Ranjan and Agrawal 2016; Mahapatra et al. 2018; Kumar and Shrimankar 2018; Kumar et al. 2022; Kumar and Shrimankar 2019; Kumar 2019, 2021), genre classification, You et al. (2010); Bhoraniya and Ratanpara (2017); Karthick et al. (2015), event detection (You et al. 2010), video scene analysis (Helm and Kampel 2019),[58] and content analysis (Hanjalic 2004), video annotation (Tong et al. 2015; Lorenzo et al. 2017), video surveillance (Chakraborty et al. 2018), video-on-demand (Rasheed and Shah 2005). But for effective implementation of any subsequent high level video analysis applications, complete video SBD

is a must. There are numerous applications which can be benefited from the extraction of temporal video units. The different applications are pictorially shown in Fig. 6.

4 Evaluation metrics

The two potential metrics "accuracy" and "processing time" are used for evaluation of the performance of the SBD algorithms. However, it is difficult to address accuracy and computational cost simultaneously through a single framework as most of the cases one metric can only be improved at the cost of the other one. This is due to the fact that increasing the detection accuracy may incur additional computational overhead.

4.1 Accuracy

For a test video the number of transitions detected correctly (N_C) are called as true positives and the number of transitions, wrongly detected by the algorithm (N_F) are known as false positive respectively. The discrepancy between the true detected transitions and the ground truth number of transitions represent transitions missed by the algorithm, N_M.

The standard performance measures (Gargi et al. 2000) indicative of detection accuracy are Recall(R), Precision(Pr) and the $F1$ measure and are used for evaluating the performances of any SBD algorithm. Mathematically

$$R = \frac{N_C}{N_C + N_M} \tag{1}$$

$$Pr = \frac{N_C}{N_C + N_F} \tag{2}$$

$$F1 = \frac{2R \times \mathrm{Pr}}{R + Pr} \tag{3}$$

Ideally R, Pr and $F1$ values are in percentage scale which represent that, all transitions are detected correctly without any miss or false detection.

4.2 Processing time

Processing time is another essential parameter for performance evaluation of the SBD algorithms (Lefevre et al. 2003). This is a representation of time required for execution of various mathematical and logical operations involved in the algorithm. However, very few papers focus on processing time for evaluation of the SBD algorithm due to the fact that the SBD algorithms generally not used for real time application.

5 Evaluation dataset

The dataset used for evaluating SBD algorithms should be chosen carefully and judiciously. Standard dataset such as TRECVid (Smeaton et al. 2010) are preferable for evaluation due to its representativeness, universal availability and availability of ground truth information. Any arbitrary video data used in the evaluation process cannot be accessed always due to domain related issues. Moreover, they may not include various challenges for evaluation leading to biased results. TRECVid (Smeaton et al. 2010) was established in 2003 for evaluation and benchmarking of various SBD tasks via open, metric-based evaluation. TRECVid is a source of large-scale collection of benchmark test videos that confirms the compatibility between description interfaces for video contents towards facilitation of subsequent highlevel tasks. It is co-sponsored by the National Institute of Standards and Technology (NIST). TRECVid has a bunch of datasets starting from 2001 to 2007 to meet various challenges in SBD process for making fair comparisons of algorithms. These contain a total of 43,33,153 number of frames which includes 24,423 number of transitions. Out of these nearly 64% are of ASTs and remaining are GST(Hassanien et al. 2017). Besides TRECVid dataset, VIVA research lab videos [65],(Dadashi and Kanan 2013) of Carleton university being available publicly, can also be used for SBD evaluation. Some authors (Baraldi et al. 2015) used the publicly available RAI dataset [68] for SBD tasks which specifically contains documentaries and talk shows. The description of different standard datasets are listed in Table 2. Beyond this, any synthetic data can also be generated from any video by interleaving randomly selected frame sequences from different videos(Gygli 2018). This may not be a suitable choice as the transition created by this process violates its definition due to the random generation process. However, it can be used for training a CNN Model.

6 Literature review

SBD is a well recognized problem having a rich and long history in the field of content based video browsing, retrieval and analysis. However, there is still room available for the researchers to achieve universal and robust models for SBD ideal for videos of any modality and complexity. The essential requirement for revealing the structure of the higher level video content is to find the boundary between shots treated as the most elementary step in the process. In this review we present an extensive review of the SBD techniques over more than twenty years.

Broadly all SBD algorithms can be classified either as uncompressed domain techniques or compressed domain techniques. However, it is rather difficult to review the existing literature to classify the algorithms either based exclusively on features or the type of approach because in most of the cases either different features or different techniques are combined to achieve higher performance of the SBD task. So overlapping of the features and algorithms is a common in SBD classification task and it is impossible to segregate the techniques of SBD to give an idea of number of literature available exclusively under a single feature or a single algorithm through graphical presentation. Due to long and rich history of SBD, a comprehensive review of the state-of-the-art methods is beyond the scope of any acceptable length paper. In order to maintain quality, we mostly focused our attention to recent and top tier journal and conference papers. Therefore, we wholeheartedly

apologize to the authors whose works could not be considered in this review. For existing reviews on SBD, readers are encouraged to refer to the articles (Koprinska and Carrato 2001; Lienhart 2001; Hanjalic 2002; Lefevre et al. 2003; Cotsaces et al. 2006; Yuan et al. 2007; Smeaton et al. 2010; Fabro and Böszörmenyi 2013; Pal et al. 2015; Singh and Aggarwal 2015; Sengupta et al. 2015; Abdulhussain et al. 2018). The summary of video SBD surveys since 2001 found in the literature is given in Table 1. The main objective of this review is to offer a comprehensive review of different SBD techniques and to present some extent of taxonomy, a high level aspect and formulation based on modeling, detection methods, benchmark data set, evaluation metrics and performance measures. The intention is to help the readers to have a clear understanding of wide variety of existing strategies and to identify current research issues. This section discusses the literature review of existing SBD techniques.

6.1 Classification of SBD approaches

SBD can be addressed either in the compressed domain or in the uncompressed domain. Compressed domain approaches are comparatively faster than uncompressed counterpart resulting a faster processing at the cost of lower performance. Compressed domain approaches are less dependable specifically for videos characterized by high COM or camera operation (Hu et al. 2011; Gao and Ma 2014; Shekar and Kirsch Uma 2015). Nonetheless, the performances of compressed domain approaches highly rely on the video compression standards. However, uncompressed domain approaches are gaining more attention and popularity for appreciable performance due to inclusion of comparatively higher visual content (Grana and Cucchiara 2007). In this section we tried to classify the SBD approaches based on some basic features as well as algorithms.

6.2 Intensity based techniques

The most basic feature is the frame intensity(Lu and Shi 2013). In the basic intensity based techniques consecutive frames are compared based on pixel intensity values at corresponding points. The difference between intensity values between successive frames are compared (Zhang et al. 1993) for final detection of the AST. Kundu in 2012 (Kundu and Mondal 2012) proposed an automatic thresholding scheme to address the false detection due to motion and illumination variation for identifying abrupt shot transition based on statistics of pixel intensity difference value.

However, intensity based techniques are highly sensitive to surrounding disturbances. The disturbances (camera/object motion (COM), illumination variation) tend to change the intensity values even in the absence of actual transition creating an illusion of transition.

6.3 Edge and gradient feature based techniques

In the literature, edge (Canny 1986) feature has been identified as a potential candidate for SBD. The most popular and the oldest edge based feature proposed is the edge change ratio(ECR) (Lienhart 2001; Jacobs et al. 2004; Zabih et al. 1995) which represents the percentage of edge pixels that enter and leave between successive frames. Yoo et al. (2006) proposed a technique exclusively for GST detection. The authors have characterized the GST by variance distribution of edge information in the frame sequences. However, it has

Table 1 Summary of video SBD surveys since 2001

Reference	Source	Content
Koprinska and Carrato (2001)	Elsevier	Overview of existing techniques of video segmentation in uncompressed and compressed domain along with recognition of camera operation
Lienhart (2001)	IJIG	Discussion on different core concepts for detection of ASTs, fades and dissolves
Hanjalic (2002)	IEEE	Detailed analysis of the existing SBD problem and proposed a novel statistical SBD detector based on minimization of the error probability.
Lefevre et al. (2003)	Elsevier	A detailed review of temporal video segmentation methods based on pixel, histogram, block, motion, in uncompressed domain including complexity analysis of the methods
Cotsaces et al. (2006)	IEEE	Common basic techniques of SBD and condensed video representation
Yuan et al. (2007)	IEEE	Exhaustive review of the existing framework and a new SBD system based on graph partitioning model
Smeaton et al. (2010)	Elsevier	Overview of the TRECVid SBD task, important approaches towards SBD, and a performance comparison for SBD algorithms in 2005
Fabro and Böszörmenyi (2013)	Springer	Overview of the existing approaches based on visual, audio, text and hybrid strategies. Brief discussion on challenges and possible applications of SBD
Pal et al. (2015)	Springer	Discussion on few SBD algorithms including histogram, DCT and motion vector based feature along with their merits and demerits
Singh and Aggarwal (2015)	Springer	Brief review of all the major and latest contributions in SBD
Sengupta et al. (2015)	IEEE	Different approaches to SBD problem
Abdulhussain et al. (2018)	Entropy	Extensive review of SBD approaches, challenges in SBD and a comparative study of merits and demerits of each approach
Chavate et al. (2021)	IEEE	Review of recent SBD approaches, challenges in SBD and a comparative study of different approaches

Table 2 Summary of video dataset used for Evaluation of video SBD

Name of the dataset	Description	Source	Used in Research work
Trec 2001	Documentary	NIST[9]	Sasithradevi et al. (2018), Mondal et al. (2017),Nandini et al. (2020), Kar and Kanungo (2018),Adjeroh et al. (2009), Sasithradevi and Roomi (2016), Sasithradevi and Roomi (2020),Nandini et al. (2021),Shekar and Kirsch Uma (2015), Idan et al. (2021), Sun and Wan (2014),Lu and Shi (2013), Youssef et al. (2017),Tong et al. (2015),Thounaojam et al. (2016), Mohanta et al. (2012),Li et al. (2009),Rashmi and Nagendraswamy (2021), Chakraborty et al. (2022), Benoughidene and Titouna (2022), Raja et al. (2022), Kar and Kanungo (2023), Yan et al. (2022), Zhou et al. (2021), Sasithradevi and Roomi (2022), Rashmi and Nagendraswamy (2018), Yuan and Zhang (2023), Sasithradevi and Nirmala (2023),Chakraborty et al. (2021)
Trec 2002	Documentary	NIST[103]	Sasithradevi et al. (2018),Kar and Kanungo (2018),Youssef et al. (2017)
Trec 2003	8 ABC/CNN	NIST[104]	Wang et al. (2014),Depalov et al. (2006), Wu et al. (2019)
Trec 2004	ABC/CNN	NIST[104]	Sasithradevi and Roomi (2020), Wu et al. (2019)
Trec 2005	8 Broadcast news and 4 NASA	NIST[104]	Idan et al. (2021),Youssef et al. (2017),Yuan et al. (2007), Wu et al. (2019),Ramli et al. (2019),Yuan and Zhang (2023)
Trec 2006	Broadcast news videos	NIST[104]	Idan et al. (2021),Dadashi and Kanan (2013), Wu et al. (2019),Ramli et al. (2019)Yuan and Zhang (2023)

Table 2 (continued)

Name of the dataset	Description	Source	Used in Research work
Trec 2007	Sound & Vision	NIST[104]	Mondal et al. (2017),Kar and Kanungo (2018), Sasithradevi and Roomi (2020), Nandini et al. (2021),Tippaya et al. (2014), Shekar and Kirsch Uma (2015), Idan et al. (2021),Lankinen and Kämäräinen (2013),Bhaumik et al. (2017), Lakshmi Priya and Domnic (2014), Baber et al. (2013),Wang et al. (2014), Tang et al. (2018),Gushchin et al. (2021),Tang et al. (2018), Rashmi and Nagendraswamy (2021), Benoughidene and Titouna (2022), Chakraborty et al. (2022) Sasithradevi and Roomi (2022),Zhou et al. (2021), Wu et al. (2019), Kar and Kanungo (2023),Ramli et al. (2019),Yuan and Zhang (2023), Chakraborty et al. (2021)
The MoCA Project	Sound & Vision	University of Mannheim University[117]	Dadashi and Kanan (2013),Fang and Jiang (2006)
VIVA/ VideoSeg dataset	10 videos: Cartoon, Horror, Action Drama, Commercial,Fiction	Carleton University[65]	Dadashi and Kanan (2013),Rashmi and Nagendraswamy (2021), Sasithradevi and Nirmala (2023)
RAI dataset	Broadcast videos: DocumentariesTalk shows	Rai Scuola Video archive[68]	Gygli (2018),Tang et al. (2018), Gushchin et al. (2021), Tang et al. (2018),Baraldi et al. (2015), Bouyahi and Ayed (2020), Benoughidene and Titouna (2022), Sasithradevi and Roomi (2022)
ClipShots	Collection of more than 20 categories of video, including sports, TV shows, animals, short videos with complex hand-held camera vibrations, large object motions and occlusion	Rai Scuola video archive[120]	Ming et al. (2021), Yan et al. (2022)
MovieShots2	15K shots from 282 movie clips	Collection created by the authors	Jiang et al. (2022)
MovieShots2	21K shots from 150 movie clips	Generated by the authors by grouping over 270K shots from 150 movies	Rao et al. (2020)

limited performance under high motion and substantial change in the video content within the shot and has higher processing time. Liang et al. (2005) proposed an enhanced SBD approach by superimposing text information along with edge information for an efficient SBD. Motivated by the fact that variations such as object motion, local illumination change in the image frame and some camera operations are local in nature and low level decomposition is dominated by global information whereas high level decomposition is dominated by localised information, Adjeroh et al. (2009) proposed a fast, robust and efficient SBD using several multilevel edge-based features and adaptive threshold suitable for real-time applications. Chan and Wong (2011), presented a GA based searching of the shot boundaries using the ECR feature for SBD. It has low performance for camera operations and miss detection were caused due to frame sub sampling. Dutta et al. (2016) obtained the variation of the content of the frames with respect to a reference frame for AST detection using edge based feature. Further, they improved the algorithm by finding the slope of the linear approximation of change in similarity values followed by least square regression and post processing method to reduce the false detection.

6.4 Histogram based techniques

Histogram is an indicative of distribution of intensity values in an image. But histogram representation ignores the spatial distribution of pixels. Hence histogram is comparatively a better choice than intensity based feature to exhibit higher invariance to local or small global motion. The basic gray histogram based similarity measure (Zhang et al. 1993) is given by

$$SM_{HD_{t,t+1}} = \sum_{i=1}^{256} |H_t(i) - H_{t+1}(i)| \tag{4}$$

where, H_t is the frame histograms at t^{th} time instant. Early histogram based approaches include histogram difference in different color space (Gargi et al. 2000; Gong and Liu 2000). Later stage algorithms merged with other approaches to produce superior performance. Qian et al. (2006) proposed a fade and flash light detection based on histogram using accumulation of histogram difference. Cernekova et al. (2007) proposed 3-D histogram in RGB space following SVD and dynamic clustering for SBD. Amiri et al. (Amiri and Fathy 2010) proposed a computationally efficient QR-decomposition and Gaussian transition-based approach for SBD using RGB color histogram feature to address large COM. The Gaussian model based shot transition representation and QR decomposition based filter successfully identifies shot transition. The accumulating histogram difference based approach of Ji et al. (2010) efficiently avoided the false detection due to flashlight. As a modification to the basic histogram based method, Lakshmi Priya and Domnic (2010) proposed a block wise histogram difference in RGB color space with automatic threshold for AST detection. Amiri and Fathy (2011) proposed an computationally efficient generalised eigen value decomposition (GED) using block wise 3-D RGB color histogram feature for SBD under large COM. Lee et al. (2011) proposed RGB histograms feature based SBD algorithm following KSVD for dimensionality reduction and Kernel-ART to improve shot detection probability by mapping to high dimensional feature space. Kernel-adaptive resonance theory(ART) clustering algorithm took the advantage of adaptive resonance theory and mercer kernel for SBD of anchor videos. Adnan and Ali (2013), proposed HSV and gray histogram based SBD through third-order polynomial curve fitting technique but

has sensitivity to dark frames, similar backgrounds and flashlight effects. Guru and Suhil (2013) proposed color histogram based approach for SBD using split and merge technique of segmentation approach. Li et al. (2016) proposed, a multi-stage approach based on the histograms difference using self-adaptive thresholds and voting technique based transition detection that successfully eliminated the noise caused due to COM, flash light and camera zoom. Yuan and Zhang (2023) proposed a two stage SBD approach to address both sudden illumination change and high speed COM. First phase focussed on color clustering changes through color histogram in small regions for AST detection followed by attention mechanism with sliding window for GST detection.

Since the histogram representation overlooks the spatial coherency between the pixels in the image, so even for extremely dissimilar content of the image frame, the histogram may have similar representation leading to missed detection in histogram based methods. Additionally, since basic histogram feature based methods fail for videos involving very high camera motion and camera operation (Hanjalic 2002) and large intensity changes and incapable to distinguish between dissolve transition and motion (Abdulhussain et al. 2018) they have been merged with other efficient techniques. Despite of the weaknesses of histogram based approaches, they are still preferable and popular owing to, acceptable trade-off between computational complexity and performance.

6.5 Information theory based techniques

In the literature information theory based approaches (Cernekova et al. 2006) has also gained popularity in SBD task. It includes frame feature extraction and formulation of inter frame information using Entropy and Entropy related parameters such as mutual information(MI), joint entropy(JE) and cross Entropy(CE). In this direction Cernekova et al. (2002, 2006) proposed MI and JE feature based SBD in uncompressed domain. A weak inter-frame dependency leads to a low MI indicating an AST and decreasing and increasing inter-frame JE value indicates a fade transitions. Bi et al. (2011) Proposed MI and JE based inter frame information and SVM based classification for AST detection. They reduced false detection occurred due to fade using JE. Inspired by Vasconcelos and Vasconcelos (2004); Vasconcelos (2003), Cooper et al. (2007), proposed MI based histogram feature selection in YUV color space through a KNN based supervised classifier with a single parameter, and has higher computational complexity. Baber et al. (2013) used entropy difference of the consecutive frames to select candidate shot following SURF (Bay et al. 2006) key point matching for SBD. But it shows poor performance under high COM and poor matching of key points with frame resizing.

6.6 Transform domain techniques

Different transforms such as DCT (Panchal and Merchant 2012) and Dual tree complex wavelet transforms (DTCWT) (Warhade et al. 2011; Mishra et al. 2013) have been successfully applied in the literature for SBD. Barbu (2009) proposed three dimensional Gabor feature based SBD through region growing based classification. It has higher computational complexity. Shekar et al. (2011), generated frame wise local feature transform (LFT) and then color texture moment based feature extraction for AST detection. Panchal and Merchant (2012) deployed DC coefficient and DC image in DCT domain. Subsequently the statistical parameters of DC images were used to obtain the threshold value for effective detection of fade and dissolve transition. Mishra et al. (2013) extracted frame wise

structural information using dual tree complex wavelet transform (DTCWT) (Selesnick et al. 2005) followed by structural similarity algorithm for similarity evaluation and AST detection. Lakshmi Priya and Domnic (2014) proposed multiple feature extraction based on Walsh Hadamard transform (WHT) kernel. The authors generated the similarity measure by combining color, edge, motion and texture strength followed by SVM based classifier for SBD and achieved good performance. Due to learning based algorithm it relied on large training data. Mondal et al. (2017) proposed a technique for simultaneous detection of AST and GST by taking the advantages of multi-scale, multi-directional and shift invariance properties of non-subsampled contourlet transform followed by dimension reduction through principal component analysis(PCA) and SVM classifier. PCA also ensured reduction in noisy features at lower computational complexity and better performance. But the method failed to handle different types of camera effects as well as flash light effects.

6.7 Soft computing and learning based classification Techniques

In the literature many machine learning approaches have been proposed (Chavez et al. 2006, 2007). In such approaches either some soft computing based feature extraction techniques are followed or some soft computing based classifiers are used for SBD. Lo and Wang (2001) proposed histogram based fuzzy c-means clustering for video segmentation which produced better performance than regular histogram based method. Qi et al. (2003) used k-nearest neighborhood classification technique for video segmentation. They combined global color histogram difference with block histogram difference between frames, in the YUV color space thus making the system insensitive to object motion. Chavez et al. (2007) introduced hybridization of SVM and K-nearest neighbor for transition detection. KNN-SVM classifier has comparatively smaller generalization error than global SVM. Xuemei et al. (2010) proposed SVM based classifier for detection of AST, GST and no transition frames. But it suffered from higher computational complexity. Kucuktunc et al. (2010) proposed fuzzy rule based fuzzy color histogram and dual threshold for AST and GST detection. Inspired by Kucuktunc, Dadashi and Kanan (2013) proposed a robust solution to AST detection in presence of flashlight effects, camera operation and COM through spatial and temporal dependency of video frames using a block wise fuzzy color histogram feature and fuzzy rule based classifier. False detection were observed mostly due to GST.

Rashmi and Nagendraswamy (2021) applied block based cumulative sum technique on fuzzified gradient frame for extracting mean cumulative sum histogram (MCSH) which is robust to noise and illumination variation. But it is threshold dependent and sensitive to complex camera and light variation affecting the overall performance of the algorithm. Moreover, it has limited efficiency during camera zooming and panning. It is also computationally expensive than global histogram techniques and false detections are encountered for frames of different shots having similar histograms. Mohanta et al. (2012) computed several local and global features for frame transition parameters for automatic SBD via artificial neural network (ANN) based classifier followed by post processing technique to reduce false transition and mis classification error. However, it failed for fast COM, shot lengths less than 30 frames and has higher computational budget due to block matching algorithm. Thounaojam et al. (2016) proposed fuzzy logic based SBD via GA based optimization. They computed membership functions through GA and shot classification by fuzzy system. However, the accuracy depends on the number of iterations of the GA optimization technique. Bhaumik et al. (2017) proposed a two phase dissolve detection technique. First phase, dealt with identification of the candidate segment through the parabolic

pattern recognition via fuzzy entropy. Second phase identifies the false detection by various parameters and threshold. Fan et al. (2017) proposed a two stage algorithm based on fuzzy color distribution chart that efficiently reduced the effect of noise, variation in illumination and insertion of words/logos and successfully distinguished the GST from fast object motion in the frames. In second stage false detection due to illumination variation and camera movement were reduced through the SIFT (Lowe 2004) feature matching algorithm. This method has a low GST performance and used several empirically selected thresholds at different stages. Sasithradevi and Roomi (2020) Proposed a low complex SBD technique for simultaneous detection of AST and GST in presence of sudden illumination changes, large motion based on pyramidal opponent color-shape model. To reduce the complexity of the SBD method the authors used suitable segment selection procedure. The selected segments were evaluated for transition detection using learning based bagged-trees classifier algorithm. Their algorithm was tested on a large dataset and shows appreciable F1-score of more than 95 %. Several convolutional neural network (CNN) based architectures (Xu et al. 2016; Bouyahi and Ayed 2020) proposed in literature for SBD. Xu et al. (2016) exploited CNN to extract segment for coarser detection of AST and GST following a pattern low level features and then high level features. They extracted candidate matching algorithm for detecting exact location of the boundary. Bhaumik et al. Hassanien et al. (2017) proposed 3D CNN based SBD through GPU by simultaneously optimizing accuracy and processing. It considered at a time a group of 16 frames as input and classified them as either AST, GST or no transition frames. The authors claimed to have good performance for AST detection and outperformed for existing GST detection techniques. Along with significant improvement in detection of wipe patterns and it offers nearly 11 times faster processing than existing ones for all the TRECVid datasets considered for the purpose. However, small datasets were not sufficient enough to train a CNN architecture (Baraldi et al. 2015). Lorenzo et al. (2017) used pretrained VGG16 model (Simonyan and Zisserman 2014) based deep neural network for scene classification by ex-ploiting multimodal features from images such as textual and audio features. Tong et al. (2015) proposed a structured network for initial detection of candidate transitions by light filtering module followed by separate detection of AST and GST where, the AST detection is based on frame similarity and the GST detection is based on temporal pattern of the frame sequence. By using TITAN GPU it achieved 30x speed. Gygli (2018) posed the SBD problem as a binary classification problem where the objective is to predict correctly, if a frame belongs to same shot as the previous frame by using CNN architecture and GPU allowing 120× faster processing for real time application. However, it fails for long dissolves, partial scene changes and fast changing scene with motion blur. Inspired by TransNet architecture of Gygli, Soucek et al. (2019) proposed multiple dilated 3D CNN architecture resulting in the larger field of view with lesser training parameters. It does not need any post processing, has nearly 100× faster processing and has better performance with the coexisting methods. Souček and Lokoč (2020) proposed improved Transnet arcitecture which offers state of the art performance. Wang et al. (2021) proposed a multistage deep CNN model based framework for SBD system with very good performance. In Benoughidene and Titouna (2022), a combination of transfer learning and long short term memory (LSTM) network has been applied for AST detection only. Here the authors extracted the spatial features by using two parallel pre-trained models followed by AST classification using a LSTM network. Sasithradevi and Roomi (2022) proposed SBD using relative entropy based pyramidal features and LSTM. The important cues such as opponent color space model and local phase quantization (LPQ) were extracted from uncompressed frames of the video in pyramidal fashion. In this framework, windowing technique is followed for deriving the dissimilarity

measure (relative entropy) to construct the temporal signature for the video. This temporal signature is used to portray the shot transitions between scenes and finds the extracts related to inter relationship between the frames. Entropy based measure and LSTM successfully classified the frames. To address the problem of SBD using a single training sample, an associative memory neural network model having online learning ability is proposed based on sport videos (Ke 2022). The authors claimed to achieve good performance for live broadcasted sports videos based on sporting events.

In threshold based approaches for SBD, a good performance can be expected only if the feature similarity values are well separated in time. However, the primary difficulty lies in setting up of appropriate thresholds for detecting different kinds of transitions for different genres of videos. In contrast, learning based classification strategies are free from the burden of threshold selection, selection of proper window length for adaptive thresholds and at the same time demands for larger training data and fine tuning of the hyper parameters.

6.8 Texture feature based techniques

Texture has been identified as an important characteristic to capture image feature and has been deployed for various tasks in image processing. It is a measure of coarseness, roughness, regularity, directionality, contrast and line likeness (Tamura et al. 1978). In this direction many authors have proposed texture (Guru et al. 2016) based method for detection of shot boundaries. Bhowmick and Chattopadhyay (2009) adapted a threshold independent, unsupervised classification technique of AST detection using texture feature based on co-occurrence matrices extracted from differences between consecutive frames. Finally AST frames were detected using clustering algorithm. Angadi and Naik (2012) extracted block color moment based texture feature vector in YCbCr color space which includes mean, standard deviation and skewness of the color space of the frame and compared to a threshold for AST detection. In contrast to the traditional image texture, Taile and Wenjun (2014) proposed dynamic texture based SBD. Video dynamic texture is identified by average gradient direction of the frame sub domain. They defined video texture which captures the dynamic global characteristics of video frames and the local characteristics of the image. LBP (Ojala et al. 2002; Heikkila et al. 2009) based approach proposed by Singh et al. (2019) for AST detection through automatic threshold selection by two stage feature selection and two stage classification technique offers good performance. They employed illumination, rotation and motion invariant block wise LBP-Histogram Fourier (LBP-HF) feature extraction followed by block wise Canny feature difference. The adaptive threshold generated from the two similarity measures first declares the most probable transitions and then confirms the AST location. As an improvement in LBP based method Chakraborty et al. (2022) proposed LTP based approach in Lab color space under high object camera motion for SBD with lower sensitivity to noise and illumination.

Only texture feature based approaches may not produce good performance for SBD, so in literature they are combined with other features to produce better results.

6.9 Motion feature based techniques

Although intensity, histogram, edge, Mutual information are simpler and faster techniques of SBD but has poor detection performance. To boost the detection performance, additionally, motion information is included by estimation of motion vector/ optical flow (Meng et al. 1995; Boreczky and Wilcox 1998; Lakshmi Priya and Domnic 2014; Yuan et al.

2004; Liu et al. 2007; Shu and Chau 2005; Kang et al. 2007; Ding and Yang 2008). Ngo et al. (2002), Porter et al. (2003), Wang et al. (2007) and Hua (2010) proposed a motion-based algorithm for SBD. Inclusion of motion information tends to make the system robust to motion, increases the accuracy at the cost of higher computational load (Baker et al. 2011; Dosovitskiy et al. 2015; Tao et al. 2012).

6.10 Sampled frame feature difference techniques

Instead of finding the consecutive frame's feature difference, many authors sampled the video in temporal domain to reduce complexity and compared the successively sampled frames to select frames of expected transition segment. In this process massive number of undesired frames (frames without transition) are filtered out thus reducing computational burden. Further refinement is applied only on the filtered segments that is expected to contain transitions.

In this line Li et al. (2009) proposed an frame sampling framework based on absolute intensity difference feature and adaptive threshold to select candidate transition for SBD. However, intensity feature makes it sensitive to flashlight, color changes and high motion. Again due to the buffering requirements of the frames, it is unsuitable for real-time applications. Lu and Shi (2013) proposed a fast SBD scheme to detect both AST and GST through candidate segment selection and applied SVD for dimension reduction making the detection process computationally efficient and faster using intensity based feature. AST detection and GST detection were employed on separate segments. Authors adapted an inverted triangular pattern matching for detecting both AST and GST with good performance. Due to higher detection speed, it is suitable for real-time application. As an improvement of the previous method, Dhiman et al. (2019) proposed a faster SBD technique for real time applications considering only the blue plane in the RGB color space through candidate segment selection followed by DCT based AST detection and finally GST detection through histogram feature based pattern matching. The authors claimed to have appreciable accuracy and low time complexity. However, the algorithm is tested only for very few frames with respect to the comparable methods and involves several empirically selected thresholds. SURF (Bay et al. 2006, 2008) key point feature based algorithm was proposed by Birinci and Kiranyaz (2014) for SBD through number of matched key points between sampled frames to find expected transition segment which is further processed for exact location of the transition, making the process faster. False detection and miss detection occurred due to frame blurring and low-intensity frames respectively. In an another approach, a novel SBD algorithm is proposed to maximize detection accuracy by analysing the transition behavior of a video in two stages (Raja et al. 2022). In first stage the video was segmented into primary segments and candidate segments by using the colour feature and the local adaptive threshold. Then AST and GST detection were achieved by fine tuning the candidate segment through SURF.

Jiang et al. (2013) proposed a dual detection model based on color histogram difference and intensity differences between the uneven blocks of sampled frames in a window, in first phase. The weighted combination of these two differences were well capable to reduce the effect of large COM. In second phase (SIFT)(Lowe 2004) is applied to exclude false detection. The re-detection round adaptively modifies the relative thresholds to achieve better performance. Gao and Ma (2014) proposed a faster technique of SBD in spatial domain, by processing only the pixels in focus regions (FR) using MI and color histogram features. Here the authors exploited adaptive frame sampling for faster processing. Moreover, by

adapting corner distribution of frames near candidate shot boundaries, the authors were able to find the accurate interval of the boundary while removing the false boundaries. Cao and Cai (2007) proposed a SVM based classifier to find the wipe patterns and digital video effects. They obtained a feature vector based on number of intra-coded and macro blocks with forward, backward and bidirectional prediction, number of blocks changed between DC coefficients in successive frames and the type of frame following SVM based classifier. Tan et al. (2007) presented block color histogram based similarity measure of sampled frames, followed by SVM based classifier for AST and GST detection through pattern matching. Chasanis et al. (2009) proposed RGB feature histogram based SBD following SVM based classifier. In another method Chasanis et al. (2009) applied shot grouping technique through spectral clustering by global k-means algorithm. The shot similarity metric is evaluated based on visual features for SBD. This algorithm preserves a good trade off between recall and precision measure. However, it suffers from over segmentation for videos with continuously changing visual content in a scene. Deepak et al. (2013) proposed color correlogram and Gauge Speeded-up robust features (Alcantarilla et al. 2011) for AST detection. Zhang et al. (2019) proposed an AST detection technique based on block-wise principal component analysis (PCA) in reduced dimension and GST detection based on bidirectional pattern matching. Initially they selected the candidate segments.However, candidate segment selection and GST detection required multiple empirically selected thresholds. Duan and Meng (2020) proposed a combined feature approach using DCT and SURF feature applied to selected frames obtained by non-overlapping interval method followed by coarse and then fine detection of shot boundaries. They claimed to achieve good performance for GST and long shots at lower computational complexity. Liu et al. (2020) proposed a combined feature based automatic SBD approach deploying fast feature descriptor of Oriented FAST and Rotated BRIEF (ORB) and SSIM. The ORB descriptor is responsible for candidate segment selection and AST detection. GST detection is achieved by finding maximum amount of the monotonous increase or decrease of feature differences in the candidate segment in the absence of AST. In another candidate selection based SBD framework Idan et al. (2021) obtained the informative content area of the frames following moment computation using orthogonal polynomials. They finally used SVM for AST detection. The authors claimed to have good performance with lower computational burden. Gushchin et al. (2021) used a combination of color histograms and object boundary for SBD and claimed to achieve superior performance.

6.11 Combined feature based tecniques

In order to take the benefit of different image features to improve the performance of the SBD technique, it is a good choice to combined multiple features (Sumiyoshi et al. 2007; Ling et al. 2008).The different features are combined in different proportions depending on their contribution. Jadon et al. (2001) proposed histogram feature for AST detection, histogram and intensity based feature for GST detection and a combination of intensity, histogram and edge-pixel count for fade detection. Huang and Liao (2001) proposed a combined feature of DC image intensity and histogram difference for SBD. They proposed an intensity statistics model using correlation feature for GST detection. Fang and Jiang (2006) proposed a bimodal hybrid scheme for SBD by integrating histogram intersection, motion information and change in texture energy via fuzzy membership value for AST detection and checked for increase/decrease of variance of edge over few continuous frames for fade in/fade out detection respectively. Moreover, when a fade-out followed a

fade-in transition it was identified as a dissolve transition. But evaluation of motion compensation feature increased computational burden. Han et al. (2007) proposed mid level features such as color and the motion vector in combination with region color histograms for sports video segmentation in soccer video. The mid level feature extraction following SVM classifier achieved F1 score of more than 95 %. Grana and Cucchiara (2007) developed an iterative SBD algorithm by combining intensity and histogram difference following adaptive threshold. It could not perform well for high speed COM (Guimaraes et al. 2009; Chen et al. 2011) and also has higher computational burden (Li et al. 2009). [209] presented a SBD algorithm by combining multiple features such as pixel, histogram and motion features to increase the robustness against camera operation. In the absence of AST, authors executed GST detection by comparing modified distance measure between nonconsecutive frames to specified thresholds. The method performed well in presence of motion and flash light. However it suffered from selecting number of threshold values at different stages. Fast dissolve transitions generated false negatives while blurred frames generated false positives. Moreover, frame sequences with fast flash light changes and COM may be mistaken as gradual transitions and dissolve with very short duration may be mistaken as AST. Chen and Hsu (2017) proposed an algorithm which is well capable to differentiate between luminance variation due to flash light and due to AST by assuming the shot duration existence of the flash light and no change in the video content before and after the occurrence of the flash light. Towards this goal they identified the position of the source of light, selected illumination insensitive features, identified the temporal relations about the expected transition, considered the color space before feature extraction and differentiated the flash lights and fade transitions by a bright frame of high intensity. This method failed when the flash light affected the last frame of the shot. Youssef et al. in 2017 Youssef et al. (2017) implemented adaptive feature extraction using the Frobenius norm of low rank approximation matrices under high speed object camera motion. Each frame's block based feature histogram is mapped into k -dimensional vector. The classification of the similarity measure is achieved via unique double thresholding technique for detection of the AST. GST detection through SVD-updating technique avoided the recalculation of the whole SVD decomposition for segment correction reducing the computational burden. Wu et al. (2019) proposed a two stage feature fusion based SBD approach. AST detection was achieved by fusing color histogram and deep features in the first stage. Then between the detected ASTs the GST detection was achieved by 3D-convolutional neural network. Zhou et al. (2021) proposed a simple and collaborative feature based approach which includes image color feature, local descriptors and motion information for SBD. Initially they selected candidate transition segments via color histogram and SURF. Then, AST detection was achieved through uneven slice matching, pixel difference, and color histogram. Finally, GST detection was achieved by the motion area extraction, scale-invariant feature transform (SIFT) and even slice matching. The authors claimed to achieve good performance with lower computation time. Another multiple feature based approach (Jose et al. 2022) proposed by Jasmine et al. combined multiple invariant features such as edge change ratio (ECR), colour layout descriptor (CLD), and scale-invariant feature transform (SIFT) key point descriptors followed by SVM classifier for SBD. Due to selection of illumination and motion insensitive features many false detection were reduced. However videos involving large videos tend to produce false detection. Kar and Kanungo (2023) proposed a two stage SBD approach in presence of motion and illumination variation.In the first phase AST detection was achieved by joint histogram of gradient magnitude and gradient orientation feature frame. Then the GST detection was applied only on the frames within two AST frames. Sasithradevi and Nirmala (2023) proposed a combined feature

based approach. The model utilized different statistical parameters to define the influencing factors of a visual significant model such as color, texture, shape and focus. This generated a visual significance model for constructing temporal signature. The final classification is achieved by random vector functional link (RVFL) networks. The authors claimed promising results for detecting the video transitions even in the presence of varying illumination conditions, fast COM.

6.12 Other relevant works

Some SBD techniques which could not be placed under either of the above mentioned categories were grouped as other relevant works. Motivated by the concept of constant false alarm rate (CFAR) in radar signal detection process Liu et al. (2004) proposed an AST detection technique for controlling false alarm rate in the detection process. The primary contribution was that instead of considering the feature difference of consecutive frames, they identified the ASTs by evaluating the number of previous frame difference lesser than the current one. Boccignone et al. (2005) proposed a video partitioning algorithm based on foveated representations of the video. It detected complete shot boundaries using a single technique, rather than a set of dedicated methods. Bescos et al. (2005) mapped the inter-frame distance values onto an multidimensional space that includes frame ordering information following an inverted triangular pattern matching technique for GST detection. Nam and Tewfik (2005) used a B-spline interpolation, curve fitting technique for the estimation of editing effects. It detected different transition behavior even under motion and other post-processing operations but under performed for transitions of very small duration, shots with substantial camera motion, dissolve transition with high degrees of correlation between consecutive frames, dissolve transition with slow zooming between same objects or background locations. Moreover, false fade detection were also encountered for very dark and homogeneous scenes having less textural feature such as scene captured under water. Rasheed and Shah (2005) converted the SBD problem into a graph partitioning problem by constructing shot similarity graph, where the node characterized a shot and edges between the shots are weighted by their similarity considering color and motion information and claimed to achieve good performance.

Joyce and Liu (2006) proposed two compressed domain algorithms for GST detection which needed partial decoding of the compressed video stream. The first algorithm evaluated dissolve trajectory in image space for dissolve transition detection. The second one relied on characteristics of image histogram during such transitions for wipe detection. Lankinen and Kämäräinen (2013) proposed SBD using object detection approach of bag of words in association with SIFT feature and k-mean clustering technique. They achieved good performance than baseline methods. Fu et al. (2013) proposed an adaptive SBD algorithm using reducing difference algorithm in order to reduce the vulnerability to motion, in HSV space following different data processing techniques for AST and GST detection. Huo et al. (2016) used histogram of inter frame difference for threshold selection towards SBD task based on poisson model. Yan et al. (2022) obtained the shot boundaries in two phases. In the first stage, they used CIEDE2000 color-difference and adaptive threshold to find the possible AST frames. Then applied BRISK feature to detect actual AST frames. In the second stage the change in the brightness of the video frames is used to detect the possible GST frame segment. Then CIEDE2000 color-difference along with cumulative frame algorithm is used to detect actual GST frames. Ramli et al. (2019) proposed Krawtchouk-Tchebichef Orthogonal polynomial moment based feature extraction followed by SVM

classifier for AST detection achieved good performance for large variety of datasets. Rao et al. (2020) proposed a video segmentation algorithm through multi-modal information across three levels, i.e. clip, segment and movie. Chakraborty et al. (2021) used Lab colour difference and mean luminance pattern for detecting AST and GST in a video. Existing approaches address the SBD problem based on the visual differences and content transitions between consecutive frames, while ignoring intrinsic shot attributes such as camera movements, scales, and viewing angles, which essentially reveal the way a shot is generated. Keeping this in view Jiang et al. (2022) proposed a new learning framework (SCTSNet) for SBD by identifying the attributes and composition of shots in videos. All SBD algorithms put in a common framework is illustrated in Fig. 7.

6.13 Key observations and research gap

The exhaustive literature review help to identify the research gap and the key observations in the field of SBD. The gap can be summerised as follows.

- Identification of any new suitable hand crafted feature space, insensitive to COM and sudden illumination variation having lower complexity. Care must be taken to address dark frames, flash light affecting single and multiple frames or blast scene.
- Finding methods to combine multiple features. Identifying other modalities and combining multimodal features for SBD. Different features and different methods can be combined to produce impressive performance. Equal weightage to all feature value cannot give optimum results. Learning based strategy can be followed to find the weight of different features. Multiple feature tend to improve the performance at the cost of higher complexity.
- Incase of large feature space, dimension reduction approaches can be followed for faster processing.
- Identify suitable similarity measure for finding frame similarity/dissimilarity. Since all similarity measure won't work for all feature space to produce impressive outcomes.
- Finding simple methods to identify potential candidate segments of a video for feature extraction, instead of applying the feature extraction to the whole video which can substantially reduce the computational burden and the execution time.
- Finding suitable and automatic threshold for AST and GST detection. Sometimes single threshold may not work. So multiple thresholds need to be found out.
- Finding some suitable method for post processing to minimise false detection.
- Designing suitable deep learning model using CNN for deep feature based SBD. CNN model needs larger training dataset.
- Combining deep features and different hand crafted features for SBD.
- Time complexity in different approaches refer to execution time. Complex feature extraction will require higher execution time. Moreover, multiple featurs as compared to single feature tend to produce better result but increases computational load. Instead of applying the feature extraction on the complete video, if first potential candidate segments can be obtained and then apply the feature extraction on the candidate segment then the execution time can be reduced substantially. Generally spatial domain approaches are simpler as compared to transform domain approach leading to lower execution time. Learning based strategy also require large dataset and higher execution time.

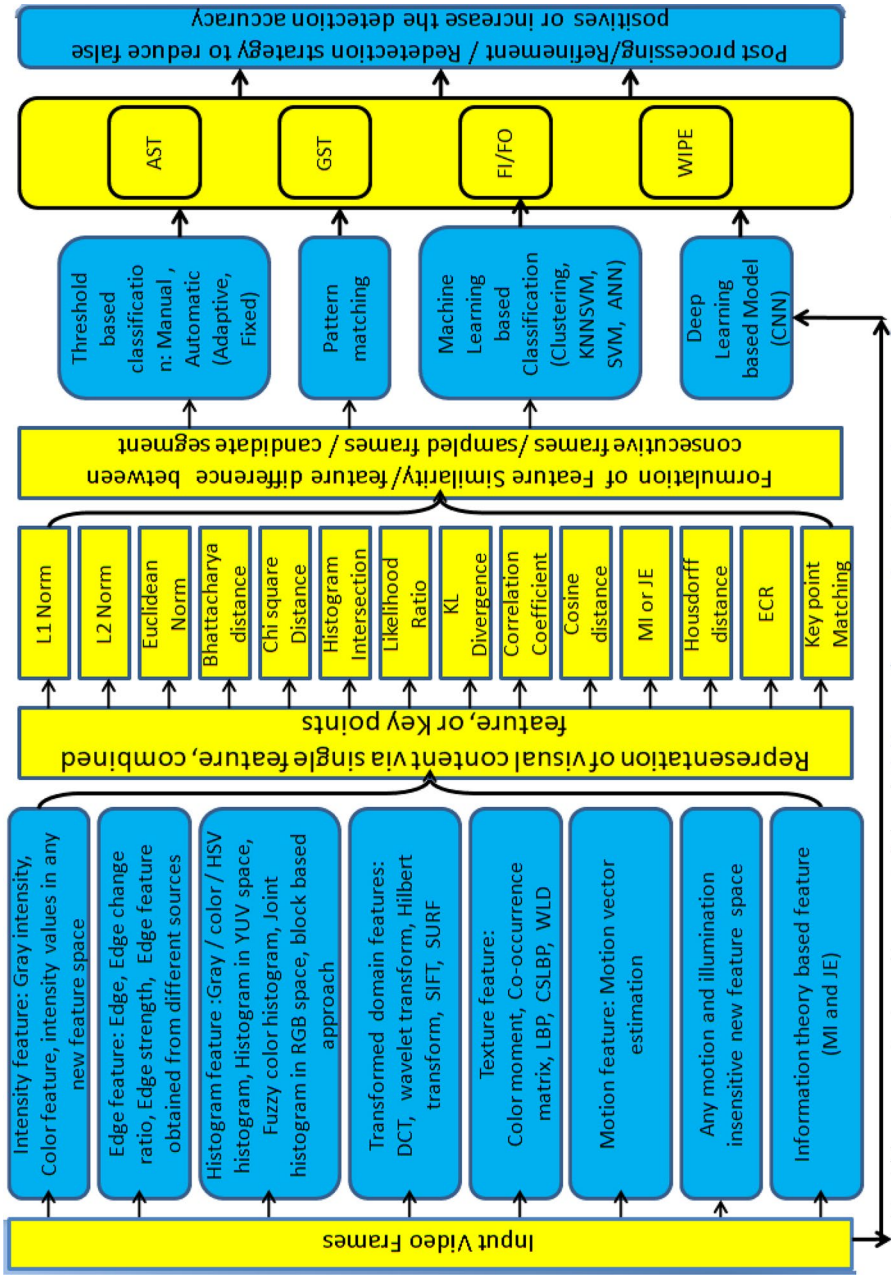

Fig. 7 A complete SBD approach

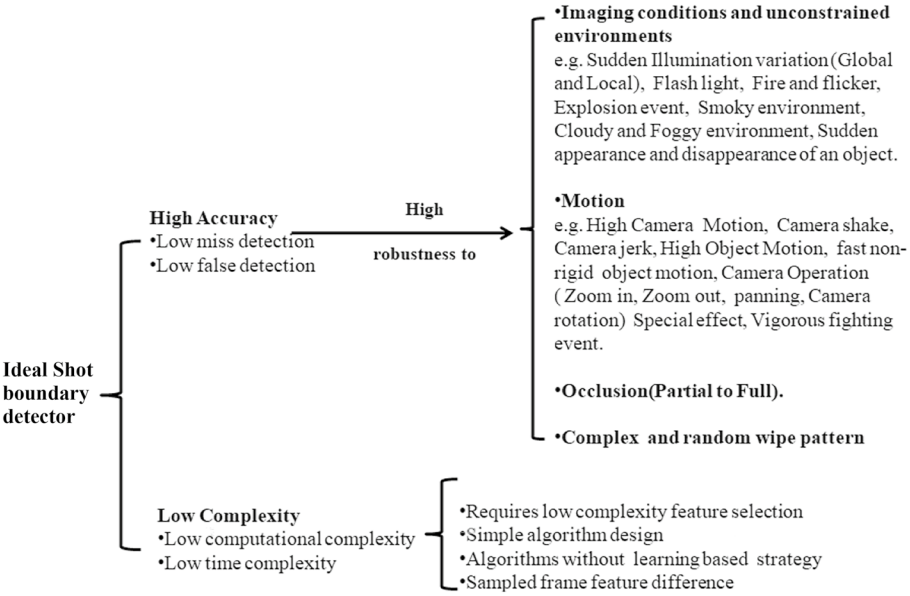

Fig. 8 Taxonomy of challenges in SBD process

7 Challenges in the field of SBD

The review of the literature paves the way to identify major contributions, important findings in the field and the primary challenges in the field of SBD. It is observed that in most of the algorithms, AST has been successfully addressed with higher accuracy than GST (Chasanis et al. 2009a). Spontaneous lighting change (flashlight) (Chen and Hsu 2017), large motion associated with COM (Warhade et al. 2011, 2013) and complex nature of the editing are the primary causes of failure in efficient detection of abrupt transitions. Flashlight effects are deliberately introduced in movies to emphasize a crisis or excitement, sense of apprehension or an indicative of some supernatural power, in order to improves the emotional experience of the viewer. Presence of sudden change in illumination or high motion tend to change the feature space causing a noticeable change in similarity measure. This may lead to either increased miss detection or false detection affecting the accuracy of any SBD process and hence performances of the SBD algorithm. To guarantee maximum efficiency of any SBD process the algorithm should be able to:

- Minimize false detection, for intra-shot frames.
- Minimize miss detection, between inter-shot frames.
- Differentiate between an actual scene break from that of the variation due to camera operation (paning/zooming/tiling etc), COM, illumination variation and special effects.
- Identify different forms of the GST such as dissolve, fade in and fade out effects even under complex nature of the editing effect.

The complete taxonomy of challenges in SBD process is pictorially illustrated in Fig. 8.

Sample frames from the English action movie "Transformer" illustrating illumination variation due to fire and blast event affecting a single frame is shown in Fig. 9a and b

Fig. 9 **a** Illustration of illumination variation due to fire and object motion in the video Transformer, **b** Illustration of illumination variation of a single frame due to blast event in the video Transformer,(c)Illustration of illumination variation due to flash light in the video Littlemiss Sunshine

Fig. 10 Illustration of substantial COM in video Transformer

respectively. Sample frames from the English movie "Little miss sunshine", illustrating flash light effect is shown in Fig. 9c. Sample frames from the English movie "Transformer" illustrating high COM is shown in Fig. 10.

8 Conclusions and Future perspectives

Video is the most significant mode of communication to convey information. It is estimated to be nearly eighty percent of our total perceived information and renders the users of multimedia systems a treasure of information due to its unique and rich information representation capabilities and its real-world mapping attributes. The dramatic growth of video material in recent times lead to the issue of accessing the desired content manually demands higher processing time. Temporal video segmentation is the elementary step towards this goal. In this paper we did an exhaustive review of the existing approaches and classified them for temporal video segmentation. This review is expected to serve a benchmark reference for research communities and will truly put light on relative merits and demerits of different SBD algorithms from different aspects. It will inspire multimedia research communities to make a tangible impact on its growth. The limitations of the existing methods unlocked new directions for research in the domain of CBIR. They can be summarized as follows

- Direction towards soft computing based approach with new features to address GST detection. Additionally, some more pattern matching techniques can be explored with new patterns for the identification of special edit effects.
- Direction for performance improvement of the SBD algorithm, beyond the visual clues through, integration of evidence from other modalities such as audio or text information.
- Identification of different motion and illumination insensitive features and their modeling towards recent SBD.
- Identification of suitable local and global feature descriptors for SBD task.
- Finding new deep learning based architecture for SBD.
- Subsequent processing of SBD outcomes to different applications discussed in Sect. 2.

So identifying new feature space, developing new similarity measure, finding fully automatic and adaptive threshold to fit different genres of videos with lesser computation time, combining video processing and computer vision with soft computing and learning based approaches to handle videos under challenging environment can be the directions of future research.

Author Contributions All Authors have equal contribution.

Declarations

Conflict of interest The authors declare no competing interests.

References

aaaa,bbbb
Abdulhussain SH, Ramli MI, Saripan AR, Mahmmod BM, Al-Haddad SAR, Jassim WA (2018) Methods and challenges in shot boundary detection: a review. Entropy 20(4):214
Abdulrahaman MD, Faruk N, Oloyede AA, Surajudeen-Bakinde NT, Olawoyin LA, Mejabi OV, Imam-Fulani YO, Fahm AO, Azeez AL (2020) Multimedia tools in the teaching and learning processes: a systematic review. Heliyon 6(11):e05312
Adjeroh D, Lee MC, Banda N, Kandaswamy U (2009) Adaptive edge-oriented shot boundary detection. EURASIP Journal on Image and Video Processing, (859371)
Adnan A, Ali M (2013) Shot boundary detection using sorted color histogram polynomial curve. Life Sci J 10(4):1965–1972
Alcantarilla PF, Bergasa LM, Davison AJ (2011) Gauge-surf descriptors. Image and Vision Comput 31(1):103–116
Amiri A, Fathy M (2010) Video shot boundary detection using qr-decomposition and gaussian transition detection. EURASIP J Adv Signal Process 2009:1–12
Amiri A, Fathy M (2011) Video shot boundary detection using generalized eigenvalue decomposition and gaussian transition detection. Comput Inform 30:595–619
Angadi SA, Naik V (2012) A shot boundary detection technique based on local color moments in ycbcr color space. Comput Sci Inform Technol 2(3):57–65
Baber J, Afzulpurkar N, Satosh S (2013) A framework for video segmentation using global and local features. Int J Pattern Recognit Artif Intell 27(05):591–594
Bajaj D, Sharma S (2016) Comparative analysis of shot boundary detection algorithms for video summarization. CSI Trans ICT 4:265–269. https://doi.org/10.1007/s40012-016-0093-0
Baker S, Scharstein D, Lewis J, Roth S, Black MJ, Szeliski R (2011) A database and evaluation methodology for optical flow. Int J Comput Vision (IJCV) 92(1):1–31
Baraldi L, Grana C, Cucchiara R (2015) A deep siamese network for scene detection in broadcast videos. ACM Multimed pages 1199–1120. https://doi.org/10.1145/2733373.2806316
Barbu T (2009) Novel automatic video cut detectin using gabor filtering. Comput Electr Eng 35(5):712–721
Bay H, Tuytelaars T, Gool LV (2006) Surf: Speeded up robust features. Computer Vision-ECCV 2006, Springer Berlin Heidelberg, pages 404–417
Bay H, Ess A, Tuytelaars T, Gool LV (2008) Surf: Speeded up robust features. Comput Vision Image Underst (CVIU) 110(3):346–359
Bendraou Y (2017) Video shot boundary detection and key-frame extraction using mathematical models. PhD Thesis
Benoughidene A, Titouna F (2022) A novel method for video shot boundary detection using cnn-lstm approach. Int J Multimed Info Retr 11(4):653–667
Bescos J, Cisneros G, Martinez JM, Menendez JM, Cabrera J (2005) A unified model for techniques on video-shot transition detection. IEEE Trans Multimed 7(2):293–307
Bhaumik H, Chakraborty M, Bhattacharyya S, Chakraborty S (2017) Detection of gradual transition in videos: Approaches and applications. Intelligent Analysis of Multimedia Information; IGI Global: Hershey, PA, USA, pages 282–318
Bhoraniya DM, Ratanpara TV (2017) A survey on video genre classification techniques. International Conference on Intelligent Computing and Control (I2C2),https://doi.org/10.1109/I2C2.2017.8321886,
Bhowmick B, Chattopadhyay D (2009) Shot boundary detection using texture feature based on co-occurrence matrices. IMPACT-2009,
Bi J, Liu X, Lang B (2011) A novel shot boundary detection based on information theory using svm. 4th International Congress on Image and Signal Processing, pages 512–516,
Birinci M, Kiranyaz S (2014) A perceptual scheme for fully automatic video shot boundary detection. Signal Process: Image Commun 29(3):410–423
Boccignone G, Chianese A, Moscato V, Picariello A (2005) Foveated shot detection for video segmentation. IEEE Trans Circ Syst Video Technol 15(3):365–377
Bommisetty RM, Khare A, Siddiqui TJ, Palanisamy P (2021) Fusion of gradient and feature similarity for keyframe extraction. Multimed Tools Appl 80:15429–15467
Boreczky JS, Wilcox LD (1998) A hidden markov model framework for video segmentation using audio and image features. IEEE International Conference on Acoustics, Speech and Signal Processing,
Bouyahi M, Ayed YB (2020) Video scenes segmentation based on multimodal genre prediction. Procedia Comput Sci 176:10–21
Brezeale D, Cook DJ (2008) Automatic video classification: a survey of the literature. IEEE Trans Syst, Man and Cybern Part C 38(3):416–430

Canny J (1986) A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell 8:679–714
Cao J, Cai A (2007) A robust shot transition detection method based on support vector machine in compressed domain. Pattern Recognit Lett 28:1534–1540
Cernekova Z, Nikou C, Pitas I (2002) Information theory based shot cut/fade detection and video summerization. Proceedings of international Conference on Image Processing, 16
Cernekova Z, Pitas I, Nikou C (2006) Information theory based shot cut/fade detection and video summerization. IEEE Trans Circuits Syst Video Technol 16(1):82–91
Cernekova Z, Kotropoulos C, Pitas I (2007) Video shot-boundary detection using singular-value decomposition and statistical tests. J Electron Imaging 16(4):51–59
Chakraborty B, Bhattacharyya S, Chakraborty S (2018) Generative model based video shot boundary detection for automated surveillance. Int J Ambient Comput Intell. https://doi.org/10.4018/IJACI.201810010 5
Chakraborty S, Thounaojam DM, Sinha N (2021) A shot boundary detection technique based on visual colour information. Multimed Tools Appl 80:4007–4022
Chakraborty S, Thounaojam DM, Singh A (2022) A novel bifold-stage shot boundary detection algorithm: invariant to motion and illumination. Visual Comput: Int J Comput Graph 38(2):445–456
Chan C, Wong A (2011) Shot boundary detection using genetic algorithm optimization. In: Proceedings of the 2011 IEEE International Symposium on Multimedia (ISM), Dana Point, CA, USA, pages 327–332, 5–7
Chasanis VT, Likas NP, Galatsanos AC (2009) Scene detection in videos using shot clustering and sequence alignment. IEEE Trans on Multimed 11(1):89–100
Chasanis V, Likas A, Galatsanos N (2009) Simultaneous detection of abrupt cuts and dissolves in videos using support vector mechines. Pattern Recognit Lett 30(1):55–65
Chavate S, Mishra R, Yadav P (2021) A comparative analysis of video shot boundary detection using different approaches. In 2021 10th International Conference on System Modeling and Advancement in Research Trends (SMART)
Chavez GC, Cord M, Precioso M (2006) Video segmentation via temporal pattern classification. 19th Brazilian Symposium on Computer Graphics and Image, pages 365–372,
Chavez GC, Precioso F, Cord M (2007) Shot boundary detection by a hierarchical supervised approach. Proceedings of the 14th. International Conference on Systems, Signals and Image Processing, pages 209–212,
Chen LH, Hsu BC (2017) A supervised learning approach to flashlight detection. Cybern Syst 48(1–12):28
Chen J, Ren J, Jiang J (2011) Modelling of content-aware indicators for effective determination of shot boundaries in compressed mpeg videos. Multimed Tools Appl 54:219–239
Cooper M, Liu T, Rieffel E (2007) Video segmentation via temporal pattern classification. IEEE Trans Multimed 9(3):610–618
Cotsaces C, Nikolaidis N, Pitas I (2006) Video shot detection and condensed representation a review. IEEE Signal Process Magaz 23(2):28–37
Cunhaa M, Mendesb R, Vilelaa JP (2021) A survey of privacy-preserving mechanisms for heterogeneous data types. Comput Sci Rev 41:100403
Dadashi R, Kanan HR (2013) Avcd-fra: a novel solution to automatic video cut detection using fuzzy-rule-based approach. Comput Vision Image Underst 117(7):807–817
Deepak CR, Babu RU, Kumar KB, Krishnan CMR (2013) Shot boundary detection using color correlogram and gauge-surf descriptors. Computing, Communications and Networking Technologies (ICCCNT),2013 Fourth International Conference, pages 1–5,
Depalov D, Pappas T, Li D, Gandhi B (2006) Perceptually based techniques for semantic image classification and retrieval. In: Rogowitz Bernice E, Pappas Thrasyvoulos N, Daly Scott J (eds) Human Vision and Electronic Imaging XI, vol 6057. International Society for Optics and Photonics SPIE, Bellingham, pp 354–363
Dhiman S, Chawla R, Gupta S (2019) A novel video shot boundary detection framework employing dct and pattern matching. Multimed Tools Appl 78(24):34707–34723
Ding JR, Yang JF (2008) Adaptive group-of-pictures and scene change detection methods based on existing h.264 advanced video coding information. IET Image Process 2(2):85–94
Dosovitskiy A, Fischery A, Ilg E, Husser P, Hazirbas C, Golkov V, Smagt P, Cremers D, Brox T (2015) Simpleflow: a noniterative, sublinear optical flow algorithm. ICCV, pp 2758–2766
Duan FF, Meng F (2020) Video shot boundary detection based on feature fusion and clustering technique. IEEE, ACCESS 8:214633–214645
Dutta D, Saha SK, Chanda B (2016) A shot detection technique using linear regression of shot transition pattern. Multimed Tools Appl 75(1):93–113

Ejaz N, Mehmood I, Baik SW (2014) Feature aggregation based visual attention model for video summarization. Comput Electr Eng 40(3):993–1005
Fabro DM, Böszörmenyi L (2013) State-of-the-art and future challenges in video scene detection: a survey. Multimedia Syst 19(5):427–454
Fan J, Zhou S, Jiang X, Siddiqui AM (2017) Fuzzy color distribution chart -based shot boundary detection. Multimed Tools Appl 76(7):10169–10190
Fu Q, Zhang Y, Xu L, Li H (2013) A shot boundary detection technique based on local color moments in ycbcr color space,. In proceedings of 9th IEEE International Conference on Computational Intelligence and Security,, pages 219–223
Gao G, Ma H (2014) To accelerate shot boundary detection by reducing detection region and scope. Multimed Tools Appl 71(3):1749–1770
Gargi U, Kasturi R, Strayer S (2000) Performance characterisation of video shot change detection methodes. IEEE Trans Circuits Syst Video Technol 10(1):1–13
Gianluigi C, Raimondo S (2006) An innovative algorithm for key frame extraction in video summarization. J. Real-Time Image Process. 1(1):69–88
Gong YH, Liu X (2000) Video shot segmentation and classifica tion. Proc. 15th Int. Conf. Pattern Recognit. 1:860–863
Grana C, Cucchiara R (2007) Linear transition detection as a unified shot detection approach. IEEE Trancs Circuits Syst Video Technolgy 17(4):483–489
Guimaraes Silvio JF, do Patrocinio Zenilton KG, Souza Kleber JF, de Paula Hugo B (2009) Gradual transition detection based on bipartite graph matching approach. In 2009 IEEE International Workshop on Multimedia Signal Processing, pages 1–6
Guru DS, Suhil M (2013) Histogram based split and merge framework for shot boundary detection. In: Prasath R, Kathirvalavakumar T (eds) Mining Intelligence and Knowledge Exploration. Lecture Notes in Computer Science, vol 8284. Springer, Cham
Guru DS, Suhil M, Lolika P (2016) A novel approach for shot boundary detection in videos. arXiv:1608.06716,
Gushchin A, Antsiferova A, Vatolin D (2021) Shot boundary detection method based on a new extensive dataset and mixed features. arXiv
Gygli M (2018) Ridiculously fast shot boundary detection with fullyconvolutional neural networks. International Conference on Content-Based Multimedia Indexing, CBMI 2018, La Rochelle, France, pp 1–4, https://doi.org/10.1109/CBMI.2018.8516556
Hameed IM, Abdulhussain SH, Mahmmod BM (2021) Content-based image retrieval: a review of recent trends. Cogent Eng 8(1):1927469
Han B, Hu Y, Wang G, Wu W, Yoshigahara T (2007) Enhanced sports video shot boundary detection based on middle level features and a unified model. IEEE Trans Consumer Electron 53(3):1168–1176
Hanjalic A (2002) Shot boundary detection: unraveled and resolved. IEEE Trans. Circuits Syst video Technol 12(2):90–105
Hanjalic A (2004) Content-based analysis of digital video. Kluwer, Academic Publishers, Boston
Hannane R, Elboushaki A, Karim A, Nagabhushan P (2016) An efficient method for video shot boundary detection and keyframe extraction using sift-point distribution histogram. Int J Multimed Info Retr 5:89–104
Hassanien A, Elgharib M, Bae SH, Hefeeda M, Matusik W (2017) Large-scale, fast and accurate shot boundary detection through spatio-temporal convolutional neural networks. arXiv preprint arXiv:1705.03281
Heikkila M, Pietikainen M, Schmid C (2009) Description of interest regions with local binary patterns. Pattern Recognit 42(3):425–436
Helm D, Kampel M (2019) Shot boundary detection for automatic video analysis of historical films,. International Conference on Image Analysis and Processing ICIAP, New Trends in Image Analysis and Processing - ICIAP 2019,Lecture Notes in Computer Science, vol 11808. Springer, Cham, 11808:137–147
https://pi4.informatik.uni mannheim.de/pi4.data/content/projects/moca/
http://trecvid.nist.gov/
http://www.scuola.rai.it
https://github.com/tangshitao/clipshots
https://trecvid.nist.gov/
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

Hua CL (2010) A hierarchical shot detection method for mpeg video. 2010 International Conference on Computer Application and System Modeling (ICCASM 2010),
Huang CL, Liao BY (2001) A robust scene-change detection method for video segmentation. IEEE Trans Circ Syst Video Technol 2419(12):1281–1288
Hui Fang Yue Feng, Jianmin Jiang (2006) A fuzzy logic approach for detection of video shot boundaries. J Pattern Recognit Soc 39:2092–2100
Huo Yi, Wang Yanfeng, Hu Haihe (2016) Effective algorithms for video shot and scene boundaries detection. In 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS), pages 1–6,
Idan ZN, Abdulhussain SH, Mahmmod BM, Al-Utaibi KA, Al-Hadad SAR, Sait SM (2021) Fast shot boundary detection based on separable moments and support vector machine. IEEE Access 9:106412–106427
Internet archieve. [online] available at: http://archive.org/details/movies
Iyer RR, Parekh S, Mohandoss V, Ramsurat A, Raj B, Singh R (2016) Content-based video indexing and retrieval using corr-lda. arxiv:1602.08581
Jacobs A, Miene A, Ioannidis GT, Herzog O (2004) Automatic shot boundary detection combin ing color, edge, and motion features of adjacent frames. TRECVID 2004:197–206
Jadon RS, Chaudhury SK, Biswas KK (2001) A fuzzy theoretic approach for video segmentation using syntactic features. Pattern Recogn Lett 22(13):1359–1369
Janwe NJ, Bhoyar K (2013) Video shot boundary detection based on jnd color histogram. 2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013), Shimla, India, 2013, pp. 476–480. https://doi.org/10.1109/ICIIP.2013.6707637.
Ji QG, Feng JW, Zhao J, Lu ZM (2010) Effective dissolve detection based on accumulating histogram difference and the support point. In Proceedings of the 2010 First International Conference on Pervasive Computing, Signal Processing and Applications (PCSPA), Harbin, China, pages 273–276, 17-19
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
Jose JT, Rajkumar S, Ghalib MR, Achyut Shankar A, Sharma P, Khosravi MR (2022) Efficient shot boundary detection with multiple visual representations. Mob Inform Syst. https://doi.org/10.1155/2022/4195905
Joyce RA, Liu B (2006) Temporal segmentation of video using frame and histogram space. IEEE Trans Multimed 8(1):130–140
Kang SJ, Cho KR, Kim YH (2007) Motion compensated frame rate up-conversion using extended bilateral motion estimation. IEEE Trans Consum Electron 53(4):1759–1767
Kar T, Kanungo P (2018) Motion and illumination defiant cut detection based on weber features. IET Image Process 12(10):1903–1912. https://doi.org/10.1049/iet-ipr.2017.1237
Kar T, Kanungo P (2023) A gradient based dual detection model for shot boundary detection. Multimed Tools Appl 82:8489–8506
Karthick S, Abirami S, Murugappan S, Sivarathinabala M, Baskaran R (2015) Automatic genre classification from videos. Artificial Intelligence and Evolutionary Algorithms in Engineering Systems. Adv Intell Syst Comput 325:389–401
Ke W (2022) Detection of shot transition in sports video based on associative memory neural network. Wireless Commun Mobile Comput. https://doi.org/10.1155/2022/7862343
Kim SH, Park RH (2002) An efficient algorithm for video sequence matching using the modified hausdorff distance and the directed divergence. IEEE Trans Crcuits Syst Video Technol 12(7):592–596
Koprinska I, Carrato S (2001) Temporal video segmentation: a survey. Signal Process: Image Commun 16(5):477–500
Krishan Kumar (2019) Evs-dk: event video skimming using deep keyframe. J Vis Commun Image Represent 58:345–352
Krishan Kumar (2021) Text query based summarized event searching interface system using deep learning over cloud. Multimed Tools Appl. 80:11079–11094
Krishan Kumar, Shrimankar Deepti D (2018) F-des: fast and deep event summarization. IEEE Trans Multimed 20(2):323–334
Krishan Kumar, Shrimankar Deepti D (2019) Esumm: event summarization on scale-free networks. IETE Tech Rev 36(3):265–274
Krishan Kumar P, Nishanth Maheep Singh, Dahiya Sanjay (2022) Image encoder and sentence decoder based video event description generating model: a storytelling. IETE J Educ 63(2):78–84
Kucuktunc O, Gudukbay U, Ulusoy O (2010) Fuzzy colour histogram-based video segmentation. Comput Vision Image Underst 114(1):125–134

Kundu MK, Mondal J (2012) A novel technique for automatic abrupt shot transition detection. 2012 International Conference on Communications, Devices and Intelligent Systems (CODIS)
Lakshmi Priya GG, Domnic S (2010) Video cut detection using block based histogram differences in rgb color space. International conference on Signal and Image Processing, pages 29–33
Lakshmi Priya GG, Domnic S (2014) Wals-hadamard transform kernel-based feature vector for shot boundary detection. IEEE Trans. on Image Process 23(12):5187–5197
Lankinen J, Kämäräinen JK (2013) Video shot boundary detection using visual bag-of-words, In Proceedings of the International Conference on Computer Vision Theory and Applications (VISAPP-2013), pp 788–791 1:788–791
Lee H, Yu J, Im Y, Gil JM, Park D (2011) A unified scheme of shot boundary detection and anchor shot detection in news video story parsing. Multimed Tools Appl 51(3):1127–1145
Lefevre S, Holler J, Vincent N (2003) A review of real-time segmentation of uncompressed video sequences for content-based search and retrieval. Real-Time Imaging 9:73–98
Li YN, Lu ZM, Niu XM (2009) Fast video shot boundary detection framework employing pre-processing techniques. IET Image process 3(3):121–134
Li L, Xu Q, Luo S, Sun X (2015) Key frame selection based on kl-divergence. IEEE International Conference on Multimedia Big Data (BigMM), pp 20-22
Li Z, Liu X, Zhang S (2016) Shot boundary detection based on multilevel difference of colour histograms. In: Proceedings of the 2016 First International Conference on Multimedia and Image Processing (ICMIP), p. 15–22
Liang L, Liu Y, Lu H, Xue X, Tan YP (2005) Enhanced shot boundary detection using video text information. IEEE Trans Consum Electron 51(2):580–588
Liao YH, Tsai CY, Su MH, Li HH, Yu PT (2011) Digital learning video indexing using scene detection. Int Conf Hybrid Learn 6837:336–344
Lienhart R (2001) Reliable transition detection in videos: a survey and practitioners guide. Int J Image Graph 13:469–486
Ling X, Chao L, Huan L, Zhang X (2008) A general method for shot boundary detection. Proceedings of international conference of Multimedia and Ubiqutous Engineering, pages 394–397
Liu T, Zhang HJ, Qi F (2003) A novel video key-frame-extraction algorithm based on perceived motion energy model. IEEE Trans Circuits Syst Video Tech 13(10):1006–1013
Liu TY, Lo KT, Zhang XD, Sinha SK, Fieguth PW (2004) A new cut detection algorithm with constant false -alarm ratio for video segmentation. J Vis Commun Image Represent 15:132–144
Liu Z, Zavesky E, Gibbson D, Shahraray B, Haffner P (2007) At & t research at trecvid 2007. TRECVID Workshop,
Liu H, Tan TH, Kuo TY (2020) A novel shot detection approach based on orb fused with structural similarity. IEEE Access 8:2472–2481
Lo CC, Wang SJ (2001) Video segmentation using a histogram based fuzzy c-means clustering algorithm. Comput Stand Interfaces 23(5):429–438
Lorenzo B, Costantino G, Cucchiara R (2017) A video library system using scene detection and automatic tagging. Italian Research Conference on Digital Libraries
Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vision 60:91–110
Lu Z, Shi Y (2013) Fast video shot boundary detection based on svd and pattern matching. IEEE Trans Image Process 22(12):5136–5145
Mahapatra D, Mariappan R, Rajan V, Yadav K, Seby A, Roy S (2018) Videoken: Automatic video summarization and course curation to support learning. In WWW (Companion Volume), pages 239–242
Meng J, Juan Y, Chang SF (1995) Scene change detection in a mpeg compressed video sequence. IS and T/SPIE Symposium Proceedings, California, 2419,
Ming B, Lyu D and Yu D (2021) Shot Segmentation Method Based on Image Similarity and Deep Residual Network (2021) IEEE 7th International Conference on Virtual Reality (ICVR), Foshan, China, pp 41–45. https://doi.org/10.1109/ICVR51878.2021.9483839
Mishra R, Singhai SK, Sharma M (2013) Video shot boundary detection using dual-tree complex wavelet transform. Advance Computing Conference (IACC), 2013 IEEE 3rd international, pages 1201–1206
Mohanta PP, Saha SK, Chanda B (2012) A model-based shot boundary detection technique using frame transition parameters. IEEE Trans Multimed 14(1):223–233
Mondal J, Kundu MK, Das S, Chowdhury M (2017) Video shot boundary detection using multiscale geometric analysis of nsct and least squares support vector machine. Multimed Tools Appl 77(7):1–23
Nagasaka A, Tanka Y (1991) Automatic video indexing and full video search for object appearances. In proc. IFIP 2nd conf. visual Data base systems, Budapest, Hungery, pages 113–127
Nam J, Tewfik AH (2005) Detection of gradual transitions in video sequences using b-spline interpolation. IEEE Trans Multimed 7(4):667–679

Nandini HM, Chethan HK, Rashmi BS (2020) Shot based keyframe extraction using edge-lbp approach. J King Saud Univ-Comput Inform Sci 34(7):4537–4545
Nandini HM, Chetan HK, Rashmi BS (2021) An efficient method for video shot transition detection using probability binary weight approach. IJCVIP 3:1–20
Ngo CW, Pong TC, Zhang HJ (2002) Motion-based video representation for scene change detection. Int J Comput Vision 50(2):127–142
Ngo CW, Ma YF, Zhang HJ (2005) Video summarization and scene detection by graph modeling. IEEE Trans Circuits Syst Video Technol 15(2):1237–1244
Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray scale and rotation invariant texture analysis with local binary patterns. IEEE Trans on Pattern Anal Mach Intell 24(7):971–987
Pal G, Rudrapal D, Acharjee S, Ray R, Chakraborty S, Dey N (2015) Video shot boundary detection A review. In Emerging ICT for Bridging the Future-Proceedings of the 49th Annual Convention of the Computer Society of India(CSI) 338:119–127
Panchal P, Merchant SN (2012) Performance evaluation of fade and dissolve transition shot boundary detection in presence of motion in video. In 2012 1st International Conference on Emerging Technology Trends in Electronics, Communication Networking, pages 1–6,
Parmar M, Angelides MC (2015) Mac-realm: a video content feature extraction and modelling framework. Comput. J. 58(9):2135–2171
Pickering MJ, Rüger S (2003) Evaluation of key frame-based retrieval techniques for video. Comput Vision Image Understand 92(2–3):217–235
Porter S, Mirmehdi M, Thosmas B (2003) Temporal video segmentation and classification of edit effects. Image Vis Comput 21(13):1097–1106
Qi Y, Hauptmann A, Liu T (2003) Supervised classification for video shot segmentation. Proceed IEEE Conf Multimed Expo 2:689–692
Qian X, Liu G, Su R (2006) Effective fades and flashlight detection based on accumulating histogram difference. IEEE Trans Circ Syst Video Technol 16(10):1245–1258
Raja Suguna M, Kalaivani A, Anusuya S (2022) 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, MDPI 14:1705–1720
Ramli AR, Mahmmod BM, Abdulhussain SH et al (2019) Shot boundary detection based on orthogonal polynomial. Multimed Tools Appl 78:20361–20382
Ranjan RK, Agrawal A (2016) Video summary based on f-sift, tamura textural and middle level semantic feature. Procedia Comput Sci 89:870–876
Rao Anyi, Xu Linning, Xiong Yu, Xu Guodong, Huang Qingqiu, Zhou Bolei, Lin Dahua (2020) A local-to-global approach to multi-modal movie scene segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 10146–10155,
Rasheed Z, Shah M (2005) Detection and representation of scenes in videos. IEEE Trans Multimed 7(6):1097–1105
Rashmi BS, Nagendraswamy HS (2018) Effective video shot boundary detection and keyframe selection using soft computing techniques. IJCVIP 2:27–48
Rashmi BS, Nagendraswamy HS (2021) Video shot boundary detection using block based cumulative approach. Multimed Tools Appl 80:641–664
Sasithradevi A, Roomi SMM (2016) Video shot boundary detection using normalized periodogram distance metric. Circuits Syst 7(10):2875
Sasithradevi A, Roomi SMM (2020) A new pyramidal opponent color-shape model based video shot boundary detection. J Vis Commun Image Represent. https://doi.org/10.1016/j.jvcir.2020.102754
Sasithradevi A, Roomi M M, Gupta S (2022) Pyramidal-relative entropy based temporal signature for video transition detection using lstm. PREPRINT (Version 1) available at Research Square
Sasithradevi A, Roomi SMM, Maheesha M (2018) Shot boundary detection in videos using saliency based statistical model. In 11th Indian Conference on Computer Vision, Graphics and Image Processing, pages 1–7,
Sasithradevi Roomi SMM, Nirmala AP (2023) Visual significance model based temporal signature for video shot boundary detection. Multimed Tools Appl 32(82):23037–23054
Selesnick W, Baraniuk RG, Kingsbury NC (2005) The dual-tree complex wavelet transform. IEEE Signal Process Magaz 22(6):123–151
Sengupta A, Singh KM, Thounaojam DM, Roy S (2015) Video shot boundary detection: A review. IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)
Sheena CV, Narayanan NK (2015) Key-frame extraction by analysis of histograms of video frames using statistical methods. Procedia Comput Sci 70:36–40

Shekar BH, Sharmila Kumari M, Holla R (2011) Shot boundary detection algorithm based on color texture moments. Comput Netw Inform Technol 142:591–594
Shekar BH, Kirsch Uma KP (2015) An efficient and accurate method Directional derivatives based shot boundary detection. Procedia Comput Sci 58:565–571
Shi Y, Yang H, Gong M, Liu X (2017) A fast and robust key frame extraction method for video copyright protection. J Electri Compute Eng 3:7. https://doi.org/10.1155/2017/1231794
Shiguo Lian (2011) Automatic video temporal segmentation based on multiple features. Soft Comput 15:469–482
Shu H, Chau LP (2005) A new scene change feature for video transcoding. IEEE Int Symp Circuits Syst. https://doi.org/10.1109/ISCAS.2005.1465652
Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scaleimage recognition. arXiv preprint arXiv:1409.1556,
Singh RD, Aggarwal N (2015) Novel research in the field of shot boundary detection - a survey. Advances in Intelligent Informatics. Advances in Intelligent Systems and Computing, vol 320. Springer, Cham, pages 457–469,
Singh A, Thounaojam DM, Chakraborty S (2019) A novel automatic shot boundary detection algorithm:robust to illumination and motion effect. Signal Image Video Process 14(4):645–653
Soucek T, Moravec J, Lokoc J (2019) Transnet: A deep network for fast detection of common shot transitions. arXiv:1906.03363v1 [cs.CV],
Souček Tomáš, Lokoč Jakub (2020) Transnet v2: An effective deep network architecture for fast shot transition detection
Sumiyoshi H, Kawai Y, Yagi N (2007) Shot boundary detection at trecvid. TRECVID 2007 workshop
Sun L, Zhou Y (2011) A key frame extraction method based on mutual information and image entropy. In: 2011 International Conference on Multimedia Technology, Hangzhou, China, 26-28 https://doi.org/10.1109/ICMT.2011.6001938
Sun J, Wan Y (2014) A novel metric for efficient video shot boundary detection. 2014 IEEE Visual Communications and Image Processing Conference, pages 45–48,
Smeaton AF, Over P, Doherty AR (2010) Video shot boundary detection: seven years of TRECVid activity. Comput Vision Image Underst 114(4):411–418
Snoek CGM, Worring M (2005) Multimodal video indexing: a review of the state-of-the-art. Multimed Tools Appl 25:5–35
Taile P, Wenjun Z (2014) Robust shot boundary detection from video using dynamic texture. Sens Transducers 167(3):104–109
Tamura H, Mori S, Yamawaki T (1978) Textural features corresponding to visual perception. IEEE Trans Syst Man Cybern 8(6):460–473
Tan W, Teng S, Zhang W (2007) Research on video segmentation via active learning. In Fourth International Conference on Image and Graphics (ICIG 2007), pages 395–400
Tang S, Feng L, Kuang Z, Chen Y, Zhang W (2018) Fast video shot transition localization with deep structured models. ACCV, Lecture Notes in Comput Sci 11361:577–592
Tang Shitao, Feng Litong, Kuang Zhanghui, Chen Yimin, Zhang Wei (2018) Fast video shot transition localization with deep structured models. arXiv,
Tao MW, Bai J, Kohli P, Paris S (2012) Simpleflow: a noniterative, sublinear optical flow algorithm. Comput Graph Forum (Eurographics). https://doi.org/10.1109/ICDSP.2013.6622827
The open video project. [online] available at :http://www.open-video.org
Thounaojam DM, Khelchandra T, Singh KM, Roy SA (2016) Genetic algorithm and fuzzy logic approach for video shot boundary detection. Comput. Intell. Neurosci, 14
Tippaya S, Sitjongsataporn S, Tan T, Chamnongthai K (2014) Abrupt shot boundary detection based on averaged two-dependence estimators learning. 14th International Symposium on Communications and Information Technologies (ISCIT), pages 522–526
Tong W, Song L, Yang X, Qu H, Xie R (2015) Cnn-based shot boundary detection and video annotation,. IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), https://doi.org/10.1109/BMSB.2015.7177222,pages 1–5
Warhade KK, Merchant SN, Desai UB (2011) Shot boundary detection in presence of fire flicker and explosion using stationary wavelet transform. Signal Image Video Process 5(4):507–515
Warhade KK, Merchant SN, Desai UB (2013) Shot boundary detection in the presence of illumination and motion. Signal, Image Video Process 7(3):581–592
Wu L, Zhang M, Jian S, Lu Z, Wang D (2019) Two stage shot boundary detection via feature fusion and spatial-temporal convolutional neural networks. IEEE Access 7(2):77268–77276
Xiong Z, Radhakrishnan R, Divakaran A, Rui Y, Huang TS A unified framework for video summarization, browsing, and retrieval. Soft computing, pages 221–235

Xuekun Jiang, Libiao Jin, Anyi Rao, Linning Xu, Dahua Lin (2022) Jointly learning the attributes and composition of shots for boundary detection in videos. IEEE Trans Multimed 24:3049–3059
Yan Fu, Renjie Guo, Ye Ou (2022) A novel shot boundary detection technique for illumination and motion effects. In: Yulin Wang, Siting Chen (eds) International Conference on High Performance Computing and Communication, volume 12162, page 1216208. International Society for Optics and Photonics, SPIE
You J, Liu G, Periks A (2010) A semantic framework for video genre classification and event analysis. Signal Process: Image Commun 25(04):287–302
Youssef B, Fedwa E, Driss A, Ahmed S (2017) Shot boundary detection via adaptive low rank and svd-updating. Computer vision and Image Underst 161:20–28
Yuan J, Wang H, Xiao L (2007) A formal study of shot boundary detection. IEEE Trans Circuits Syst Video Technol 17(2):168–186
Yuan Y, Zhang J (2023) Shot boundary detection using color clustering and attention mechanism. ACM Trans Multimed Comput Commun Appl 19(6):1–23
Vasconcelos N (2003) Feature selection by maximum marginal diversity:optimality and implications for visual recognition. Proc. IEEE Computer Society Conf. Comput Vision Pattern Recognit 1:762–772
Vasconcelos N, Vasconcelos M (2004) Scalable discriminant feature selection for image retrieval and recognition. Proc. IEEE Computer Society Conf. Comput Vision Pattern Recognit 2:770–775
Vinicius VMC, Pedrini H (2018) Viscom: a robust video summarization approach using color co-occurrence matrices. Multimed Tools Appl 77:857–875
VIVA Research Lab.[Online] Available at: http://www.site.uottawa.ca/ laganier/videoseg/
Wang X, Wang S, Chen H (2007) A fast algorithm for mpeg video segmentation based on macroblock. Fourth International Conference on Fuzzy Systems and Knowledge Discovery,2007,
Wang DH, Tian Q, Gao S, Sung WK (2014) News sports video shot classification with sports play field and motion features. International Conference on Image Processing, ICIP 4 : 2247-2250. ScholarBank@NUS Repository
Wang T, Feng N, Yu J, He Y, Hu Y, Chen YP (2021) Shot boundary detection through multi-stage deep convolution neural network. MultiMedia Modeling. MMM 2021. Lect Notes Comput Sci 12572:456–468
Warhade KK, Merchant SN, Desai UB (2011) Performance evaluation of shot boundary detection metrics in the presence of object and camera motion. IETE J Res 57(5):461–466
Xu J, Song L, Xie R (2016) Shot boundary detection using convolutional neural networks. In 2016 Visual Communications and Image Processing (VCIP), pages 1–4
Xuemei Sun, Xiaoyu Lv, Mingwei Zhang (2010) Novel shot boundary detection method based on support vector machine. pages 56–59, 3-5
Yoo HW, Ryoo HJ, Jang D (2006) Gradual shot boundary detection using localised edge blocks. Multimed Tools Appl 28:283–300
Yuan J, Zheng W, Ding L, Wang D, Tong Z, Wang H, Wu JLJ, Lin F, Zhang B (2004) Shot boundary detection and high-level feature extraction. TRECVID Workshop,
Zabih R, Miller J, Mai K (1995) A feature based algorithm for detecting and classifying scene breaks. In Proceedings of the Third ACM International Conference on Multimedia; San Francisco, CA, USA,, 95:189–200, 5-9
Zhang HJ, Kankanhalli A, Smoliar SW (1993) Automatic partitioning of full motion video. Multimed Syst 1(1):10–28
Zhang D, Lei W, Zhang W, Chen X (2019) Shot boundary detection based on block-wise principal component analysis. J Electr Imaging, SPIE 28(2):1–11. https://doi.org/10.1117/1.JEI.28.2.023029
Zhou X, Wu S, Qi Y et al (2021) Video shot boundary detection based on multi-level features collaboration. SIViP 15:627–635

Authors and Affiliations

T. Kar[1] · P. Kanungo[2] · Sachi Nandan Mohanty[3] · Sven Groppe[4] · Jinghua Groppe[4]

✉ T. Kar
tkarfet@kiit.ac.in

P. Kanungo
pkanungo@gmail.com

Sachi Nandan Mohanty
sachinandan09@gmail.com

Sven Groppe
sven.groppe@uni-luebeck.de

Jinghua Groppe
jinghua.groppe@uni-luebeck.de

[1] School of Electronics Engineering, KIIT Deemed to be University, Bhubaneswar, Odisha, India

[2] C. V. Raman Global University, Bhubaneswar, Odisha, India

[3] School of Computer Science & Engineering (SCOPE), VIT-AP University, Amaravati, Andhra Pradesh, India

[4] Institute of Information Systems (IFIS), University of Lübeck, Lübeck, Germany