Signal, Image and Video Processing
https://doi.org/10.1007/s11760-019-01593-3

ORIGINAL PAPER

A novel automatic shot boundary detection algorithm: robust to illumination and motion effect

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

Received: 21 February 2019 / Revised: 16 October 2019 / Accepted: 23 October 2019

Abstract

Many researches have been done on shot boundary detection, but the performance of shot boundary detection approaches is yet to be addressed for the videos having sudden illumination and object/camera motion effects efficiently. In this paper, a novel dual-stage approach for an abrupt transition detection is proposed which is able to withstand under certain illumination and motion effects. Firstly, an adaptive Wiener filter is applied to the lightness component of the frame to retain some important information on both frequencies and LBP-HF is extracted to reduce the illumination effect. From the experimentation, it is also confirmed that the motion effect is also reduced in the first stage. Secondly, Canny edge difference is used to further remove the illumination and motion effects which are not handled in the first stage. TRECVid 2001 and TRECVid 2007 datasets are applied to analyze and validate our proposed algorithm. Experimental results manifest that the proposed system outperforms the state-of-the-art shot boundary detection techniques.

Keywords LBP-HF · Shot boundary detection · Abrupt · Adaptive threshold

1 Introduction

Nowadays, at the rate at which digitalization is increasing, the sharing of multimedia data (especially video) over the Internet is also increasing at an exponential pace. Due to this excess growth, an effective tool is required for video indexing and retrieval. For designing an effective tool, the contents of the video should be properly recognized and thus the temporal video segmentation is necessary. Temporal video shot boundary detection (SBD) is the segmentation of video into a meaningful shots by spotting the transition between consecutive frames, and a transition marks the boundary between two consecutive shots [1]. A transition can be of two types, namely abrupt and gradual. An abrupt transition is a sudden change in the contents of the video where there is no intersection of frames between the boundaries of two shots. But in gradual transition, the content of the frames changes slowly, both the consecutive shots have intersection of some frames and the duration of these frames is termed as a gradual transition period. The abrupt transition seems very easy to detect but there are lot of challenges in an abrupt transition detection such as sudden illumination and OCM which cause high false positive results.

There are many approaches for an abrupt transition detection, but the histogram-based approach [20] is widely used because of its computational cost and motion-invariant property [1]. In [11], color histogram is used for SBD.

In [27], DCT is used to remove the effect of illumination and then the histogram difference of the transformed frame is evaluated to detect transitions. In [28], an approach using logarithmic transform and DCT is proposed to suppress illumination effect in preprocessing step. In [29], cross-correlation coefficient and stationary wavelet transform (SWT) are used. This method is mainly focused on SBD under fire, flicker and explosion. In [30], dual-tree complex wavelet transform and structural similarity are used to reduce the false positive results due to illumination and OCM effects. In this, an adaptive threshold is used.

In [22], histogram of mid-range LBP features is extracted and a sliding window of 21 frames is used to declare an abrupt transition based on an adaptive threshold. Similarly, in [13], a basic LBP feature histogram is used for SBD. To reduce an illumination effect, block-based center-symmetric LBP (BBCSLBP) is used in [12].

✉ Alok Singh
alok.rawat478@gmail.com

1 Computer Vision Lab, Department of Computer Science and Engineering, National Institute of Technology, Silchar, Assam, India

Published online: 10 November 2019

Edge-based features are also used to reduce an illumination and motion effect [1]. A boundary is declared when the number of edges appearing in the current frame and previous frame exhibits a large change in location. In [6,10], Canny edge detector is used to reduce or remove the motion effect. In [5], the motion strength is calculated between the original frame and compensated frame using a block matching algorithm for reducing motion effects.

A genetic algorithm and fuzzy logic-based approach is proposed for shot detection in [25]. A fast SBD algorithm is proposed in [18] for an abrupt transition detection using a pixel-based approach.

In [23], a simple approach for SBD using SSIM and standard deviation is proposed and to reduce the illumination and motion effects, features like quantized HSV color space [6], histogram intersection [4], absolute sum gradient oriented feature differences [14], etc., are used for SBD.

In [8,9], object tracking method is used for SBD, where a time stamp is attached to each object for locating the number of frames in which the particular object appears. The drawbacks of object-based tracking SBD are sudden disappearance of object from the frame, large object movement which is mistaken as wipe transition and uneven illumination in a video [9]. The uneven illumination effect can be removed by using the object tracking algorithms which are proposed in [15–17].

A CNN-based SBD approach is proposed in [26] which uses an adaptive threshold for selecting candidate segment in preprocessing steps. [7,24,32] also use CNN for SBD.

From the above literature review, illumination and OCM are one of the major challenges in abrupt transition detection. The frames suffering from these challenges are often misunderstood as an abrupt change. Thus, it is a challenging task to propose an approach that is invariant to both of these challenges. In this paper, we proposed a dual-stage-based approach using block-wise LBP-HF and Canny edge detector which is resilient to sudden illumination and OCM effects.

The novel contributions of the paper are:

(i) We proposed a dual-stage-based approach that is resilient under sudden illumination and OCM effects.
(ii) We proposed three adaptive thresholds, where two thresholds are used in initial stage and one in final stage, for detecting an abrupt transition.

The paper is organized as follows: Sect. 2 explains the feature used in the proposed system in detail. In Sect. 3, detailed explanation of the proposed system is given, followed by discussion and conclusion in Sects. 4 and 5, respectively.

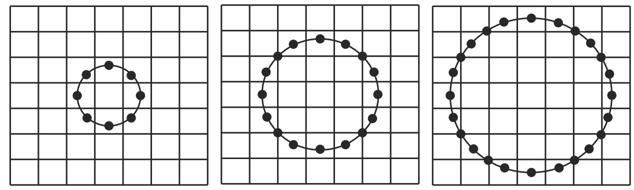

Fig. 1 Three circular neighborhoods: (8,1), (16,2) and (24,3)

2 Background knowledge

In the proposed approach, block-based LBP-HF is used to extract features from the frames in a video. LBP-HF is a texture-based rotation-invariant feature extraction approach in which discrete Fourier transforms (DFTs) of LBP histogram of each frame are taken [2]. LBP provides global rotation- and illumination-invariant feature descriptor [12, 13,21,33].

2.1 Local binary pattern histogram Fourier

Local binary pattern (LBP) is a powerful texture descriptor for an image. The original version of LBP works on a block of 3×3 pixels in an image, where the center pixel is compared with all its neighboring 8 pixels, and according to a threshold, a binary pattern is generated and each value is weighted by a power of two and then all these values are summed.

In the extended LBP, different sizes of neighborhoods can be used. In this paper, a circular neighborhood (P, R) is used as shown in Fig. 1. Here, P is the number of neighbors and R is the radius of neighborhood. The neighbors around the center index (x, y) lie at coordinates $(x_n, y_n) = ((x + R\cos(2\pi n/P)), y - R\sin(2\pi n/P))$; if the coordinates of neighboring pixels are not integer, then they are estimated by interpolation. Now, for the center pixel (x, y) of an image $I(x, y)$, LBP is calculated by

$$\mathrm{LBP}_{P,R}(x, y) = \sum_{n=1}^{P-1} L(I(x, y) - I(x_n, y_n))2^n \tag{1}$$

where $I(x, y)$ is center pixel and $I(x_n, y_n)$ is neighboring pixel and thresholding function $L(s)$ is defined as:

$$L(s) = \begin{cases} 1, & s \geq 0 \\ 0, & s < 0 \end{cases} \tag{2}$$

The DFT-based descriptor is invariant to circular shift in input vector [2]. Therefore, to achieve rotation invariance, we take the DFT of the uniform LBP histogram.

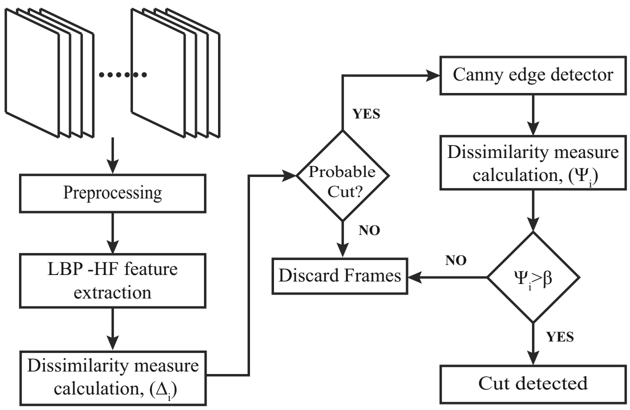

Fig. 2 Block Diagram for proposed system

The DFT of LBP histogram ($h_I(U_P(n,r))$) is denoted by $H(n,.)$ and evaluated by using Eq. 3:

$$H(n,u) = \sum_{r=0}^{P-1} h_I(U_P(n,r))e^{\frac{-i2\pi ur}{P}}. \tag{3}$$

[2] shows that the magnitude of the Fourier spectrum of the histogram rows results into a features that are invariant to input image ($I(x,y)$) rotation.

$$|H(n,u)| = \sqrt{H(n,u)\overline{H(n,u)}} \tag{4}$$

3 Proposed system

This section discusses the proposed approach.

Figure 2 shows the flow of the proposed approach.

3.1 Preprocessing

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

1. Converting the color image to HSV space and taking lightness component of the image.
2. Resizing each frame to $S \times S$ and then dividing it into blocks of size $B \times B$, where $S = 60$ and $B = 10$ in the proposed system. The number of blocks is given by $nb = (S \times S)/(B \times B)$. The block division reduces the effects of illumination and OCM.
3. Applying adaptive Wiener filter to increase the discriminative capability of LBP-HF.

3.2 Feature extraction and dissimilarity measure

After preprocessing, LBP-HF features of each block in a frame are calculated. The reason for using LBP-HF feature is that it is rotation invariant and it is also invariant to small OCM, as it is histogram-based approach. Euclidean distance is used to calculate the dissimilarity differences (δ and ψ) between the corresponding blocks of consecutive frames in a video. In the first stage, the dissimilarity between the consecutive frames is calculated using block-wise LBP-HF feature which is represented as δ. And in the second stage, block-wise canny edge difference is calculated using Euclidean which is represented as ψ.

The final difference value of both stages, Δ and Ψ, are calculated using Eqs. 5 and 6, respectively:

$$\Delta_i = \sum_{j=1}^{n} \delta_j \tag{5}$$

$$\Psi_i = \sum_{j=1}^{n} \psi_j \tag{6}$$

3.3 Thresholding

For determining an abrupt change between the consecutive frames, a threshold is used to declare an abrupt transition if the distance between the consecutive frames is beyond the threshold. It is also important to select a right threshold for ensuring high accuracy in detecting the abrupt changes. As the behavior and contents vary from one video to another, it is very difficult to set a unique threshold (hard threshold) which will work effectively for all kinds of videos. So, a threshold is required which can adapt according to the video characteristics. In the proposed approach, three adaptive thresholds (γ, Γ and β) are proposed for detecting possible and actual abrupt transition, respectively.

The threshold γ is calculated by using Eq. 7, where μ_Δ and σ_Δ are mean and standard deviation of Δ and κ_1 is constant whose appropriate range is [1, 3].

$$\gamma = \mu_\Delta + \sigma_\Delta \times \kappa_1 \tag{7}$$

The threshold Γ is calculated by using Eq. 8:

$$\Gamma = \frac{\gamma}{\sigma_\Delta} \tag{8}$$

Similarly, threshold β is calculated with the help of Eq. 9 where σ_Ψ is the standard deviation of Ψ and κ_2 is a constant whose range is [1, 3].

$$\beta = \sigma_\Psi \times \kappa_2 \tag{9}$$

3.4 Abrupt transition detection

For declaring an abrupt transition, a dual stage is applied where the initial stage is to extract the possible transition frames and the final stage is the confirmation stage to reduce the false detection.

3.4.1 Possible transition detection

After extracting block-based LBP-HF features, dissimilarity measure between the corresponding blocks of consecutive frames is calculated using Eq. 5. From the experimentation, it is observed that when abrupt change is encountered, the value of Δ_i is greater than the threshold γ.

Also the difference between the dissimilarity values of the current frame (i) with the neighboring frames $i-1$ and $i+1$, respectively, is strictly greater than Γ. Using these concepts, the possible transition frames are separated from the nontransition frames (Eq. 10):

$$P_i = \begin{cases} \text{possible tran.,} & (\Delta_i \geq \gamma)\&\&(\Delta_i - \Delta_{i-1} > \Gamma) \\ & \&\&(\Delta_i - \Delta_{i+1} > \Gamma) \\ \text{non-tran.,} & \text{Otherwise} \end{cases} \tag{10}$$

One of the advantages of the possible transition detection is that the nontransition frames detected in this stage are discarded and thus they are not considered in confirmation stage.

3.4.2 Confirmation stage

After getting all probable abrupt transitions by performing the first stage, it is observed from the experimental results that most of the illumination and motion effects are reduced, but there is a possibility that some motion-affected frames are classified as a probable abrupt transition. Thus, it increases the importance of post-processing stage (or confirmation stage) for determining correct transition and reducing false.

In the confirmation stage, block-wise Canny edge difference between the frames $P_i+\eta$ and $P_i-\eta$ is evaluated where η signifies the frames index before and after P_i and η is taken as 4 for experimentation.

After extracting block-wise Canny edge difference, Ψ is calculated for each frame using Eq. 6. From the experimental results, it is observed that if the frames belong to different shots, then the dissimilarity value (Ψ) of the frames $P_i+\eta$ and $P_i-\eta$ is greater than the threshold β. So, Eq. 11 is used for ensuring the confirmation of all probable abrupt transition (P_i).

$$\text{Final\_cut}_i = \begin{cases} \text{True,} & \Psi_i > \beta \\ \text{False,} & \text{Otherwise} \end{cases} \tag{11}$$

The time complexity of the proposed system is $O(nb)$ where n is total number of frames in a video and b is the number of blocks in a frame. A pseudocode of the proposed system is given in Algorithm 1.

Algorithm 1 Pseudocode for the Proposed System.

```
Input: Video, V
Output: Shot Boundaries, Final_cut
procedure SHOT_DETECTION(V)
    F ← VideoReader(V);
    c ← 60; bs ← 10;
    mapping ← getmaplbp(8);                          ▷ provides index for 59 patterns.
    hist₁ ← Blockwise_hist(rgb2hsv(F₁), c, bs, mapping);
    for i = 2 to length(F) do
        hist₂ ← Blockwise_hist(rgb2hsv(Fᵢ), c, bs, mapping);
        for j = 1 to nb do                           ▷ nb is no. of blocks in a frame
            histograms(1, :) ← [his₁(j, :); his₂(j, :)];
            l ← constructhf(histograms, mapping);
            Δᵢ₋₁ ← Δᵢ₋₁ + δⱼ;                         ▷ dissimilarity difference
        his₁ ← his₂;
    for i = 2 to length(Δ) − 1 do
        if Δᵢ ≥ γ&&(Δᵢ − Δᵢ₋₁) > Γ&&(Δᵢ − Δᵢ₊₁) > Γ then
            Pᵢ ← record i^th frame for possible transition;
    for i = 1 to length(Pᵢ) do
        I_p ← imresize(canny(F_{Pᵢ+η}), [c c]);
        I_f ← imresize(canny(F_{Pᵢ−η}), [c c]);
        [B_p, B_f] ← block(I_p, I_f, [bs bs]);
        Ψᵢ ← sum(ψⱼ);
    Ψ ← Ψ/max(Ψ);
    for i = 1 to length(Ψ) do
        if Ψᵢ > β then
            record Final_cutᵢ;
Function Blockwise_hist(frame, c, bs, mapping)
im ← imresize(frame(:, :, 3), [c c]);
IB ← block(im, [bs bs]);
for i = 1 to nb do
    A1 ← wiener2(IBᵢ);
    L ← lbp(A1, 1, 8, mapping, 'h');
    histogramsᵢ ← L/sum(L);
return histograms
```

4 Experimental results and discussion

4.1 Dataset

To analyze the effectiveness of the proposed approach and to prove the advantages of the proposed system over the state-of-the-art techniques, TRECVid 2001 and TRECVid 2007 datasets, two small clips of movie "Transformer" and a song of movie "Masoom" are used. The proposed system is experimented using HP-Z220 workstation. Details of the selected videos are given in Table 1.

4.2 Evaluation parameters

The performance of the proposed approach is analyzed by using recall (R), precision (P) and $F1$ score $(F1)$ which are evaluated by using Eqs. 12, 13 and 14, respectively.

$$R = \frac{\text{Correctly Detected}}{\text{Correctly Detected} + \text{Miss Detected}} \tag{12}$$

$$P = \frac{\text{Correctly Detected}}{\text{Correctly Detected} + \text{Wrongly Detected}} \tag{13}$$

$$F1 = \frac{2 \times R \times P}{R + P} \tag{14}$$

Table 1 Ground truth data of the test videos

Video	Frames	Transition			Sources
		Abrupt	Gradual	Total	
$D2$	16,586	42	31	73	TRECVid 2001
$D3$	12,304	39	64	103	
$D4$	31,389	98	55	153	
$D5$	12,508	45	26	71	
$D6$	13,648	40	45	85	
$anni001$	914	–	8	8	
$anni001$	2492	12	–	12	
BG_3027	49,815	127	1	128	TRECVid 2007
BG_3097	44,991	91	–	91	
BG_3314	35,802	44	1	128	
BG_16336	2466	127	1	128	
BG_28476	23,238	176	3	179	
BG_36136	29,426	88	21	109	
BG_37309	9639	11	10	21	
BG_37770	15,836	8	29	37	
$ClipV1$	1183	20	–	20	Movie
$ClipV2$	924	19	–	19	Transformer
Masoom	9193	41	–	41	Movie song

Table 2 Adaptive thresholds for different videos

Video	Possible stage		Confirmation stage
	(γ)	(Γ)	(β)
$D2$	8.0293	4.1419	0.3004
$D3$	7.6483	4.4991	0.3619
$D4$	6.7610	4.2594	0.2304
$D6$	5.7123	3.9680	0.1865

Table 3 System performance using hard thresholds and the proposed adaptive thresholds

Videos	Proposed system					
	With hard thresholds $\gamma = 8, \Gamma = 4, \beta = 0.3$			With adaptive thresholds		
	R	P	$F1$	R	P	$F1$
$D4$	0.88	0.95	0.91	0.89	0.94	0.92
$D5$	1.00	0.93	0.96	1.00	0.95	0.97
BG_3027	0.96	0.97	0.96	0.95	0.95	0.95
BG_3314	0.86	0.80	0.83	0.84	0.86	0.85
Average	0.92	0.91	0.91	0.92	0.92	0.92

4.3 Parameters selection

The performance of the system totally depends on the proper selection of the parameters used in the proposed approach. In the proposed approach, we have used three adaptive thresholds γ, Γ and β which are discussed in Sect. 3.3. The thresholds γ and Γ are used for extracting probable abrupt changes, and β 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.

From the experimentation, it is observed that the appropriate range of both constants κ_1 and κ_2 used in Eqs. 7 and 9 is [1 3]. For the experimentation, we set the value of κ_1 and κ_2 as 1.9 and 2.5, respectively.

Table 3 shows the performance of the proposed system using the proposed adaptive thresholds and experimental thresholds (hard thresholds). Two videos from each of the datasets, TRECVid 2001 and TRECVid 2007, are selected for comparison.

4.4 Discussion

In this section, we will discuss the importance of the adaptive Wiener filter and Canny edge detector in the proposed system.

4.4.1 Wiener filtering for increasing discriminative power of LBP-HF

Illumination falls in the low-frequency part of a signal. So, to remove the illumination effect, using a high-pass filtering is not a wise idea as some relevant features of the signal may

Table 4 Experimental results without adaptive Wiener filter and with adaptive Wiener filter

Videos	Proposed system					
	Without adaptive Wiener Filter			With adaptive Wiener Filter		
	R	P	$F1$	R	P	$F1$
$D2$	0.64	0.97	0.77	0.90	0.97	0.93
$D3$	0.46	1.00	0.63	0.89	1.00	0.94
$D4$	0.36	0.92	0.52	0.89	0.94	0.92
$D6$	0.62	0.80	0.70	0.92	0.97	0.94
Average	0.52	0.92	0.65	0.90	0.97	0.93

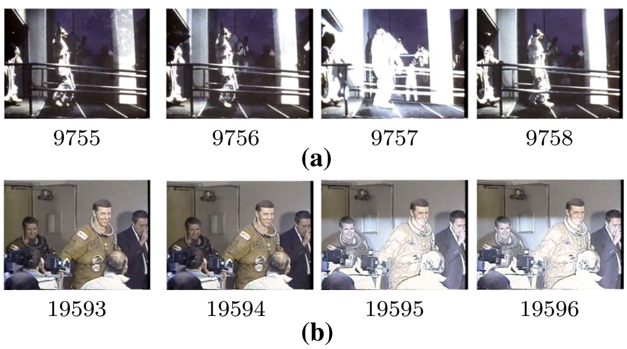

Fig. 3 Flash effect : **a** uniformly distributed and **b** nonuniformly distributed

also fall in the low frequency [31]. To retain information in both frequencies, an adaptive Wiener filter [19] is used in the proposed system.

To show the importance of the adaptive Wiener filter, an experimentation is carried out without the use of confirmation stage. Table 4 shows the performance of the single-stage system with and without adaptive Wiener filter.

From Table 4, it is observed that the use of adaptive Wiener filter significantly improves the *recall* and $F1$ score of the proposed approach.

From the experimentation, it is also observed that if the illumination effect is distributed throughout the frame, the proposed system can handle it. An example of this type of situation is given in Fig. 3a (from $D2$). Figure 3b (from $D4$) shows the partial illumination in multiple consecutive frames where the proposed system is not able to handle it and thus a possible abrupt transition is declared between frames 19594 and 19595.

Further, in Fig. 4, small clips of the movie "Transformer" are shown which have both illumination and OCM effects where the proposed system can handle the changes.

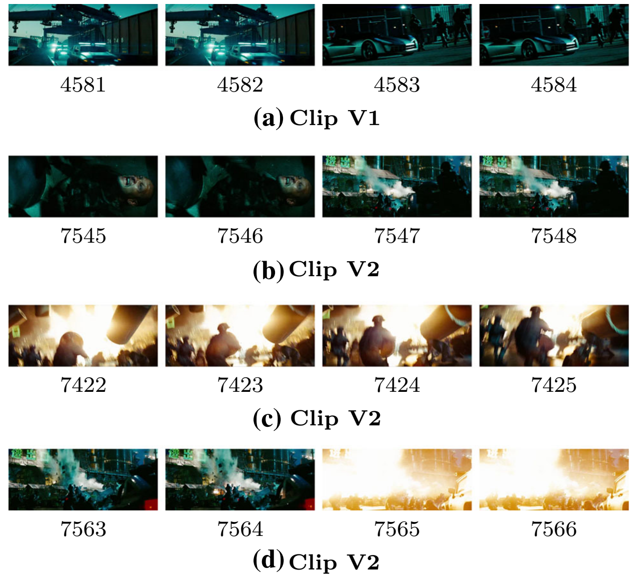

Fig. 4 Video clips from the movie "Transformer" affected by sudden illumination and OCM effect. **a**, **b** Correctly detected boundaries where motion effect is high. **c**, **d** Illumination effects successfully handled by our proposed system

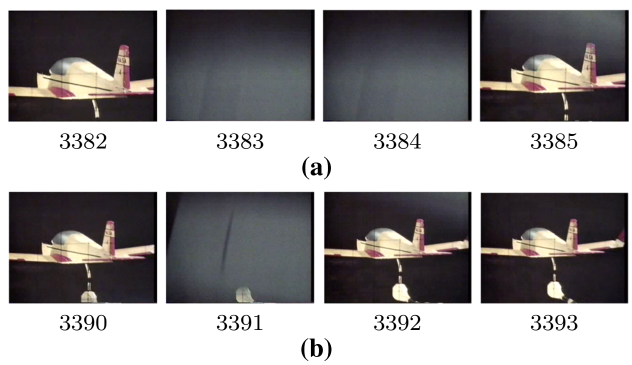

Fig. 5 Obstacle in front of camera **a** in multiple frames and **b** in single frame

4.4.2 Increasing robustness of the system with confirmation stage

From Table 4, we have observed that the adaptive Wiener filter is able to increase the discriminative property of LBP-HF effectively by retaining the important features of both frequencies but still there are some problems due to large object motion. An example from video $D4$ is shown in Fig. 5 where an object (fan) is obstructing the object (dummy airplane) in front of the camera. Figure 5a shows the obstacle in multiple consecutive frames, and Fig. 5b shows the obstacle in a single frame. In confirmation stage, the problem in Fig. 5b is handled, whereas the problem in Fig. 5a still persists due to the object effect in multiple frames, i.e., large object effect for long duration.

Table 5 Experimental results without and with confirmation stage

Videos	Proposed system					
	Without confirmation stage			With confirmation stage		
	R	P	$F1$	R	P	$F1$
$D2$	0.90	0.97	0.93	0.90	1.00	0.95
$D3$	0.89	1.00	0.94	0.89	1.00	0.94
$D4$	0.89	0.94	0.92	0.89	0.94	0.92
$D6$	0.92	0.97	0.94	0.92	0.97	0.94
Average	0.90	0.97	0.93	0.90	0.98	0.94

Table 5 shows the performance of the proposed system with and without confirmation stage.

The videos $D2$ and $D4$ have most of the illumination and motion effects. But from Table 5, it is clearly seen that the proposed system is very effective in dealing with most of the illumination and motion effects. In fact, when Table 5 is minutely observed, the $F1$ score of all the sample videos increases due to the increase in the $precision$ of the proposed system. The problem in Fig. 5b can be handled by the proposed system by varying the value of η but it will impact on the performance of the system.

Table 6 Performance of the system for TRECVid 2001 and TRECVid 2007

Videos	Parameter measure			Computation time in sec. (approx.)	Frame rate
	R	P	$F1$		
$D2$	0.90	1.00	0.95	820	30
$D3$	0.89	1.00	0.94	630	30
$D4$	0.89	0.94	0.92	1542	30
$D5$	1.00	0.95	0.97	678	30
$D6$	0.92	0.97	0.94	670	30
$anni002$	0.92	1.00	0.96	45	30
BG_3027	0.95	0.95	0.95	2999	25
BG_3097	0.91	0.95	0.93	2680	25
BG_3314	0.84	0.86	0.85	2136	25
BG_16336	0.90	1.00	0.94	158	25
BG_28476	0.92	0.85	0.89	1373	25
BG_36136	1.00	0.98	0.99	1826	25
BG_37309	1.00	0.84	0.91	603	25
BG_37770	1.00	1.00	1.00	888	25
$Clip1$	0.80	1.00	0.95	65	24
$Clip2$	0.68	1.00	0.81	58	24
Massom	0.97	1.00	0.98	540	25
Average	0.91	0.95	0.93	1041	26

4.5 System performance

The overall performance of the proposed system is given in Table 6.

For TRECVid 2001 dataset, the average recall, $precision$, $F1$ score and computation time of the proposed system are

Table 7 Comparison of the proposed system with the state-of-the-art techniques

Algorithm	Evaluation parameter	Videos				Average
		$D2$	$D3$	$D4$	$D6$	
Proposed	R	0.90	0.89	0.89	0.92	0.90
	P	**1.00**	**1.00**	0.94	0.97	0.98
	$F1$	**0.95**	0.94	0.92	0.94	**0.94**
[5]	R	**0.97**	0.82	0.88	0.95	0.91
	P	0.85	0.86	0.90	0.97	0.90
	$F1$	0.91	0.84	0.89	0.96	0.91
[14]	R	0.80	0.82	0.78	0.92	0.83
	P	0.94	**1.00**	0.96	0.84	0.93
	$F1$	0.87	0.90	0.86	0.88	0.88
[29]	R	**0.97**	**0.97**	0.93	**1.00**	**0.97**
	P	0.06	0.08	0.07	0.08	0.07
	$F1$	0.12	0.16	0.13	0.16	0.14
[18]	R	0.57	0.46	0.75	0.89	0.67
	P	**1.00**	**1.00**	0.98	**1.00**	**0.99**
	$F1$	0.72	0.63	0.85	0.94	0.79
[12]	R	0.78	0.69	0.41	0.85	0.68
	P	0.78	0.72	0.34	0.87	0.68
	$F1$	0.78	0.71	0.37	0.86	0.68
[3]	R	**0.97**	0.92	**1.00**	**1.00**	**0.97**
	P	0.82	**1.00**	0.89	**1.00**	0.92
	$F1$	0.89	**0.96**	**0.94**	**1.00**	**0.94**
[7]	R	0.89	0.92	0.85	0.87	0.88
	P	0.87	**1.00**	**1.00**	**1.00**	0.96
	$F1$	0.88	**0.96**	0.92	0.93	0.93

Bold values indicate that the corresponding value is highest among all compared approaches

0.92, 0.97, 0.94 and 723 s, respectively. For TRECVid 2007 dataset the average recall, precision, $F1$ score and computation time of the proposed system are 0.94, 0.93, 0.93 and 1582.9 s, respectively. Similarly, the overall performance of the proposed system is 0.91, 0.95, 0.93 and 1041 s.

4.6 Comparison

To show the superiority, the proposed system is compared with the state-of-the-art techniques such as WHT-SBD [5], gradient-oriented feature distance (GOFD) [14], stationary wavelet transform (SWT) [29], fast framework [18], block-based center-symmetric LBP (BBCSLBP) [12], PSO-GSA [3] and ST-CNN [7] as shown in Table 7.

From Table 7, it is observed that the $F1$ score of all the selected videos from TRECVid 2001 dataset is better than the other techniques which shows that the proposed system outperforms the other techniques.

5 Conclusion

The illumination and motion effects in a video affect the performance of an abrupt transition detection process due to the misinterpretation of these effects as an abrupt transition. In this paper, we proposed a dual-stage-based approach for an effective illumination and motion-resilient abrupt SBD. The first stage uses adaptive Wiener filter and LBP-HF to eliminate the illumination and some motion effects, and the second stage is used for the verification of all the probable abrupt changes and the declaration of actual abrupt transition. The proposed system also proposes a novel three adaptive thresholds for the detection of probable and actual abrupt changes where the effectiveness of these thresholds is shown in the experimental section.

In the future work, the proposed system can be improved by eliminating the false positive due to the partial illumination effect.

References

1. Abdulhussain, S.H., Ramli, A.R., Saripan, M.I., Mahmmod, B.M., Al-Haddad, S.A.R., Jassim, W.A.: Methods and challenges in shot boundary detection: a review. Entropy **20**(4), 214 (2018)
2. Ahonen, T., Matas, J., He, C., Pietikäinen, M.: Rotation invariant image description with local binary pattern histogram fourier features. In: Salberg, A.B., Hardeberg, J.Y., Jenssen, R. (eds.) Image Analysis, pp. 61–70. Springer, Berlin (2009)
3. Chakraborty, S., Thounaojam, D.M.: A novel shot boundary detection system using hybrid optimization technique. Appl. Intell. (2019). https://doi.org/10.1007/s10489-019-01444-1
4. Chen, L.H., Hsu, B.C., Su, C.W.: A supervised learning approach to flashlight detection. Cybern. Syst. **48**(1), 1–12 (2017)
5. Domnic, S.: Walsh–Hadamard transform kernel-based feature vector for shot boundary detection. IEEE Trans. Image Process. **23**(12), 5187–5197 (2014)
6. 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)
7. 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)
8. 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)
9. 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)
10. 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)
11. Kaabneh, K., Alia, O., Suleiman, A., Abuirbaleh, A.: Video segmentation via dual shot boundary detection (DSBD). In: International Conference on Information and Communication Technologies, vol. 1. IEEE, pp. 1530–1533 (2006)
12. 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)
13. 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)
14. Kar, T., Kanungo, P.: A motion and illumination resilient framework for automatic shot boundary detection. Signal Image Video Process. **11**(7), 1237–1244 (2017)
15. 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)
16. 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)
17. 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. (2019). https://doi.org/10.1109/TIE.2019.2898618
18. Li, Y., Lu, Z., Niu, X.: Fast video shot boundary detection framework employing pre-processing techniques. IET Image Process. **3**(3), 121–134 (2009)
19. Lim, J.S.: Two-Dimensional Signal and Image Processing. Prentice Hall, Englewood Cliffs, NJ (1990)
20. 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)
21. 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)
22. 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)
23. 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)
24. Tang, S., Feng, L., Kuang, Z., Chen, Y., Zhang, W.: Fast video shot transition localization with deep structured models. CoRR arXiv:1808.04234 (2018)

25. 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)
26. 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)
27. Waghmare, M.S.P., Bhide, A.: Shot boundary detection using histogram differences. Int. J. Adv. Res. Electron. Commun. Eng. **3**, 1460–1464 (2014)
28. 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)
29. 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)
30. 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)
31. Xie, X., Zheng, W.S., Lai, J., Yuen, P.C.: Face illumination normalization on large and small scale features. In: Conference on Computer Vision and Pattern Recognition, pp. 1–8. IEEE (2008)
32. Xu, J., Song, L., Xie, R.: Shot boundary detection using convolutional neural networks. In: Visual Communications and Image Processing, pp. 1–4 (2016)
33. Zou, X., Kittler, J., Messer, K.: Illumination invariant face recognition: a survey. In: First IEEE International Conference on Biometrics: Theory, Applications, and Systems, pp. 1–8 (2007)