A Proposed Algorithm for Shot Boundary Detection Using Various Features

Matheel Emadaldeen[1] and Eman Hato*[2]

[1] Computer Science Department – University of Technology
[2] Computer Science Department – University of Mustansiriyah

*Emanhato@yahoo.com

Received: 26 July 2018 Accepted: 18 December 2018

Abstract

The Shot is often used as a basic meaningful unit for analysing, summarization and indexing video. In this paper, an effective algorithm is proposed to detect abrupt and gradual shot boundary automatically. The proposed algorithm consists of three main phases: fade in/out detector, abrupt detector, and gradual detector. Fade in/out detector is handled in the first step of the proposed algorithm using Maximally Stable Extremal Regions (MSER) descriptors to avoid misdiagnosis caused by black frame. Second, abrupt transitions are detected by analysis of gist features similarity and adaptive local threshold. Lastly, gradual transitions are found out between neighbouring abrupt using gist features as introductory detection of the gradual shot while the color histogram features are employed for finishing detection to reduce the false detection. Experimental results indicate that the proposed algorithm provides 0.981 F-measure, 0.977 precision and 0.986 recall rates for selected testing videos. The performance of the proposed algorithm is satisfactory for detecting different models of transitions under motion effects and camera operations.

Keywords: Abrupt detector, Color histogram, Gist features, Gradual detector, MSER.

خوارزمية مقترحة للكشف عن حدود مشهد الفيديو بإستخدام ميزات مختلفة

مثيل عماد الدين[1] و إيمان هاتو[2]

[1] قسم علوم الحاسوب – الجامعة التكنولوجية

[2] قسم علوم الحاسوب – الجامعة المستنصرية

ألخلاصة

يعتبر المشهد الوحدة الأساسية للفيديو والتي تسخدم في تحليل وتلخيص وفهرسة الفيديو. تم في هذا البحث اقتراح خوارزمية فعالة للكشف عن حدود مشاهد الفيديو التي تتغير بشكل مفاجئ وتدريجي تلقائيا. تتكون الخوارزمية المقترحة من ثلاث مراحل رئيسية: الكشف عن التلاشي التدريجي والكشف عن التغيير المفاجئ والكشف عن التداخل التدريجي. المرحلة الأولى من الخوارزمية المقترحة يتم فيها الكشف عن التلاشي التدريجي بإستخدام ميزات. المرحلة الثانية من الخوارزمية يتم الكشف عن حدود المشاهد التي تتغير بشكل مفاجئ باستخدام ميزات (Gist) وقيمة العتبة التي تحدد وفق معادلة تتغيير وفق محتويات الفيديو. أما المرحلة الثالثة والاخيرة من الخوارزمية فيتم فيها الكشف عن التداخل التدرجي في المشاهد التي تكون بين مشاهد التغيير المفاجئ، وتستخدم ميزات (Gist) كتحديد أولي وبعدها يتم استخدام (Color Histogram) لإنهاء عملية الكشف الاولي وتقليل التشخيص الخاطئ. أظهرت النتائج التجربية أن الخوارزمية المقترحة توفر دقة عالية في تحديد وكشف حدود مشاهد الفيديو المختلفة (F = 0.981, precision = 0.977 and recall = 0.986) ومتجاوزة تأثير حركة الكائن او الكاميرا وبهذا فانها تقدم أداء جيد ومرضي .

الكلمات المفتاحية: الكاشف المفاجئ، المدرج التكراري، ميزات جيست، الكاشف التدريجي، المناطق المستقرة.

Introduction

Segmenting the video is the first step in an automatic video indexing process. The aim of video segmentation is to divide the video into essential and meaningful parts called shots in order to detect and locate the desired information in a large amount of video data. Each video shot represents a continuous series of actions, which corresponds to a series of frames captured by a single camera without significant changes in visual content [1]. The basis of any shot detection techniques is to find the boundary between the shots, where the boundary is the discontinuity frame(s) that describes the transition from one shot to the subsequent shot. However, it is quite difficult to detect a video shot transition due to many reasons such that object motions, camera movements and noise addition which often may change the video content dramatically and cause a false detection on shot transition [2].

Shot boundary falls into two classes: Abrupt Transition (AT) and the Gradual Transition (GT). The shots are separated by a single frame called AT, while the shots are separated by several frames having highly interrelated visual information called GT [3, 4]. A GT is obtained using some effects, for example, fade in where a new shot appears gradually with an increase in the brightness from a black color frame, fade out is reverse of a fade in, and dissolve where two consecutive shots overlap [5]. It is easy to detect abrupt transitions because the two successive frames involved in the transition are totally uncorrelated. In contrast, the GT is difficult to detect because the similarity between successive frames is slightly reduced.

Moreover, object motion and camera operation (zoom, rotate, tilt) display features similar to transition effects such as dissolves. A reliable shot boundary detection algorithm must be able to recognize the effect of dissolve without misinterpretation of camera motion as GT.

The objective of the proposed method is to provide a new algorithm to detect abrupt and gradual transitions in the video stream that are robust for camera operation and object movement. The remainder of this paper is structured as follows: in the next section, literature reviews are discussed. In section 3 the description of features extraction techniques is presented. The proposed shot boundary detection algorithm is presented in section 4. Section 5 illustrates the experimental results obtained. Lastly, the conclusions are contained in section 6.

Literature Review

The video shot detection techniques have become one of the most important research areas in content-based video analysis and retrieval. Researchers have attempted to detect the shot boundaries using video content features by focusing on visual interruptions between frames. In abrupt detection techniques (hard cuts), the interruptions can be detected by searching for two consecutive frames that are totally dissimilar. Various useful features extracted from the video frames are used to distinguish between transition and non-transition frames by applying similarity measures that are then compared to a threshold that can be local, global or adaptive threshold [6]. These techniques are simple but choosing a threshold is critical. The most commonly used feature for these techniques is histogram which is calculated using a gray level or color space such as RGB, YUV, HSV, CMYK and YCbCr [7,8]. In 2015 [9], Hong Shao et al. exploit Saturation Value (HSV) color histogram and Histogram of Gradient (HOG) features to detect AT. HSV color histogram is used for introductory detection of video shot while HOG feature is adopted for secondary detection in order to improve the algorithm performance. Chi-square measure is used to calculate the difference of adjacent two frames which is then compared to the global adaptive threshold.

Other useful features extracted from video frames are also employed in shot boundary detection such that pixel level difference of consecutive frames [10], motion-based features [11], texture [12], edge and gradient-based features [13,14]. For example, T. Kar and P. Kanungo in 2017 [15] are employed absolute sum gradient orientation feature difference which is compared to threshold generated from the local and global threshold for detecting AT. Their proposed method is robust to the object /camera motion and illumination change due to invariable features used in the method. In 2016 [16] Dalton Meitei Thounaojam et al. used the features extracted by Gray Level Co-occurrences Matrix (GLCM) and correlation measure to calculate the difference between two consecutive GLCM frames of the video. Their proposed algorithm used a local threshold to detect AT in an uncompressed domain.

Sliding windows is a common technique used to detect GT (dissolves [17], fades [18], and wipes [19]) in video sequences. The success of this technique depends on the appropriate selection of window size. According to the recent literature, a large number of techniques have been proposed to recognize both abrupt and gradual transition [20].

A shot detection method presented by Rachida Hannane et al. in 2016 [21] combines local and global features by using a distance of SIFT point distribution histogram of consecutive frames and an adaptive threshold for detection the cut and gradual transition. Ali Amiri and Mahmood Fathy in 2009 [22] are presented algorithm utilizing the properties of QR decomposition and Gaussian model. The algorithm finds the candidates gradual transition using three-dimensional color histograms and properties of QR decomposition. Then candidate's gradual transitions are filtered to find the correct transition by model each candidate's transition using a Gaussian model. The detection of shot boundaries based on machine learning techniques is also common such as fuzzy clustering [23], decision tree-based classification [24], neural network [25, 26], support vector machine [27, 28]. For example, Jaydeb Mondal et al. in 2017 [28] presented a Least Squares Support Vector Machine (LS-SVM) classifier as a modified version of SVM to reduce the computational cost. Feature vectors are extracted by the multiscale geometric analysis of non-subsampled contourlet transform and reduced by principal component analysis. LS-SVM is classified the frames sequence based on the feature vectors into No-Transition, AT and GT classes. The performance of their proposed algorithm is efficient because robust feature vector produced. In 2014 [29] Lakshmi Priya G. and Domnic S. extract feature vectors from multiple features (color, texture, edge and motion) using Walsh Hadamard transform and then Procedure Based shot transition Identification process (PBI) as classifier the shot into abrupt and gradual transitions.

After reviewing the shot boundary detection algorithms, it concludes that the main challenges to detect correct shot boundary are the flashlight events, fast object motion, large camera movement and the presence of special effects

Feature Extraction

During the detection process, it is essential to extract visual features such as the features based on motion, color, shape and texture or their combination that measures the similarity between frames. A brief description for types of the feature extraction techniques that are used in the proposed algorithm is introduced in the following section.

1. Gist Feature

The global Gist descriptor is a low dimensional representation of the scene, which does not require any form of segmentation. This representation refers to the Gist of a scene, which includes all levels of visual information ranging from low-level features (contours, color, spatial frequencies) to intermediate features (texture, shapes) and high-level features (objects, activation of semantic knowledge). Thus, Gist can be represented as Perceptual Gist (structural representation of a scene) and Conceptual Gist (semantic information that is deduced while viewing a scene) [30]. The Gist descriptor is scaled invariable feature and shows good results for scene recognition and searching.

2. Maximally Stable Extremal Region (MSER)

Maximally Stable Extremal Regions (MSER) is a local feature detector that extracts a number of co-variant regions from a Grayscale image called MSERs. The term Extreme Region (ER) refer to sets of connected pixels have intensity values higher (bright extremely regions) or lower (dark extremely regions) than their outer boundaries. An extremism region is maximally stable when it shows invariance to the transformation of image intensities over a large range of intensity threshold levels. MSER is very efficient for multi-scale detection, low computational complexity and it widely used for finding the correspondence between the two images [31].

3. Color Histogram

The histogram represents the color distribution of digital images; it gives the number of pixels that contain colors within the same value range. Histogram techniques are effective and robust to camera effects, object motion and changes in scale and rotation within a frame sequence. The limitation of color histogram is described the relative quantity of each color in the image while ignoring the spatial relationships between different colors, thus ineffective to the applications in which texture or shape is important [32].

Proposed Video Shot Detection

The structure of the proposed algorithm is composed of three main phases: Abrupt detector, Gradual detector, Fade In/Out detector as shown in Figure1. First, the frame images are extracted from the input video and resize each image into 256 × 265. Next, the extracted frame images are converted to a Grayscale image to be the input to the detector of shot boundaries. To avoid the false detection caused by a black frame and ensure the removal of unwanted information from the video, the black frame is necessary to discard. But for the Fade In/Out transition, the blank frames play an essential role in detecting the fade effect which is part of the gradual transition, therefore Fade In/Out detector are handled in the onset of the proposed algorithm separately from the gradual detector. Whenever the fade transition is detected, the remaining processing is not performed for the abrupt and gradual transition. Abrupt transitions are detected before processing for the gradual transitions detection that is done between neighboring cuts to reduce missed detections. A detailed explanation of all these phases is provided in the following sub-sections:

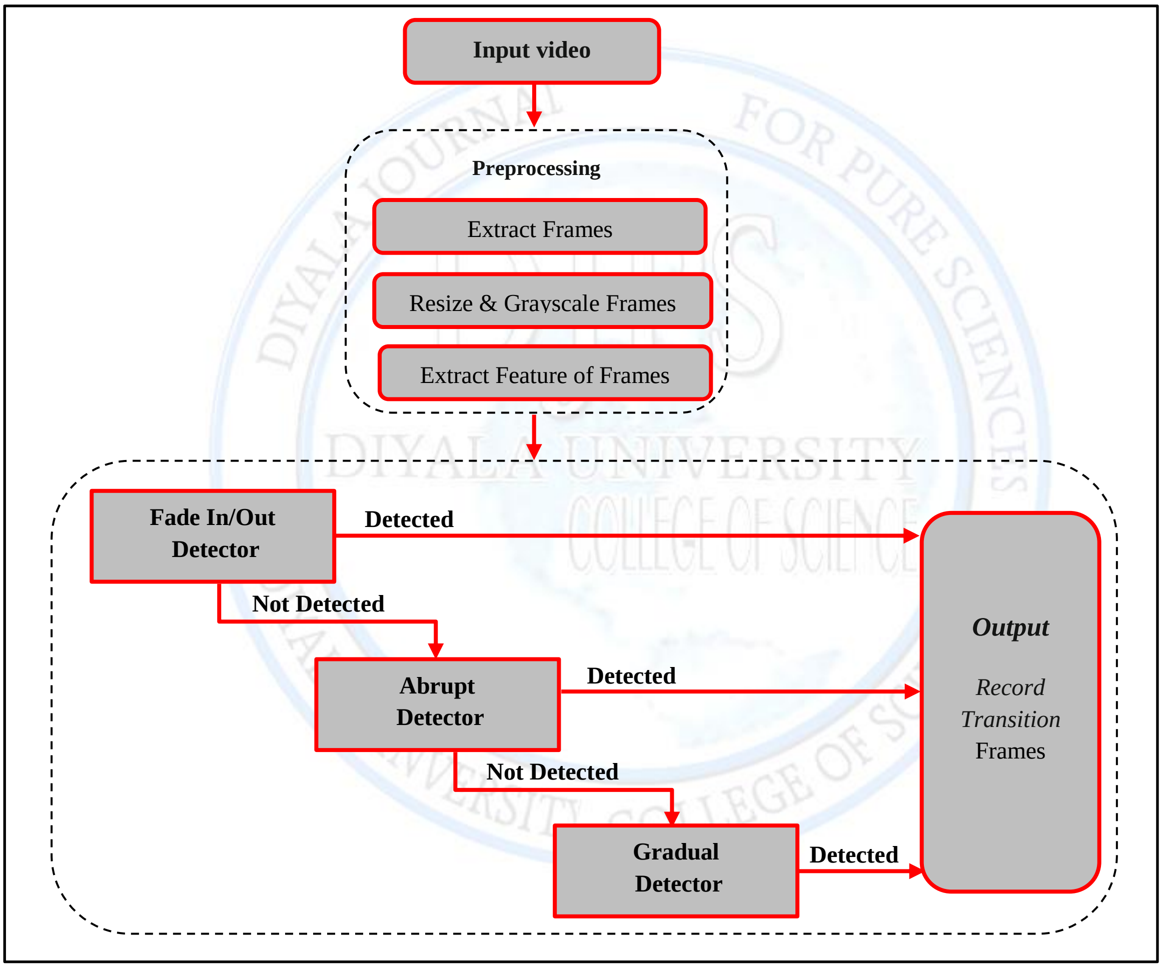


Figure 1 :Structure of proposed algorithm

1. Fade In / Out detector

A Fade/In is a transition where the brightness of the frame gradually increases from the black frame. On the contrary, a Fade/Out is a transition where the brightness gradually decreases towards a black frame. The proposed method detects fades by focusing on this black frame by using MSER regions extracted from a frame. It should be noted that the black frame has zero MSER regions. A large number of video files are analyzed using the MSER in order to detect the pattern and behavior of Fade In/Out transition. Graphical visualization of the example of video sequences with fade transitions and the normal scene is illustrated in Figure2 according to the number of MSER regions. One can observe that MSER varies in an increase from zero or a decrease toward zero for fade transition, unlike the normal scene variation. Therefore, proposed Fade In/Out detector search for increasing or decreasing pattern that starts or ends in the black frame respectively (i.e. MSER regions equal to zero) in any video to diagnose Fade In/Out transition.

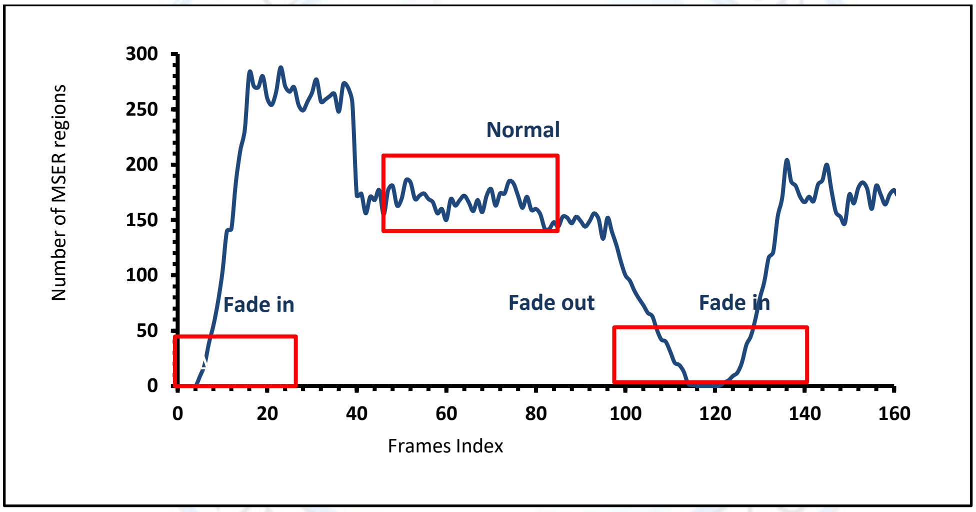
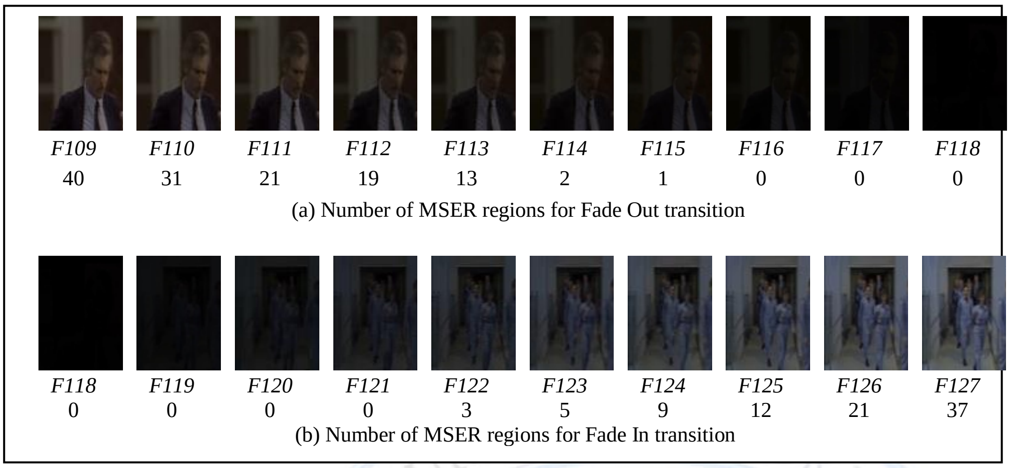


Figure 2: Multiple Fade In/Out transitions for the track of CNN1 video

To avoid the sequence of black frames or sequence of dark frames where MSER regions closed to zero that may be found in the video and caused in false detection, it is essential to check transition frames in order to determine if there is a black frame or not as a post-processing steps for proposed Fade In/Out detector. An example of Fade In/Out detection is illustrated in Figure 3 and Algorithm 1 shows the implemented steps to Fade In/Out detector.



(a) Number of MSER regions for Fade Out transition

(b) Number of MSER regions for Fade In transition

Figure 3: Fade In/Out transitions for the track of CNN1 video file
(a) Example of Fade Out transition detection (b) Example of Fade In transition detection

Algorithm 1: Fade In/Out transition detection

Input: Gray scale images of video frames (f_1, f_2, f_3, ……. $f_{NumFram}$).

Output: Fade IN/Out transition frames index.

```
j←0                        //counter for Fade IN/Out transition frames
Max ← 10                   // maximum MSER regions in the increasing/ decreasing patterns
For i = 1: NumFram   // NumFram: number of frames in input video
    Reg ← extract MSER features (fi)
    NumMSER (i) ← count of MSER regions (Reg)
End For
Find all Conindex that less than Max in NumMSER  //Conindex: consecutive indices
Record each Conindex in FadeList
For i = 1 to length of FadeList
    If ((not all values in Conindex(i)==0) & (not all values in Conindex(i) >0))
        FadeIN/OutIndex(j) ← start index of Conindex(i)
        FadeIN/OutIndex(j) ← end index of Conindex(i)
        j ← j+2
    End If
End For
```

2. Abrupt Detector

In abrupt (hard cut) shot the last frame of a shot is very different from the first frame of the next shot, so the similarity measure between two consecutive shots is always very low. The abrupt detector consists of three stages: feature frame generation, similarity index calculation, and an adaptive threshold. In the generation stage of the feature frame, Gist descriptor is computed to all video frames to obtain the Gist features frames (GF). It is worth to mention that all frames in a preliminary stage are pre-processed by resizing them to 256x256 and converting them into Grayscale images. The Gist is scaled invariable feature because Gist descriptor gives the scene of the image (it does not represent the details of an image), so changing the image size will not alter the Gist features.

Because the correlation measure has high accuracy in matching, it is used to calculate the similarity index value between two consecutive Gist feature frames. If the correlation coefficient is closer to one, indicates the similarity between frames is high, while the correlation coefficient closest to zero indicates the dissimilarity between frames is high. The last stage of the abrupt detection is the selection of the threshold as an important parameter to achieve high accuracy in recognizing the abrupt transition when comparing changes between the two frames. The predefined global thresholds are not efficient because of the video information change dramatically, as well as it is impossible to find a universal threshold that is fit all variety class of videos. Thus, a local adaptive threshold is proposed which is changing along with video content to detect camera breaks. The mean value μ_{GF} and standard deviation σ_{GF} are calculated according to the similarity index values of the window size m, then threshold Th_{cut} is generated as:

$$Th_{cut} = (\mu_{GF} + \sigma_{GF}) * p \qquad \textbf{(1)}$$

Where the value of m is determined as 100 frames and p is assumed to be 0.88 through a lot of experiments. The Th_{cut} values are saved in an array of length N (N= number of frames in input video /100) used for detection purposes.

An abrupt is present between G_{Fi} and G_{Fi+1} if their correlation similarity value is smaller than a threshold (Th_{cut}). To minimize the presence of high activities such as flashlight effect or rapid object motion that causes the appearance of transition or disappearance, the condition of the abrupt detection method is defined as:

$$CorGist_i = \begin{cases} true, (CorGist_i < th) \ \& \ ((CorGist_{i+1}, CorGist_{i+2}, CorGist_{i-1}, CorGist_{i-2}) > th_{cut}) \\ false, & otherwise \end{cases} \quad (2)$$

Where the $CorGist_i$ correlation value between adjacent Gist future frames, $CorGist_{i+1}$ and $CorGist_{i+2}$ are two correlation values preceding current correlation value $CorGist_i$, while $CorGist_{i-1}$ and $CorGist_{i-2}$ are two correlation values previous current correlation value $CorGist_i$. An example of abrupt detection is illustrated in figure 4 while the procedure of the abrupt detector is described in Algorithm 2.

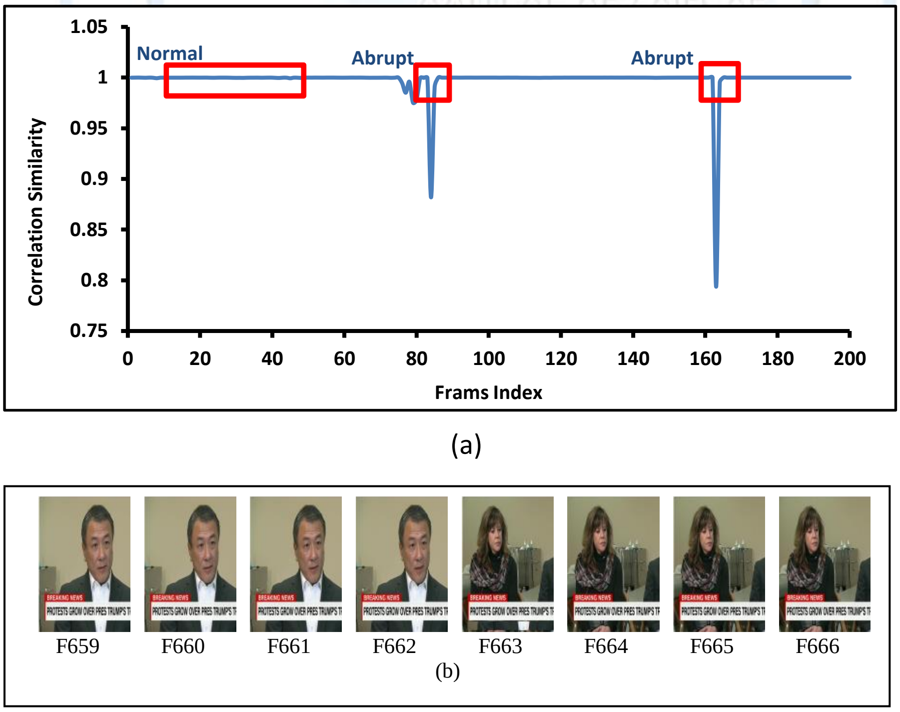


Figure 4: Abrupt transition for the track of CNN2 video file
(a) Example of abrupt and normal transition (b) Example of abrupt detection

Algorithm 2: Abrupt transition detection

Input: An array of threshold value (Th_{cut}), Gist features frames (GF).

Output: Abrupt transition frames index.

```
s←1   //counter for Thcut array
j←1   //counter for cut transition frames
For i = 1 : NumFram-1    // NumFram: number of frames in input video
    CorGist(i) ← Correlation (GF(i), GF (i+1))
End For
For i = 1 to length of CorGist
    If  (Eq. (2) with Thcut (s))
        Abrupt Index(j) ← i
        j ← j+1
    End If
If  ( i  mod 100) = = 0
    s←s+1   // sliding the threshold window
End If
End For
```

3. Gradual Detector

Determining gradual transitions is more difficult when compared to determine the abrupt transitions. This difficulty arises because of the slow varying frame features usually takes place over many frames. Gist feature can keep good invariance in geometric distortion and it can also reflect semantic information. Therefore, Gist features are used for introductory detection of video gradual shot (dissolve) in proposed gradual detector algorithm while color histogram feature is employed for finishing detection to eliminate the wrong shot boundaries caused by object motion or camera operation. Analysis the behavior of dissolve transition patterns for large video files using Gist features concludes that dissolve behavior varies in an increasing or a decreasing pattern as shown in figure 5.

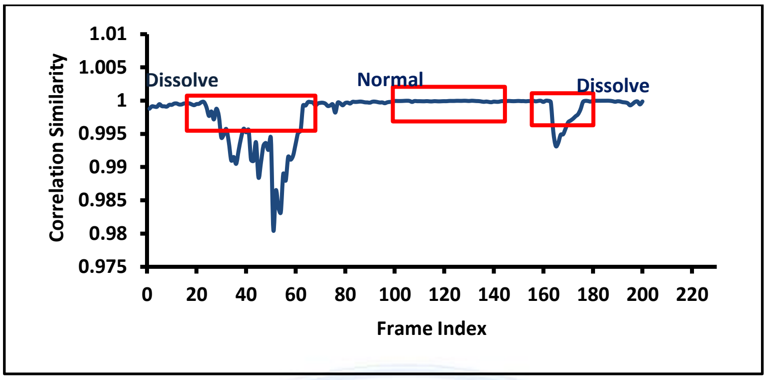


Figure 5: Multiple dissolve transitions for the track of BBC3

The proposed algorithm defines gradual transition considering that any increasing or decreasing pattern in sequence of Gist feature frames as candidates, for example, consider frames f1, f2 and f4 are present but frame f3 is missing after the above operation; then the pattern of the frames will be (f1, f2, f3, f4). The candidates' gradual transition must satisfy the condition in Eq3 otherwise the possible gradual transition is discarded:

$$Correlation(\mathrm{GF}_s - 1, \mathrm{GF}_e + 1) \le Th_{grad} \quad \textbf{(3)}$$

where GF_s -1 and GF_e+1 are the Gist feature frame index of the previous frame of the starting (GF_s) and preceding frame of the end (GF_e) of the possible gradual transition of length L respectively. The length (L) of gradual transition must be longer than six frames; this inherently assumes that all gradual transition frames appear in the video frames sequence longer than six frames, which is a valid assumption for most domains. Th_{grad} is a threshold used for detecting of the gradual transition, which is determined experimentally. Although the gradual detection using the Gist features is robust for the local motion of the object, it is not proper for object movement over multiple frames and large camera motion thus Gist features may not distinguish between gradual transitions and camera motion changes. Therefore, color histogram features are used for finishing preliminary gradual boundary detection due to its robust to the video frame translation, rotation and camera movement and also, it's simple to calculate.

Filtering is done by considering two windows of length not more than 10 frames. The left window content the previous frames sequences of the starting frame index of the possible gradual transition (GF_s) while the right window content the preceding frames sequences of the end frame index of the possible gradual transition (GF_e). The color histogram similarity is calculated for each frame from (GF_s -Wleft) to (GF_e+ Wright) where the Wleft and Wright is the length of left and right window respectively.Before declaring a gradual transition, the following two conditions are must satisfy otherwise the candidates' gradual transition is rejected. The first condition is finding at least one value of color histogram similarity within range (GF_s - Wleft) and within range (GF_e + Wright) that must be larger than a threshold value of Th_{hist}. The indexes of these values represent the new boundary of a gradual transition. The second condition is to check the uniformity of color histogram similarity between the new indexes if satisfy the first condition. To verify the uniformity, the mean value of the color histogram similarity between the new indexes is must be larger than a threshold value of $Th_{histmean}$. Th_{hist} and $Th_{histmean}$ threshold values are chosen experimentally. An example of the filtering process to wrong and right detection of gradual transition using color histogram feature is illustrated in figure 6 while figure 7 shown the frames of the corresponding shots in figure 6. The procedure of the gradual shot detection process is illustrated in Algorithm 3.

GFs-L	……	GFs-1	GFs							GFe	GFe+1	……	GFe+L
Left windows			Candidate transition frames								Right windows		

(a)

F4620	F4621	F4622	F4623	F4624	F4625	F4626	F4627	F4628	F4629	F4630	F4631	F4632	F4633	F4634	F4635
	0.9805	0.9864	0.9839	0.9832	0.9889	0.9879	0.9915	0.9911	0.9917	0.9934	0.9952	0.9955			
	Correlation similarity of gist features														

F4620	F4621	F4622	F4623	F4624	F4625	F4626	F4627	F4628	F4629	F4630	F4631	F4632	F4633	F4634	F4635
0.9491	0.9555	0.7421	0.7386	0.6945	0.6768	0.6615	0.6536	0.6496	0.6455	0.6333	0.6266	0.6157	0.5907	0.9551	0.9459
Wleft		Correlation similarity of histogram features												Wright	
Satisfy the two condition: Accept shot															

(b)

F2303	F2304	F2305	F2306	F2307	F2308	F2309	F2310	F2311	F2312	F2313	F2314	F2315	F2316	F2317	F2318
		0.9982	0.9987	0.9985	0.9953	0.9986	0.9987	0.9991	0.9986	0.9982	0.9993	0.9994	0.9995		
		Correlation similarity of gist features													

F2303	F2304	F2305	F2306	F2307	F2308	F2309	F2310	F2311	F2312	F2313	F2314	F2315	F2316	F2317	F2318
0.7991	0.7667	0.8342	0.8468	0.8324	0.7914	0.8704	0.8448	0.8703	0.8678	0.8321	0.8836	0.9111	0.9281	0.9507	0.9321
Wleft		Correlation similarity of histogram features												Wright	
Not satisfy the two condition: Reject shot															

(c)

Figure 6: Calculate of gradual transition detection for BBC3 video file (a) General idea of filtering process.
(b) Example of correct gradual detection. (c) Example of incorrect gradual detection.

Algorithm 3: Gradual transition detection.
Input: Correlation matrix of Gist (CorGist), Correlation matrix of Histogram (CorHis), Th_{grad}, $Th_{histmean}$, Th_{hist}.
Output: Gradual transition frames index.

```
Wleft ← length of the left window
Wright ← length of the right window
Min  ← 6                          // Minimum number in the gradual transition frames
Find all Inindex in CorGist     // Inindex: increasing consecutive indices series
Find all Deindex in CorGist     // Deindex: decreasing consecutive indices series
Gtran ← union(Inindex, Deindex)
Fill the missing number between two consecutive numbers series in Gtran
For i = 1: i+2: length of Gtran
    If  ( differences (Gtran (i), Gtran(i+1)) > Min )
        Cor ← Correlation (GF(Gtran(i)-1), GF(Gtran(i+1)+1))
        If (Cor < Thgrad )
            Gindex(j) ← Gtran(i)
            Gindex (j) ← Gtran(i+1)
            j ← j+2
        End If
    End If
End For
  k ← 1                        // finishing preliminary gradual boundary detection
For i= 1:2: length of Gindex
   Garindex (k) ← Verify-Histogram (Gindex(i)- Wleft, Gindex(i), CorHis)
   k← k+1
   Garindex (k) ← Verify-Histogram (Gindex(i), Gindex (i+1) + Wright, CorHis)
    k← k+1
End For
k ← 1
For i=1:2:length of   Garindex
    meanhist ←   mean (CorHis (Garindex(i) to Garindex(i+1))
    If meanhist  < Thhistmean
        GardualIndex (k) ← Garindex(i)
        GardualIndex (k+1) ← Garindex(i+1)
        k ← k+2
    End If
End For
```

Continue of Algorithm 3

```
Function Verify-Histogram (f, l, CorHis)
Temp←0
For  j= CorHis (f):  CorHis (l)
        If CorHis (j) > Thhist
            temp ←j
            break;
        End If
End For
Return temp
End Function
```

Experimental Results

Evaluation tests are performed on the proposed algorithm using selected video files as test material. The video files are obtained from the BBC archive [33] and CNN archive [34] and the Internet [35], Table1 shows the details of all selected video files which includes abrupt, dissolve and fade in/out transitions. The measures used in this work to evaluate the performance of the proposed detection algorithm are Recall (R), Precision (P) and F- measure (F) as defined in the following equation:

$$\text{Recall(R)} = \frac{\text{True}}{\text{True+Miss}} \tag{4}$$

$$\text{Precision(P)} = \frac{\text{True}}{\text{True+False}} \tag{5}$$

$$\text{F}-\text{measure(F)} = \frac{2*\text{Precision}*\text{Recall}}{\text{Precision+Recall}} \tag{6}$$

Table1: Specifications of video files used for proposed SBD.

Video	Number of frames	Transition model			Total	Source
		Abrupt	Dissolve	Fade		
BBC1	8775	43	4	0	47	BBC archive https://www.bbc.co.uk
BBC2	7670	31	0	0	31	
BBC3	7401	29	3	1	33	
BBC4	5243	14	6	0	20	
BBC5	15379	26	6	0	32	

BBClearn1	14578	68	0	0	68	
BBClearn2	12453	45	0	0	45	
CNN1	1800	38	0	2	40	CNN archive https://edition.cnn.com
CNN2	4397	14	9	0	23	
MT1	4180	71	0	5	76	Movie Trailer https://www.youtube.com
MT2	3421	84	0	21	105	

Where True, Miss, False is the number of correct, missed and false transitions detected respectively. High precision (low rate of false boundary detection) and high recall (low rate of miss in detecting a boundary) indicate perfect shot transition detection. In order to obtain the optimal thresholds for gradual detection an experiment was performed by changing their values until the best overall performance is achieved, the thresholds are set to Th_{grad}=0.92, Th_{hist}=0.85 and $Th_{histmean}$= 0.75. The results of the experiments are presented in Table 2, while the performance of the proposed algorithm according to correctly detect the whole shot boundary. The recall, precision and F rate of in Table 3 are very high, indicating the high performance of the proposed algorithm. Recall, Precision and F measure are presented in Table3.

Table 2: Proposed SBD algorithm result

Video	Abrupt Transition			Dissolve Transition			Abrupt Transition		
	True	False	Miss	True	False	Miss	True	False	Miss
BBC1	43	0	0	4	0	0	0	0	0
BBC2	31	0	0	0	0	0	0	0	0
BBC3	29	0	0	3	0	0	1	0	0
BBC4	14	0	0	4	2	0	0	0	0
BBC5	26	0	0	6	0	0	0	0	0
BBClearn1	45	0	0	0	0	0	0	0	0
BBClearn2	68	0	0	0	0	0	0	0	0
CNN1	36	1	0	0	1	0	2	0	0
CNN2	12	0	2	8	2	1	0	0	0
MT1	70	0	1	0	0	0	5	0	0
MT2	83	1	1	0	0	0	21	0	0

It is clear from Table 2 that the effect of the proposed SBD algorithm is good and it is able to correctly detect the whole shot boundary. The recall, precision and F rate of in Table 3 are very high, indicating the high performance of the proposed algorithm. One can observe the performance of the Abrupt and Fade In/Out detector is better in comparison to the gradual detector. The Abrupt detector algorithm handles the object and camera motion (zoom and pan) very well without false detection in a long shot as shown in figure 8. Thus, the Abrupt detector algorithm has some robustness for object /camera motion and illumination changes.

Table 3: Proposed SBD algorithm performance

Video	Recall	Precision	F measure
BBC1	1.000	1.000	1.000
BBC2	1.000	1.000	1.000
BBC3	1.000	1.000	1.000
BBC4	1.000	0.900	0.947
BBC5	1.000	1.000	1.000
BBClearn1	1.000	1.000	1.000
BBClearn2	1.000	1.000	1.000
CNN1	1.000	0.950	0.974
CNN2	0.869	0.909	0.888
MT1	0.986	1.000	0.992
MT2	0.991	0.991	0.990
Average	0.986	0.977	0.981

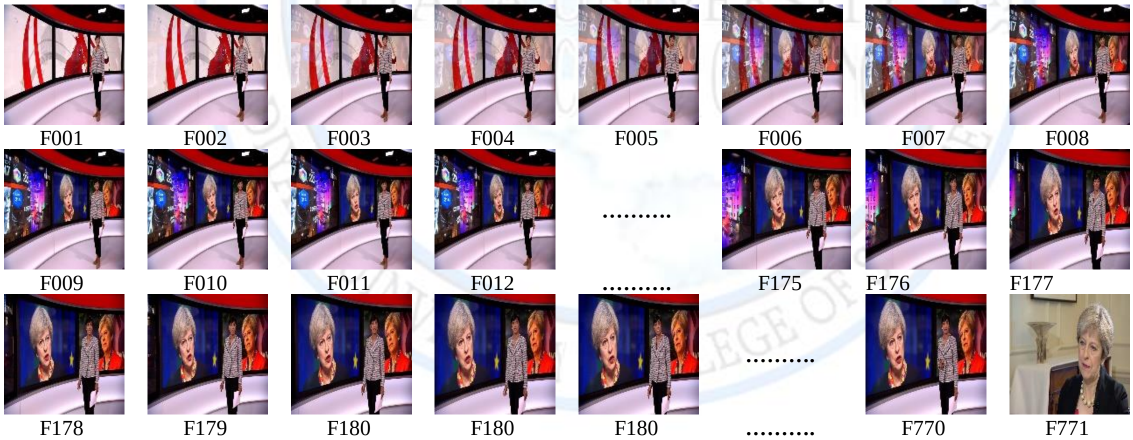


Figure 8: Abrupt detection to long shot with object & camera motion (BBC2 video)

However, gradual detector algorithm is relatively low. The false or miss detections of gradual transition are mainly due to high-speed camera motions or slow color changes the change. For example, one of a gradual transition in BBC4 video, some objects move so fast, so it is easy to be detected falsely as shown in figure 9a. On the other hand, a gradual transition in CNN2 video, the color changes are not clear, so the detection is lost as shown in figure 9b.

(a) BBC4: Fast object motion

(b) CNN2: Color changing slowly

Figure 9: Example of wrong gradual detection
(a) False gradual transition detection (b) Missing gradual transition detection

Conclusions

A new algorithm for abrupt and gradual shot boundary detection is presented in this paper. The algorithm has the ability to distinguish between different types of boundaries: abrupt, fades, and dissolves. The MSER descriptors are employed in the proposed algorithm to detect a fade transition, while Gist features are employed in the proposed algorithm to detect abrupt and dissolve transitions due to its ability to reflect semantic information. AT is detected by evaluating the correlation similarities between successive Gist features frames according to the local adaptive threshold. Gist features may interpret camera movements as GT. To overcome this limitation, color histogram feature is used to finish detection of gradual boundary due to its robust against camera and object movement. The experimental results illustrate the efficiency of the proposed algorithm that can achieve high precision and recall due to the reduction of undesirable effect, camera changes and object movement. Although the use of more features makes the situation better, this does not solve the problem radically, especially to detect the gradual transitions in the presence of sudden camera shake and very high motion, therefore further work is still needed.

References

1. Albanese, M.; Chianese, A.; Moscato, V.; Sansone, L. Multimedia Tools and Applications 2004, 24(3), 253-272.
2. Janwe, N. J.; Bhoyar, K. K. Video shot boundary detection based on JND color histogram, Proceedings of Image Information Processing, IEEE 2013, 476-480.
3. Singh, R. D.; Aggarwal, N. in Advances in Intelligent Informatics, El-alfy, E. S., Eds.; Springer, Cham, 2015, pp457-469.
4. Lee, H.; Yu, J.; Im, Y.; Gil, J.M.; Park, D. Multimedia Tools and Applications 2011, 51(3), 1127-1145.
5. Qian, X.; Liu, G.; Su, R. IEEE Transactions on Circuits and Systems for Video Technology 2006, 16(10), 1245-1258.
6. Huo, Y.; Wang, Y.; Hu, H. Effective algorithms for video shot and scene boundaries detection Proceedings of International Computer and Information Science, IEEE 2016, 1-6.
7. Domnic, S. Video Shot Cut Detection Using Least Square Approximation Method, in Computer Networks and Intelligent Computing, Venugopal, K. R.; Patnaik, L.M., Eds.; Springer Berlin Heidelberg 2011, 161-170.
8. Liu, Y.; Zhang, D.; Lu, G.; Ma, W. Pattern recognition 2007, 40(1), 262-282.
9. Shao, H.; Qu, Y.; Cui, W. Shot boundary detection algorithm based on HSV histogram and HOG feature, Proceedings of the 5th International Advanced Engineering Materials and Technology 2015, 951-957.
10. Ling, X; Chao, L.; Huan, L.; Zhang, X. A general method for shot boundary detection, Proceedings of International Multimedia and Ubiquitous Engineering, IEEE 2008, 394-397.
11. Amel, A.M.; Abdessalem, B. A.; Abdellatif, M. Journal of Telecommunications 2010, 2(1), 54-59.

12. Whitehead, A.; Bose, P.; Laganiere, R. Feature based cut detection with automatic threshold selection, in Image and Video Retrieval, Enser P.; Eds.; Springer, Berlin, Heidelberg 2004, 410–418.

13. Huan, Z.; Xiuhuan, L.; Lilei, Y. Shot boundary detection based on mutual information and canny edge detector, Proceedings of the International Computer Science and Software Engineering, IEEE 2008, 12(2),1124-1128.

14. Yoo, H. W.; Ryoo, H. J.; Jang, D. S. Multimedia Tools and Applications 2006, 28(3), 283-300.

15. Kar, T.; Kanungo, P. Signal, Image and Video Processing 2017,11(7), 1237–1244.

16. Thounaojam, D. M.; Roy, S.; Singh, K. M. Indian Journal of Science and Technology 2016, 9(7), 1-5.

17. Volkmer, T.; Tahaghoghi, S. M. M.; Williams, H. E. Gradual transition detection using average frame similarity, Proceedings of Computer Vision and Pattern Recognition Workshop, IEEE 2004, 139-139.

18. Huang, C.; Liao, B. IEEE Transactions on Circuits and Systems for Video Technology 2001, 11(12), 1281-1288.

19. Zhang, H.; Hu, R.; Song, L. A shot boundary detection method based on color feature, Proceedings of international Computer Science and Network Technology, IEEE 2011, 2541-2544.

20. Xiao, Y.; Xia, L.; Zhu, S.; Huang, D.; Xie, J. Mathematical Problems in Engineering 2013, 1-10.

21. Hannane, R; Elboushaki, A.; Afdel, K.; Naghabhushan, P.; Javed, M. International Journal of Multimedia Information Retrieval 2016, 5(2), 89-104.

22. Amiri, A.; Fathy, M. *EURASIP Journal on Advances in Signal Processing* 2010, 2009(1), 1-12.

23. Fan, J.; Zhou, S.; Siddique, M.A. Multimedia Tools Applications 2017,76(7), 10169–10190.

24. Zhao, S.; Zhuo, L.; Xiao, Z.; Shen, L. A Data-Mining Based Video Shot Classification Method, Proceedings of Image and Signal Processing, IEEE 2009, 1-4.

25. Tong, W.; Song, L.; Yang, X.; Qu, H.; Xie, R. CNN-Based Shot Boundary Detection and Video, Proceedings of Broadband Multimedia Systems and Broadcasting, IEEE 2015, 1-5.

26. Xu, J.; Song, L.; Xie, R. Shot boundary detection using convolutional neural networks, Proceedings of Visual Communications and Image Processing, IEEE 2016, 1-4.

27. Sun, X.; Zhao, L.; Zhang, M. A novel shot boundary detection method based on genetic algorithm-support vector machine, Proceedings of international Intelligent Human-Machine Systems and Cybernetics, IEEE 2011, 144-147.

28. Mondal, J.; Kundu, M.K.; Das, S.; Chowdhury, M. Multimedia Tools and Applications 2017,77(7), 8139-8161.

29. Domnic, S.; Priya, L.G.G. IEEE Transactions on Image Processing 2014, 23(12), 5187-5197.

30. Douze, M.; Jégou, H.; Sandhawalia, H.; Amsaleg, L.; Schmid, C. Evaluation of gist descriptors for web-scale image search, Proceedings of International Image and Video Retrieval 2009, 1-8.

31. Sharma, N.; Nidhi Indian Journal of Science and Technology 2017, 10(17), 1-12.

32. Wang, X. Y.; Wu, J. F.; Yang, H. Y. Multimedia Tools and Applications 2010, 49(2), 323-345.

33. https://www.bbc.co.uk.

34. https://edition.cnn.com.

35. https://www.youtube.com.