The Visual Computer (2022) 38:445–456
https://doi.org/10.1007/s00371-020-02027-9

ORIGINAL ARTICLE

A novel bifold-stage shot boundary detection algorithm: invariant to motion and illumination

Saptarshi Chakraborty[1] · Alok Singh[1] · Dalton Meitei Thounaojam[1]

Accepted: 12 November 2020 / Published online: 2 January 2021

Abstract
Shot boundary detection is mainly considered as a stepping stone in the broad arena of content-based video retrieval. Ample systematic investigation has been carried out in the terrain of shot boundary detection. The attainment of shot boundary detection procedures is greatly hindered due to the presence of unforeseen illumination change and motion effects in a video. This work proposes a novel bifold-stage technique to recognize abrupt transition in videos, invariant to motion, and illumination effects. In the first stage, the local ternary patterns feature is used to extract information from each frame in a video. Then, a set of novel adaptive thresholds such as γ and β are used to find the possible transition frames. In the confirmation stage, Lab color difference along with an adaptive threshold δ is used to extract actual transition frames. The experimental result depicts that the motion effect is also scaled down in the initial stage. The Lab color difference passed down in the second stage also handles the illumination and motion effects which are not managed in the initial stage. Experimentation is done using TRECVid 2001 and 2007 standard datasets and palpable that the proposed technique outperforms most of the contemporary shot boundary detection approaches.

Keywords LTP · Shot boundary detection · Abrupt · Adaptive threshold · Illumination invariance · Object motion

1 Introduction

In today's age of the Internet and the advancement of cutting edge technologies, the evolution of multimedia data is increasing by leaps and bounds. From social networking to online shopping, everywhere the demand for digital data is increasing exponentially. So it is very cumbersome to handle or effectively process this sort of digital data. The lion share of these digital data (in size per data) comprises of multimedia videos. Due to the inherent unstructured property of the video, it is very difficult to manage and retrieve a video. The use of the name as a prime attribute for indexing and accessing data items has become obsolete nowadays.

To sketch a solution, for the above problem the structural characteristics of the video are effectively used for indexing and retrieval which ascends to the necessity of content-based video retrieval system. In excerpting the entire gamut of a video, a video is subdivided into consequential intra-relevant frames defined as shots. In an abridgment, a shot can be defined as a series of correlated information for a specific chronological length of frames in a video. The activity of extracting the periphery among two chronological shots is labeled as shot boundary detection (SBD). In the context of SBD, a transition is of basically two types: abrupt and gradual transitions [1,2]. Abrupt transition is also known as cut transition; this type of transition suddenly crops up among two successive frames. A gradual transition is created due to the incorporation of editing effects in multiple consecutive frames. A gradual transition is predominantly clustered into a fade, dissolve and wipes. This breed of transition prolongs to a sanguine length of frames. Fade transition is broadly classified as fade-in and fade-out transitions. When a shot materializes deliberately from a monochromatic frame is considered as fade-in transition; on the other hand, when a shot dissipates to a monochromatic frame is termed as fade-out transition. The protruding of fade-out and fade-in (without the monochromatic frames), where the current shot starts receding and the next shot starts breezing in simultaneously, is named as dissolve transition. A wipe transition

✉ Saptarshi Chakraborty
chakraborty0007@gmail.com

[1] Computer Vision Laboratory, Department of Computer Science and Engineering, National Institute of Technology Silchar, Assam, Silchar, India

occurs with an animated effect, and it is generally found in sports and news videos.

The major challenges still persisting in the shot boundary detection arena are to develop an illumination, object, and camera motion (OCM) invariant shot boundary detection approach and to define an adaptive threshold to classify transition and non-transition frames across videos. Sudden illumination and OCM challenges in the detection of abrupt transition result in high false positive [1]. Many researchers in the field of shot boundary detection prefer a histogram-based approach due to its low computational cost and motion invariant advantages [1,3,4]. Discrete cosine transform (DCT) transform plays an efficient role to reduce sudden illumination changes in a video and then after shot boundaries are detected using histogram difference approaches [2]. Some researchers have also experimented with DCT and wavelet transform to undermine the illumination changes [5]. It also used an adaptive threshold for detecting shot boundaries. Some algorithms are proposed based on the cross-correlation coefficient and stationary wavelet transform features [6]. In such approaches shot boundaries are detected using a combination of the local and adaptive threshold. In some cases, a dual-tree complex wavelet transform is used for analyzing the structural feature of each and every frame [7]. Here, shot boundaries are computed using these structural similarity features along with an adaptive threshold. In some cases, a block-based center symmetric local binary feature vector is also recycled to identify shot boundaries [8]. Few sparse representation-based approaches [9,10] are also proposed for boundary detection. [9] proposed a boundary detection approach using sparse coding for selecting a keyframe efficiently for video summarization.

Edge-based features play a crucial role in neglecting the effect of sudden illumination and motion effect [1]. A transition is announced when the edges of the prevailing frame display a hefty variation with the edges of the preceding frame that have dissipated [11,12]. A block matching algorithm is used to compute the motion strength to reduce motion effects [13]. A fast SBD algorithm using pixel information is proposed in [14].

Machine learning techniques also have some applications in shot boundary detection. In [15] a genetic and fuzzy logic-based technique is proposed to detect boundaries. In some cases combining with the adaptive threshold, a convolution neural network is used to extract the possible candidate segment in prepossessing steps [16–19].

Many researchers have employed some essential features like structural similarity index and standard deviation [20], quantized hue, saturation and value (HSV) color space feature [12], histogram intersection [21], along with feature difference using absolute sum gradient [22] to reduce the illumination and motion effects in a video. In object-based SBD approaches, a time stamp is attached to an object for identifying that object in the number of frames [23,24]. Some of the drawbacks in object-oriented SBD are the exodus of an object from the frame, an enormous object moving which is erroneous as wipe transition, and irregular illumination in a video [24]. Some object tracking algorithms are developed in [25–27] to tackle the sudden illumination effect in videos. The multi-view spatiotemporal feature points and grid-based matching approach used for video stitching in [38] can also be helpful in SBD.

In a few approaches, local binary pattern (LBP) is used as an illumination invariant feature to detect shot boundaries [28–30]. Due to some pitfalls persisting in LBP, some researchers preferred the local ternary pattern (LTP) feature, a conjecture of LBP [31]. LTP is basically a local texture descriptor that is further discriminant and lowers susceptible to noise in a uniform suburb. Due to its less sensitivity to noise, the effects of sudden illumination change is nullified, hence preserving its essential properties.

The literature review correctly articulates that the sudden illumination and OCM effects are the sizable hurdles in abrupt transition detection. The frames suffering from these hurdles are falsely classified as abrupt transitions thereby reducing the precision of the system. To propose an illumination and OCM invariant method is a challenging task. The paper proposes a bifold stage abrupt transition detection approach where the LTP feature is used at the initial stage and Lab color difference in the confirmation stage which is invariant to irregular illumination and OCM effects.

The notable findings of the paper are:

i. A bifold-stage shot boundary detection technique is scripted to counter the sudden illumination and motion effect across videos.
ii. Adaptive thresholds have been proposed to efficiently classify possible transition frames in the initial stage and actual transition frames in the confirmation stage.

Further, the organization of paper is as follows: Sect. 2 gives a brief background knowledge of the features used in the proposed approach. In Sect. 3, detail of proposed approach is given. Section 4 presents a detailed discussion of experimental results and parameter settings, followed by Sect. 5 which reports the conclusion of the work done.

2 Background knowledge

This section briefly discusses about the frame features used in the proposed system.

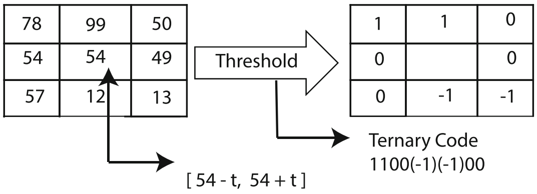


Fig. 1 Illustration of the basic LTP operator

2.1 Local ternary patterns

In the field of texture classification features which are very much reliable and highly discriminative in nature are preferred. LBP features having those characteristics tend to resist against lighting effects, which fall under this category [32]. In most of the cases, it is found that LBP features are resistant against monotonic gray-level transformations. Research clearly depicts that in some cases LBP lacks in context to sensitivity against noise, most probably in near-uniform image regions.

Due to the above problem, a more robust texture operator, i.e., local ternary pattern (LTP), is formulated which is more resistant to noise [31]. LTP has formulated a new pattern where the neighboring pixels are concealed to a three-valued code, i.e., $(-1, 0, 1)$, in comparison with the LBP-based approach in which a two valued code $(0, 1)$ is depicted. This process is carefully processed using a user-defined threshold t. In LTP, the three-valued code is calculated in accordance to the centre pixel i_c as given in Eq. 2. The value generated after comparing with the center pixel higher than the threshold yields + 1 else − 1. The generated value is considered 0 if the difference falls inside the range of the threshold.

Here the gray value of the centre pixel along with the neighboring pixel are denoted as i_c and i_p $(p = 0, \ldots, P-1)$ in LTP. The radius of the circle formed incorporating the neighboring pixel is denoted as R and p to define the count of the neighboring pixels. To make sure that the neighboring pixels do not fall at the center of the pixel, an estimation method is defined and known as bilinear interpolation:

$$\mathrm{LTP}_{P,R} = \sum_{p=0}^{P-1} 2^p s\left(i_p - i_c\right), \tag{1}$$

$$s(x) = \begin{cases} 1, & x \geq i_c + t \\ 0, & i_c - t < x < i_c + t \\ -1, & x \leq i_c - t \end{cases} \tag{2}$$

where x is the neighbor pixel values.

Here, the use of threshold makes LTP invariant to noise. The LTP encoding procedure is illustrated in Fig. 1 where the threshold is set to $t = 5$, so the tolerance interval is [49, 59].

2.2 CIEDE 2000 color difference

CIEDE 2000 is an efficient color-difference formula that correctly distinguishes different colors perceived through human eyes. This formula (ΔE) is basically based on CIELab color space [33]. The difference between each pair of color values in the context of CIELab space is computed using Eq. 3:

$$\Delta E_{00} = \Delta E_{00}\left(L_1^*, a_1^*, b_1^*, L_2^*, a_2^*, b_2^*\right)$$
$$= \sqrt{\left(\frac{\Delta L'}{K_L S_L}\right)^2 + \left(\frac{\Delta C'}{K_C S_C}\right)^2 + \left(\frac{\Delta H'}{K_H S_H}\right)^2 + R_T \left(\frac{\Delta C'}{K_C S_C}\right)^2 + \left(\frac{\Delta H'}{K_H S_H}\right)^2} \tag{3}$$

Here, lightness, chroma, and hue differences among the pair of samples in CIEDE2000 are efficiently calculated using $(\Delta L')$, $(\Delta C')$, and $(\Delta H')$, respectively. In the blue region, an interaction between chroma and hue difference is encountered or efficiently executed using a rotation function (R_T). For a modification of a^* in the case of CIELab some specific weighting functions, parameter related factors along with a rotation term that basically affects the colors which have low chroma values are also added. The CIEDE2000 that has been inducted in CIELab has five major corrections. As compared to the previous version of the formula the S_L, S_C, and S_H are Compensation for lightness, chroma, and the hue, respectively, and K_L, K_C, and K_H are weighting parameters. This modification which has been undertaken in Eq. 3 correctly maps to the primed values in context to lightness, chroma, and hue differences. The similarity measure between frames in a video sequence is computed using CIEDE 2000 color difference.

3 Proposed approach

This section discusses the proposed approach. Figure 2 shows the block diagram of the proposed approach.

3.1 Preprocessing

Preprocessing is the first step in the proposed approach which includes:

1. Converting the video frames from RGB color space to grayscale.
2. Resizing each frame to $R \times S$ where $R = 130$ and $S = 150$.

3.2 Abrupt transition detection

For properly classifying abrupt transition an innovative automatic bifold-stage shot boundary detection approach is

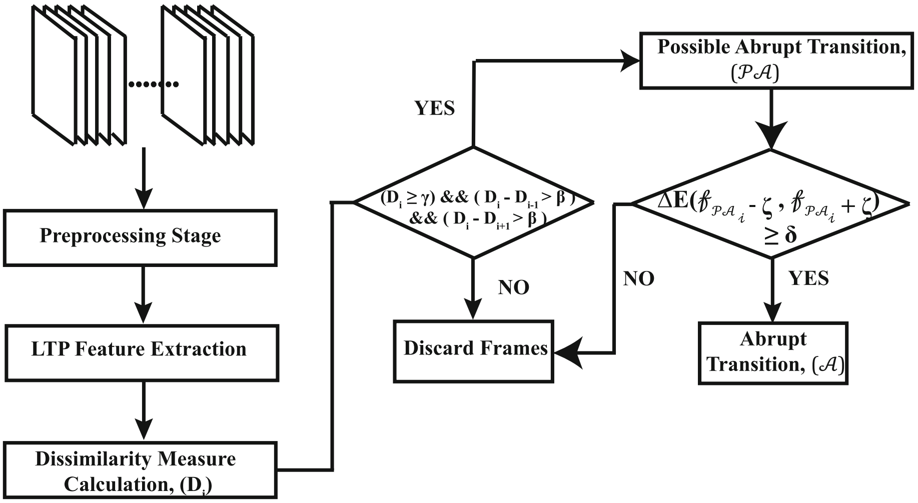

Possible Abrupt Transition, $(\mathcal{PA})$
YES
Preprocessing Stage
$(D_i \geq \gamma)$ && $(D_i - D_{i-1} > \beta)$ && $(D_i - D_{i+1} > \beta)$
$\Delta E(\mathfrak{f}_{\mathcal{PA}_i} - \zeta, \mathfrak{f}_{\mathcal{PA}_i} + \zeta) \geq \delta$
NO
NO
YES
LTP Feature Extraction
Discard Frames
Abrupt Transition, $(\mathcal{A})$
Dissimilarity Measure Calculation, (D_i)

Fig. 2 Block diagram for proposed system

scripted. The whole detection procedure is divided into possible stage and confirmation stage. The function of the possible stage is to eliminate non-transition frames, on the other hand, the function of the confirmation stage is to detect actual transition frames. The whole process drastically reduces the false detection rate.

3.2.1 Possible stage

This section briefly explains the possible stage of the proposed system which consists of feature extraction, similarity measure, threshold selection, and possible transition detection.

3.2.1.1 Feature extraction and similarity measure
After properly accomplishing the prepossessing stage of the proposed system, LTP features of each and every frame are computed for a video. The reason to incorporate the LTP feature is that it has an advantage of more discriminant and less sensitive to noise in uniform regions as compared with the histogram-based approaches.

The dissimilarity difference D_i between the extracted LTP codes histogram of consecutive frames is calculated using Chi-square histogram distance [34] as given in Eq. 4. The possible abrupt transitions are marked based on the thresholds discussed in Sect. 3.2.1.2:

$$D_i(x, y) = \sqrt{\frac{1}{2} \sum_{j=1}^{n} \frac{(h_j(x) - h_j(y))^2}{(h_j(x) + h_j(y))}} \tag{4}$$

Here, h_j depict the histogram value of jth bin and $D_i(x, y)$ will store the dissimilarity difference value between the consecutive frames x and y.

3.2.1.2 Threshold selection
An important criterion in shot boundary detection is to properly classify transition and non-transition frames. For attaining this goal a suitable threshold is to be determined. The role of a threshold is to declare an abrupt transition when the distance difference between the consecutive frames is beyond some value. The importance of selecting the right threshold will yield high accuracy in classifying abrupt transition. As the structure and behavior of video data are very random, so it is very cumbersome to set a unique threshold (hard threshold) to work efficiently for all videos. So an adaptive sort of threshold is need of the hour which can adapt according to the structure of the video. In our proposed algorithm, a set of adaptive threshold is selected, namely γ and β, to properly filter out non-transition frames as given in Eqs. 5 and 6, respectively:

$$\gamma = \mu_D + \sigma_D \times \kappa_1 \tag{5}$$

$$\beta = \sigma_D + \mu_D \tag{6}$$

where μ_D and σ_D are the mean and standard deviation of the dissimilarity measure of a video. κ_1 is a user-defined constant which is set as 1.9 experimentally.

3.2.1.3 Possible transition detection
In the possible stage initially, LTP features of each and every frame of a video are extracted. Then, a similarity measure between the corresponding frames is evaluated using Eq. 4. From the dissimilarity values D, it is observed that an abrupt transition is encountered when the dissimilarity value of the ith and $i + 1$th frames given by D_i is greater than equal to threshold γ. Further, it is also observed that the difference of the dissimilarity values between D_i and $D_{i\pm 1}$ is greater than the adaptive threshold β as shown in Eq. 7:

$$\mathscr{PA} = \begin{cases} \text{true}, & (D_i \geq \gamma)\&\&(D_i - D_{i-1} > \beta) \\ & \&\&(D_i - D_{i+1} > \beta) \\ \text{false}, & \text{Otherwise} \end{cases} \tag{7}$$

Due to the elimination of a large number of non-transition frames, a handful of frames are left for consideration in the confirmation stage.

3.2.2 Confirmation stage

Similarly, this section briefly explains the confirmation stage of the proposed system which consists of feature extraction, dissimilarity measure, threshold selection, and actual transition detection.

3.2.2.1 Feature extraction and dissimilarity measure
In this stage, only the possible transition frames are considered. The detected frames are converted to Lab color space, f. Lab color feature is drafted in such a fashion to resemble human vision. It strives for perceptual uniformity, and its L component meticulously contests the human perception of lightness. It can be used to make authentic color equity alterations by modifying output curves in the a and b components or to regulate the lightness contrast using the L component.

Then, the Lab color difference is calculated using Eq. 3 which is represented as ΔE.

3.2.2.2 Threshold selection
For the confirmation stage, an adaptive threshold δ is proposed which is given in Eq. 8:

$$\delta = \mu_{\Delta_E} + \kappa_2 \times \sigma_{\Delta_E} \tag{8}$$

where μ_{Δ_E} and σ_{Δ_E} are the mean and standard deviation of the dissimilarity measure of a video. κ_2 is a user-defined constant which is set as 0.8 experimentally.

3.2.2.3 Actual transition detection
From the possible stage, a handful of possible transition frames are mined. The sole work of the confirmation section is to verify that all the possible transition frames can be effectively mapped into actual transition frames. Due to the use of the initial screening stage, most of the sudden illumination and motion effects frames are drastically reduced. Then, also there may be a rare possibility that some of the illumination and motion effects frames may creep into as a member of the possible transition frames set that are fed into the later stage. So, a post-processing stage (or confirmation stage) for determining correct transition and false reduction is highly required.

In the confirmation stage, the frames $f_{\mathscr{PA}_i - \zeta}$ and $f_{\mathscr{PA}_i + \zeta}$ are used for ensuring the conformity of the possible transition frames where $f_{\mathscr{PA}_i}$ is the possible abrupt frame given by index $\mathscr{PA}_i$. Thus, $f_{\mathscr{PA}_i \pm \zeta}$ are preceding and upcoming frames given by ζ from $f_{\mathscr{PA}_i}$ where the value ζ is fixed as 4. It is observed that $f_{\mathscr{PA}_i}$ is an actual abrupt transition if the Lab color difference (ΔE) between $f_{\mathscr{PA}_i - \zeta}$ and $f_{\mathscr{PA}_i + \zeta}$ is greater than or equal to threshold δ. So, Eq. 9 is used for ensuring the confirmation of all probable abrupt transition ($\mathscr{PA}_i$).

$$\mathscr{A} = \begin{cases} \text{true}, & \text{if } \Delta E\left(f_{\mathscr{PA}_{i-\zeta}}, f_{\mathscr{PA}_{i+\zeta}}\right) \geq \delta \\ \text{false}, & \text{otherwise.} \end{cases} \tag{9}$$

The pseudocode of the proposed system is clearly depicted in Algorithm 1.

Algorithm 1 Pseudocode for the Proposed System.

```
Input: Video, V
Output: Shot Boundaries, $Final\_cut$
procedure SHOT_DETECTION($V$)
    $F \leftarrow VideoReader(V)$;
    $f_1 \leftarrow preprocessing(F_1)$
    $l_1 \leftarrow ltp\_frame(f)$  ▷ Possible Stage ▷ LTP feature extraction
    for $i = 2$ to $length(F)$ do
        $f_2 \leftarrow preprocessing(F_i)$
        $l_2 \leftarrow ltp\_frame(f_2)$
        $D_{i-1} \leftarrow dist\_chi(l_1, l_2)$    ▷ Finding chi square distance between the consecutive frames
        $l_1 \leftarrow l_2$;
    $D \leftarrow D/max(D)$
    for $i = 2$ to $length(D) - 1$ do
        if $D_i \geq \gamma \&\& (D_i - D_{i-1}) > \beta \&\& (D_i - D_{i+1}) > \beta$ then  ▷ Possible Abrupt
            $\mathscr{PA} \leftarrow record\ possible\ abrupt\ between\ i$th $and\ (i+1)$th $frame$;
        else
            $Discard\ Frame$;
    for $i = 1\ to\ length(\mathscr{PA}) - 1$ do    ▷ Confirmation Stage
        if $\Delta E(f_{\mathscr{PA}_{i-\zeta}}, f_{\mathscr{PA}_{i+\zeta}}) \geq \delta$ then    ▷ Actual Abrupt
            $\mathscr{A}_i \leftarrow record\ abrupt\ transition\ between\ i^{th}\ and\ (i+1)$th $frame$;
            $Final\_cut \leftarrow \mathscr{A}$;
        else
            $Discard\ Frame$;
```

4 Experimental results and discussion

4.1 Database

The database plays an important metric for validating the results mined through the proposed approach. Here, the database videos consist of selected TRECVid 2001 and 2007 videos. To make our database videos more dynamic, we have included some video and small movie clips which are subject to more lighting, illumination, and motion effect; for example, "Transformer (T1)," "Mission impossible (M1)" and a song of the movie "Massom" are used. Our experiments are carried out using HP-Z220 workstation. The overall details of the selected videos of our dataset are given in Table 1.

Table 1 Ground truth data of the test videos

Video	Frames	Transition			Sources
		Abrupt	Gradual	Total	
$D2$	16586	42	31	73	TRECVid 2001
$D3$	12304	39	64	103	
$D4$	31389	98	55	153	
$D5$	12508	45	26	71	
$D6$	13648	40	45	85	
BG_3027	49815	127	1	128	TRECVid 2007
BG_3097	44991	91	–	91	
BG_3314	35802	44	1	128	
BG_16336	2466	127	1	128	
BG_28476	23238	176	3	179	
BG_36136	29426	88	21	109	
BG_37309	9639	11	10	21	
BG_37770	15836	8	29	37	
$ClipM1$	3444	63	–	63	Mission impossible
$ClipT1$	7721	38	–	38	Transformer
Massom	9193	41	–	41	Movie song

4.2 Evaluation parameters

The performance evaluation of the proposed system is computed using recall (R), precision (P) and $F1$ score ($F1$) performance metrics through Eqs. 10, 11 and 12, respectively:

$$R = \frac{\text{Correctly detected}}{\text{Correctly detected } + \text{ Miss detected}} \times 100 \tag{10}$$

$$P = \frac{\text{Correctly detected}}{\text{Correctly detected } + \text{ Wrongly detected}} \times 100 \tag{11}$$

$$F1 = \frac{2 \times R \times P}{R + P} \tag{12}$$

4.3 Parameters selection

The performance of the system totally depends on the proper selection of the parameters used in the proposed approach. We have used three adaptive thresholds γ, β and δ which are discussed in Sects. 3.2.1.2 and 3.2.2.2.

The thresholds γ and β are used for extracting probable abrupt changes and δ is used in the confirmation stage for ensuring conformity of the probable abrupt changes where the adaptation of these thresholds can be seen in Table 2.

The experimental results depicts that the appropriate range of constants κ_1 and κ_2 used in Eqs. 5 and 8 are [1 3] and [0.5 1.5], respectively. Throughout the experimentation the values of κ_1 and κ_2 are set as 1.9 and 0.8, respectively.

Table 2 Adaptive thresholds for different videos

Video	Possible stage		Confirmation stage
	(γ)	(β)	(δ)
$D2$	0.2601	0.2022	0.2897
$D3$	0.1946	0.1541	0.2206
$D4$	0.1839	0.1453	0.2136
$D6$	0.2124	0.1519	0.2259

4.4 System performance

The overall performance of the proposed system using TRECVid 2001 and 2007 selected videos is shown in Table 3.

Table 3 depicts the performance analysis of selected videos of TRECVid 2001 and TRECVid 2007 dataset. In TRECVid 2001 92.3%, 99.1%, 95.5% and 757 s are the recorded average *recall*, *precision*, *F1 score* and *computation time* in the proposed system. Similarly, 99.3%, 96.7%, 97.9% and 1012 s are the recorded average values for the TRECVid 2007 dataset, respectively. Finally, 96.7%, 98.0%, 96.8% and 812 s are the overall average performance depicted for the proposed system. Figure 3 shows correctly detected abrupt transitions between frame numbers 468 and 469 from "D6.mpg" video.

4.5 Discussion

Our proposed algorithm has correctly detected most of the abrupt transitions present in the clip. The proposed algorithm

Table 3 Performance of the system for TRECVid 2001 and 2007

Videos	Parameter measure			Computation time in seconds (approx.)
	R	P	$F1$	
$D2$	92.9	100.0	96.3	700
$D3$	87.2	100.0	93.2	510
$D4$	86.8	97.8	92.0	1350
$D5$	100.0	98.0	99.0	678
$D6$	95.0	100.0	97.4	550
BG_3027	100.0	99.2	99.6	1559
BG_3097	98.0	100.0	99.0	1406
BG_16336	100.0	100.0	100.0	79
BG_28476	98.8	98.3	98.5	728
BG_36136	98.8	100.0	99.4	1826
BG_37309	100.0	100.0	100.0	603
BG_37770	100.0	80.0	89.0	888
$ClipM1$	93.6	100.0	96.7	195
$ClipT1$	92.1	97.2	94.6	677
Massom	97.5	100.0	98.7	484
Average	96.7	98.0	96.8	812

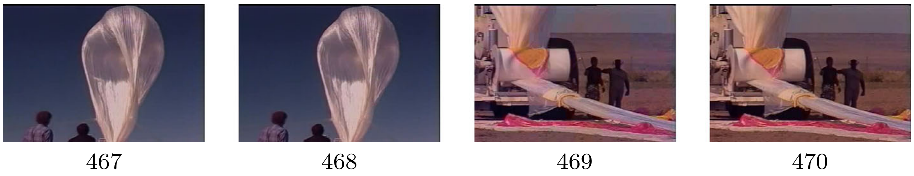


Fig. 3 An example of correctly detected abrupt transition from video

has discarded the illumination change and object motion frames. To show the advantages of the adaptive threshold, comparison between the proposed system using adaptive and hard threshold is done as shown in Table 4. From Table 4 it is clearly shown that the proposed system performs better using the proposed adaptive thresholds. To show the advantage of the confirmation stage, a comparison between the performance of the proposed system with and without using the confirmation stage is done and the results are shown in Table 5.

The reason for choosing videos $D2$ and $D4$ for experimentation is that both videos present sudden illumination and motion effects. Table 5 clearly depicts that the proposed method is very much successful in overcoming most of the challenges mentioned above. Another interesting fact which is revealed when minutely observed is that the $F1$ $score$ of all the videos in Table 5 is enhanced subject to the increase of $precision$ in the proposed system. Figure 4 shows the major challenges such as uniform and non-uniform illumination changes from video $D2$ and $D4$, respectively; these problems are easily handled by our proposed system.

An example from video *D4* is shown in Fig. 5 where a large object (fan) is obstructing the object (dummy airplane) in front of the camera. Figure 5a, b shows the obstacle in multiple consecutive frames and obstacle in a single frame, respectively. The problems in Fig. 5a, b are handled effectively in the confirmation stage of the proposed system. Figure 6 shows the problem of flashlight effect in videos due to which frames are miss-classified as abrupt transition. Our proposed approach has effectively solved this scenario in shot boundary detection.

Our proposed algorithm has correctly detected all the abrupt boundaries present in the *Clip T1* as shown in Fig. 7. The proposed algorithm has successfully discarded most of the illumination and object motion frames, thereby increasing the precision of the system.

4.6 Comparison

For the comparison with the proposed system, some state-of-the-art techniques are considered—WHT-SBD [13], gradient-oriented feature distance (GOFD) [22], stationary wavelet transform (SWT) [6], fast framework [14], PSO-GSA [35], ST-CNN [16], a dual-stage-based approach using LBP-HF and canny edge difference [30], an adaptive low rank and svd-updating approach in [36] and SBD using color histogram and modified multilayer perceptron neural network [37]. Table 6 shows the comparison of the proposed system with the state-of-the-art techniques.

Table 4 System performance using hard thresholds and the proposed adaptive thresholds

Videos	Proposed system					
	With hard threshold $\gamma = 0.2$, $\beta = 0.15$, $\delta = 0.2$			With adaptive thresholds		
	R	P	F1	R	P	F1
*D*2	90.5	97.4	93.8	92.9	100.0	96.3
*D*3	82.1	100.0	90.1	87.2	100.0	93.2
*D*4	80.6	97.5	88.2	86.8	97.8	92.0
*D*6	90.0	100.0	94.7	95.0	100.0	97.4
Average	86.0	99.0	92.0	90.5	99.5	94.7

Table 5 Experimental results without and with confirmation stage

Videos	Proposed system					
	Without confirmation stage			With confirmation stage		
	R	*P*	*F*1	*R*	*P*	*F*1
*D*2	92.9	95.1	94.0	92.9	100.0	96.3
*D*3	87.2	91.9	89.5	87.2	100.0	93.2
*D*4	86.8	94.5	90.5	86.8	97.8	92.0
*D*6	95.0	92.7	93.8	95.0	100.0	97.4
Average	90.5	93.6	92.0	90.5	99.5	94.7

Fig. 4 An example of **a** uniform illumination and **b** non-uniform illumination

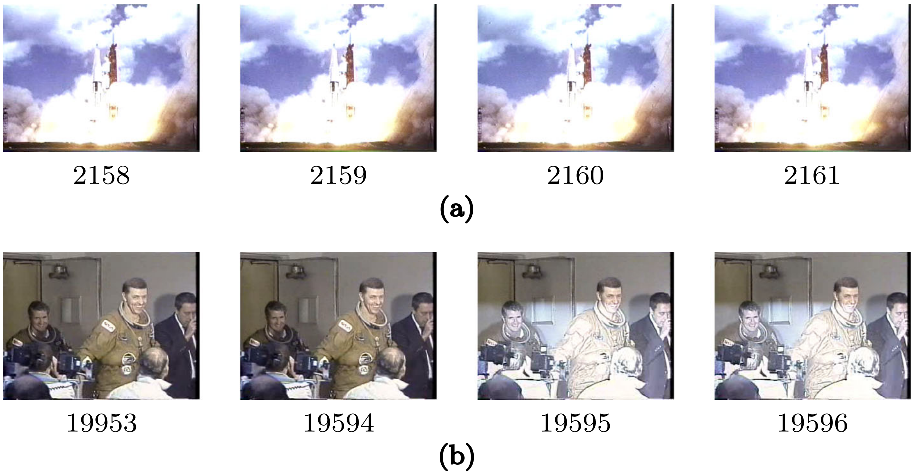
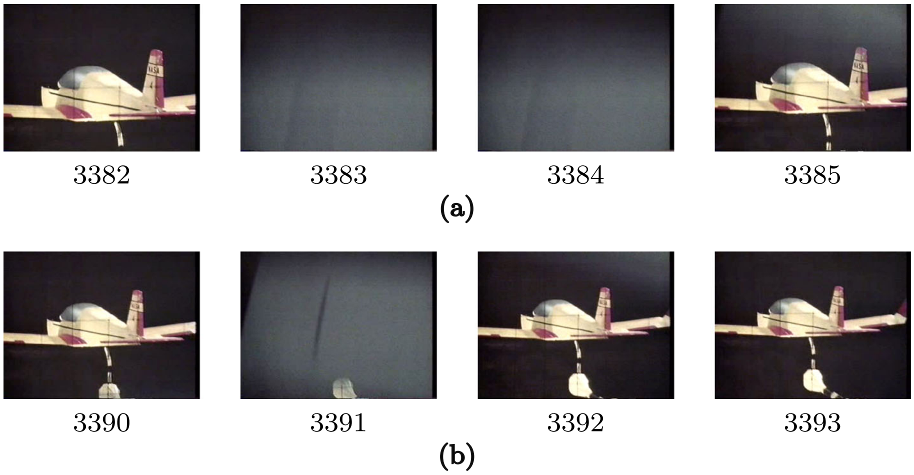


Fig. 5 An example of obstacle in front of camera: **a** multiple frames, **b** single frame



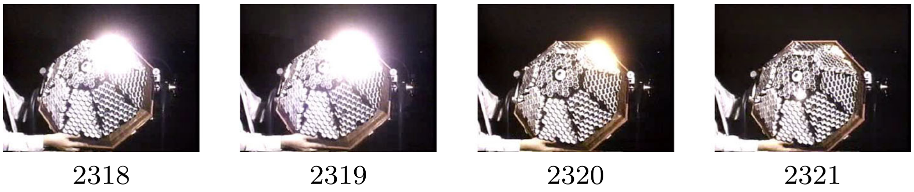


Fig. 6 An example of correctly discarded flashlight effect in detecting abrupt transition

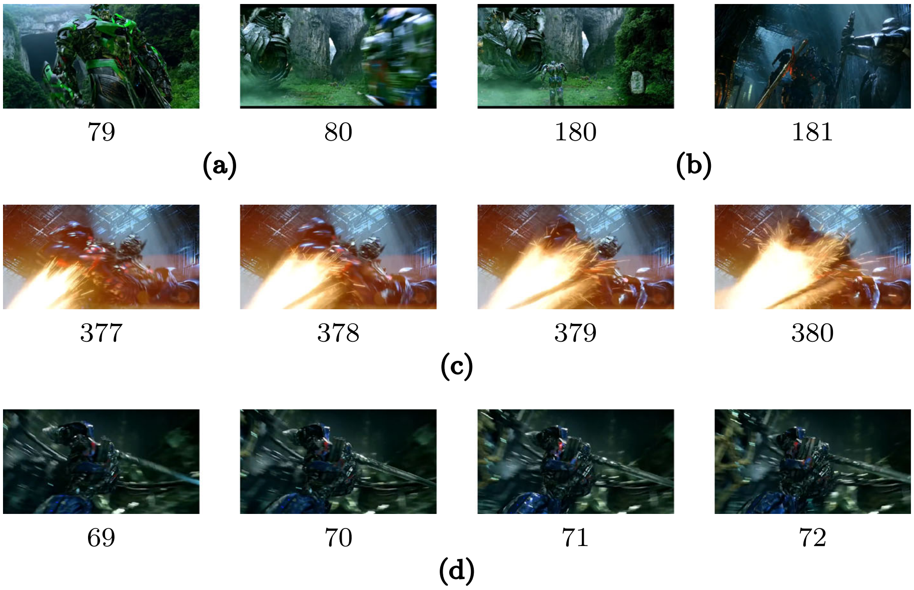


Fig. 7 An example of **a** and **b** correctly detected abrupt transitions, **c** correctly discarded Illumination effect and **d** correctly discarded large Object motion

Table 6 Comparison of the proposed system with state-of-the-art techniques

Algorithm	Evaluation parameter	Videos				Average
		$D2$	$D3$	$D4$	$D6$	
Proposed	R	92.9	87.2	86.8	95.0	90.4
	P	**100.0**	**100.0**	97.8	**100.0**	99.4
	$F1$	**96.3**	93.2	92.0	97.4	94.7
[13]	R	**97.0**	82.0	88.0	95.0	90.5
	$F1$	85.0	86.0	90.0	97.0	90.0
	$F1$	91.0	84.0	89.0	96.0	90.0
[22]	R	80.0	82.0	78.0	92.0	83.0
	P	94.0	**100.0**	96.0	84.0	94.0
	$F1$	87.0	90.0	86.0	88.0	88.0
[6]	R	**97.0**	**97.0**	93.0	**100.0**	**97.0**
	P	6.0	8.0	7.0	8.0	7.0
	$F1$	12.0	16.0	13.0	16.0	14.0
[14]	R	57.0	46.0	75.0	89.0	67.0
	P	**100.0**	**100.0**	98.0	**100.0**	99.6
	$F1$	72.0	63.0	85.0	94.0	79.0
[35]	R	**97.0**	92.0	**100.0**	**100.0**	**97.0**
	P	82.0	**100.0**	89.0	**100.0**	92.0
	$F1$	89.0	**96.0**	**94.0**	**100.0**	94.0
[16]	R	89.0	92.0	85.0	87.0	88.0
	P	87.0	**100.0**	**100.0**	**100.0**	96.0
	$F1$	88.0	**96.0**	92.0	93.0	93.0

Table 6 continued

Algorithm	Evaluation parameter	Videos				Average
		*D*2	*D*3	*D*4	*D*6	
[30]	*R*	90.0	89.0	89.0	92.0	90.0
	P	**100.0**	**100.0**	**94.0**	**97.0**	98.0
	$F1$	95.0	**94.0**	92.0	94.0	94.0
[36]	*R*	92.8	92.3	91.8	100.0	94.2
	P	90.6	**100.0**	**93.7**	**100.0**	96.0
	$F1$	91.6	**95.9**	93.2	100.0	**95.2**
[37]	*R*	88.1	97.4	87.8	97.5	92.7
	P	94.9	88.4	91.5	92.9	92.0
	$F1$	91.4	92.7	89.6	95.1	92.2

Bold value signifiy the best values in the corresponding videos

Table 6 overwhelmingly depicts that the performance of the proposed approach is comparable to other state-of-the-art techniques which is possible due to illumination invariant nature of LTP features and the dual-stage-based approach which is employed for the dual confirmation of the boundaries helps in ensuring comparable precision and $F1$ score. From Table 6, it is observed that [36] has the highest $F1$ score which is possible due to a good recall and in addition decent precision. But the target of the proposed system is to eliminate the illumination and motion effect, that is, to reduce the false positive of the system. Again a good precision of the system ensures that a system is free from false positive, and [6] has the highest precision, but in order to attain a good precision they have compromised their recall which reduces the $F1$ score of the system drastically. However, our proposed system has a comparable recall and precision which yields to a good $F1$ score.

5 Conclusion and future work

A solution has been carved out in the field of shot boundary detection in case of sudden illumination and object motion across videos. This bifold-stage SBD technique clearly depicts the highest precision among all the other state-of-the-art approaches. In the first stage, a possible abrupt transition ($\mathscr{PA}$) is extracted using illumination invariant feature LTP and a set of adaptive thresholds γ and β, respectively. The CIEDE2000 features are extracted from the possible abrupt transition frames in the confirmation section along with an adaptive threshold δ to effectively classify actual abrupt transition frames ($\mathscr{A}$). The advantage of the *Lab* color space is that it can map all the colors perceived by the human visual system, enabling it to differentiate the color features more accurately. The experimental results manifest the illumination and object motion effects effectively handled by the proposed system across videos. All the persisting challenges are effectively addressed such as non-uniform illumination effect and obstruction in front of the camera for multiple frames because a high precision is recorded by the proposed system.

The future work is to improve the recall of the proposed system along with effectively detecting gradual and wipe transition, which will in turn make our system complete to be applied in content-based video retrieval systems.

Compliance with ethical standards

Conflict of interest The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

1. Abdulhussain, S.H., Ramli, A.R., Saripan, M.I., Mahmmod, B.M., Al-Haddad, S.A.R.: Methods and challenges in shot boundary detection: a review. Entropy **20**(4), 214 (2018)
2. Waghmare, M.S.P., Bhide, A.S.: Shot boundary detection using histogram differences. Int. J. Adv. Res. Electron. Commun. Eng. **3**, 1460–1464 (2014)
3. Liu, T., Chan, S.: Automatic shot boundary detection algorithm using structure-aware histogram metric. In: International Conference on Digital Signal Processing, pp. 541–546 (2014)
4. Kaabneh, K., Alia, O., Suleiman, A., Abuirbaleh, A.A.A.A.: Video segmentation via dual shot boundary detection (DSBD). In: 2006 2nd International Conference on Information and Communication Technologies, vol. 1, pp. 1530–1533. IEEE (2006)
5. Warhade, K.K., Merchant, S.N., Desai, U.B.: Avoiding false positive due to flashlights in shot detection using illumination suppression algorithm. In: International Conference on Visual Information Engineering, pp. 377–381 (2008)
6. Warhade, K.K., Merchant, S.N., Desai, U.B.: Shot boundary detection in the presence of fire flicker and explosion using stationary wavelet transform. Signal Image Video Process. **5**(4), 507–515 (2011)
7. Warhade, K.K., Merchant, S.N., Desai, U.B.: Shot boundary detection in the presence of illumination and motion. Signal Image Video Process. **7**(3), 581–592 (2013)

8. Kanungo, P., Kar, T.: Cut detection using block based center symmetric local binary pattern. In: International Conference on Man and Machine Interfacing, pp. 1–5 (2015)
9. Li, J., Yao, T., Ling, Q., Mei, T.: Detecting shot boundary with sparse coding for video summarization. Neurocomputing **266**(C), 66–78 (2017)
10. Pingping, C., Guan, Y., Ding, X., Yu, Z.: Shot boundary detection with sparse presentation. In: 2016 IEEE 13th International Conference on Signal Processing (ICSP), pp. 900–904. IEEE (2016)
11. Huan, Z., Xiuhuan, L., Lilei, Y.: Shot boundary detection based on mutual information and canny edge detector. Int. Conf. Comput. Sci. Softw. Eng. **2**, 1124–1128 (2008)
12. Fu, Q., Zhang, Y., Xu, L., Li, H.: A method of shot-boundary detection based on HSV space. In: Ninth International Conference on Computational Intelligence and Security, pp. 219–223 (2013)
13. Lakshmi Priya, G.G., Domnic, S.: Walsh-Hadamard transform kernel-based feature vector for shot boundary detection. IEEE Trans. Image Process. **23**(12), 5187–5197 (2014)
14. Li, Y., Lu, Z., Niu, X.: Fast video shot boundary detection framework employing pre-processing techniques. IET Image Process. **3**(3), 121–134 (2009)
15. Thounaojam, D.M., Khelchandra, T., Singh, K.M., Roy, S.: A genetic algorithm and fuzzy logic approach for video shot boundary detection. Comput. Intell. Neurosci. **2016**, 14 (2016)
16. Hassanien, A., Elgharib, M.A., Selim, A., Hefeeda, M., Matusik, W.: Large-scale, fast and accurate shot boundary detection through spatio-temporal convolutional neural networks. *CoRR*, arXiv: 1705.03281 (2017)
17. Tong, W., Song, L., Yang, X., Qu, H., Xie, R.: CNN-based shot boundary detection and video annotation. In: IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, pp. 1–5 (2015)
18. Xu, J., Song, L., Xie, R.: Shot boundary detection using convolutional neural networks. In: Visual Communications and Image Processing, pp. 1–4 (2016)
19. Tang, S., Feng, L., Kuang, Z., Chen, Y., Zhang, W.: Fast video shot transition localization with deep structured models. *CoRR*, arXiv:1808.04234 (2018)
20. Srilakshmi, B., Sandeep, R.: Shot boundary detection using structural similarity index. In: Fifth International Conference on Advances in Computing and Communications (ICACC), pp. 439–442 (2015)
21. Chen, L.-H., Hsu, B.-C., Chih-Wen, S.: A supervised learning approach to flashlight detection. Cybern. Syst. **48**(1), 1–12 (2017)
22. Kar, T., Kanungo, P.: A motion and illumination resilient framework for automatic shot boundary detection. Signal Image Video Process. **11**(7), 1237–1244 (2017)
23. Heng, W.J., Ngan, K.N.: The implementation of object-based shot boundary detection using edge tracing and tracking. IEEE Int. Symp. Circuits Syst. VLSI **4**, 439–442 (1999)
24. Heng, W.J., Ngan, K.N.: An object-based shot boundary detection using edge tracing and tracking. J. Vis. Commun. Image Represent. **12**(3), 217–239 (2001)
25. Lan, X., Zhang, S., Yuen, P.C., Chellappa, R.: Learning common and feature-specific patterns: a novel multiple-sparse-representation-based tracker. IEEE Trans. Image Process. **27**(4), 2022–2037 (2018)
26. Lan, X., Ye, M., Shao, R., Zhong, B., Yuen, P.C., Zhou, H.: Learning modality-consistency feature templates: a robust RGB-infrared tracking system. IEEE Trans. Ind. Electron. **66**(12), 9887–9897 (2019)
27. Lan, X., Ye, M., Shao, R., Zhong, B., Jain, D.K., Zhou, H.: Online non-negative multi-modality feature template learning for RGB-assisted infrared tracking. IEEE Access **7**, 67761–67771 (2019)
28. Rashmi, B.S., Nagendraswamy, H.S.: Video shot boundary detection using midrange local binary pattern. In: International Conference on Advances in Computing, Communications and Informatics, pp. 201–206 (2016)
29. Kar, T., Kanungo, P.: A texture based method for scene change detection. In: International Conference on Power, Communication and Information Technology Conference, pp. 72–77 (2015)
30. Singh, A., Thounaojam, D.M., Chakraborty, S.: A novel automatic shot boundary detection algorithm: robust to illumination and motion effect. Signal Image Video Process. **14**, 1–9 (2019)
31. Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Trans. Image Process. **19**(6), 1635–1650 (2010)
32. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
33. Sharma, G., Wencheng, W., Dalal, E.N.: The ciede2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations. Color Res. Appl. **30**, 21–30 (2005)
34. Pele, O., Werman, M.: The quadratic-chi histogram distance family. In: European Conference on Computer Cision, pp. 749–762. Springer (2010)
35. Chakraborty, S., Thounaojam, D.M.: A novel shot boundary detection system using hybrid optimization technique. Appl. Intell. **49**, 1–14 (2019)
36. Youssef, B., Fedwa, E., Driss, A., Ahmed, S.: Shot boundary detection via adaptive low rank and SVD-updating. Comput. Vis. Image Underst. **161**, 20–28 (2017)
37. Thounaojam, D.M., Thongam, K., Jayshree, T., Roy, S., Singh, K.M.: Colour histogram and modified multi-layer perceptron neural network based video shot boundary detection. Int. Arab J. Inf. Technol. **16**, 686–693 (2019)
38. Krishnakumar, K., Gandhi, S.I.: Video stitching based on multi-view spatiotemporal feature points and grid-based matching. Vis. Computer. **36**, 1–10 (2019)

Saptarshi Chakraborty was born in Tripura, India. He did M.Tech from Tripura University, Tripura, India, in 2011 in Computer Science and Engineering. He is pursuing his PhD in Computer Science and Engineering from National Institute of Technology, Silchar. His research interests include image processing, video processing, artificial neural network, and machine learning applications.

Alok Singh is a PhD scholar in the Department of Computer Science and Engineering, National Institute of Technology Silchar, India. He has done M.Tech. in Computer Science and Engineering from National Institute of Technology Silchar, India. His research interests include video processing and machine learning applications.

Dalton Meitei Thounaojam was born in Manipur, India. He did M.E. and PhD from Anna University, Coimbatore, and Assam University, Silchar, India, in 2009 and 2017, respectively, in Computer Science and Engineering and Information Technology. He is an assistant professor in the Department of Computer Science and Engineering, NIT Silchar. His research interests include image processing, video shot boundary detection, fuzzy system and artificial neural network.