



## Temporal Video Segmentation Using Optical Flow Estimation

Eman Hato

Computer Science Department, College of Science, Mustansiriyah University, Baghdad, Iraq

Received: 30/3/2021

Accepted: 2/8/2021

### Abstract

Shot boundary detection is the process of segmenting a video into basic units known as shots by discovering transition frames between shots. Researches have been conducted to accurately detect the shot boundaries. However, the acceleration of the shot detection process with higher accuracy needs improvement. A new method was introduced in this paper to find out the boundaries of abrupt shots in the video with high accuracy and lower computational cost. The proposed method consists of two stages. First, projection features were used to distinguish non boundary transitions and candidate transitions that may contain abrupt boundary. Only candidate transitions were conserved for next stage. Thus, the speed of shot detection was improved by reducing the detection scope. In the second stage, the candidate segments were refined using motion feature derived from the optical flow to remove non boundary frames. The results manifest that the proposed method achieved excellent detection accuracy (0.98 according to F-Score) and effectively speeded up detection process. In addition, the comparative analysis results confirmed the superior performance of the proposed method versus other methods.

**Keywords:** Abrupt transition, Candidate transition, Projection, Optical flow, Video Segmentation.

### تقسيم الفيديو باستخدام تخمين تغيير التدفق البصري

إيمان هاتو

قسم علوم الحاسوب، كلية العلوم، الجامعة المستنصرية، بغداد، العراق

### الخلاصة

أكتشاف حدود اللقطة هو عملية تقسيم مقطع فيديو إلى وحدات أساسية تُعرف بإسم اللقطات من خلال أكتشاف الإطارات الإنتقالية بين اللقطات. تم إجراء الأبحاث لإكتشاف حدود اللقطة بدقة. ومع ذلك ، فإن تسريع عملية أكتشاف اللقطات بدقة أعلى يحتاج إلى تحسين. في هذا البحث، تم تقديم طريقة جديدة لإكتشاف حدود اللقطات المفاجئة في الفيديو بدقة عالية وتكلفة حسابية أقل. تتكون الطريقة المقترحة من مرحلتين. أولاً ، تم استخدام ميزات الإسقاط للتمييز بين الإنتقالات غير الحدودية والإنتقالات المرشحة التي قد تحتوي على حدود مفاجئة. يتم الإحتفاظ فقط بالإنتقالات المرشحة للمرحلة الثانية. وهذا يساعد على تحسين سرعة اكتشاف اللقطة عن طريق تقليل نطاق الكشف. في المرحلة الثانية ، تم تنقيح المقاطع المرشحة باستخدام ميزة الحركة المشتقة من التدفق البصري لإزالة الإطارات غير الحدودية. أظهرت النتائج أن الطريقة المقترحة حققت دقة كشف ممتازة (98 حسب مقياس F ) وتسريع عملية الكشف بشكل فعال. بالإضافة إلى ذلك ، أكدت نتائج المقارنة مع البحوث الأخرى الأداء المتفوق للطريقة المقترحة مقابل الطرق الأخرى.

## 1. Introduction

Video is one of the important forms of multimedia. It is widely consumed in large applications like education, medicine, surveillance, sports and entertainment. Video is a multi-dimensional signal structured as a series of still images called frames. The frames content changes in time form a smooth movement for the observer. Successive frames without significant changes in visual content are grouped together and compose the so-called shots. Each video shot represents a frames series captured by a single camera. While a set of shots with semantic related content is called a scene [1].

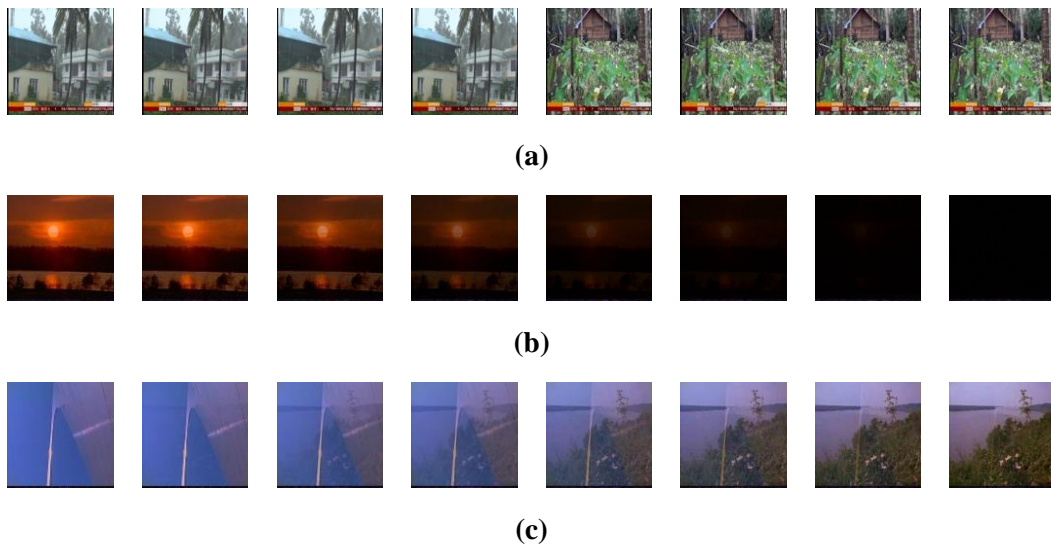
The aim of temporal video segmentation is to split a video into shots as meaningful unit of the desired information in a large amount of video information. It represents the first step of many video processing for different applications. The essence of video temporal segmentation techniques (known as shot boundary detection techniques) is to discover the shot boundaries. The boundaries are the discontinuity frame(s) that describe the transition from one shot to the next [2].

The types of shot transitions in videos can be classified into the following categories [3, 4]:

- Abrupt transition (Hard transition): sudden changes of content from one shot to the next, which means that the shots are separated by a single frame.
- Gradual transition (Soft transition): gradual changes of content from one shot to the next, which means that the shots are separated by several frames with very closely related visual content.

The gradual transition can be furthermore categorized as follows:

- Fade - in: a new shot appears by increasing gradually the brightness of black color frame.
- Fade - out: a new shot disappears by decreasing gradually the brightness of frame.
- Dissolve: a current shot is overlapping with an incoming one. Examples of different types of shot transitions are shown in Figure 1.



**Figure-1** Shot transitions types (a) Abrupt transition (b) Fade - out transition (c) Dissolve transition

The main challenges of shot boundary detection are camera and objects movements because they may alter video content dramatically, giving an impact similar to transition effects and causing incorrect detection of shot transition. Additionally, most algorithms rely on frame-by-frame comparison to discover the boundaries of the shot, which causes a high computational cost. A shot detection algorithm should have the ability to reduce computation costs and

recognize the shots boundaries without misconstruction the object or camera motion as the shot transition [5, 6].

This paper presents a new method to detect abrupt transition between successive shots in the video with high detection accuracy and lower execution time as much as possible, regardless of the complexity of the video in terms of camera operations, motion effects, and illumination changes. The paper is organized as follows: the literature review is discussed in the next section. In section 3 projection features are described. Optical flow estimation is presented in section 4. The proposed method is described in section 5. In section 6 the experimental results obtained are discussed. Finally, conclusions are contained in section 7.

## 2. Literature Review

Video segmentation is widely used as first step for many applications such as video categorization, video indexing, video summarization and content based video retrieval. Many researchers have tried to detect the shot transitions depending on the comparison of consecutive frames by focusing on the visual features of video frames. First, the features are extracted from each frame, then distance or similarity measures are applied and compared to a specific threshold to distinguish the abrupt transition [7].

Histogram is a commonly used feature for shot detection techniques which are calculated using color space or a gray level. Liu and Dai [8] presented a method of shot boundary detection using grey model based on set sequence (the model was denoted as SGM) and Hue Saturation Intensity (HSI). The frame was converted to HSI and then a histogram was computed and sampled by SGM. Based on absolute mean error with thresholds, the abrupt shot was detected. Hong Shao *et al.* [9] exploited color histogram with Hue Saturation Value (HSV) and features of Histogram of Gradient (HOG) to detect abrupt shot.

In order to decrease the influence of camera/object motion, Kar and Kanungo [10] used absolute sum gradient orientation features with threshold generated from the local and global thresholds for detecting abrupt shots. Thounaojam *et al.* [11] employed features extracted from Gray Level Co-occurrences Matrix (GLCM) and used correlation measure to similarity between two consecutive GLCMs of the video. Zajic *et al.* [12] formed a structure to measure color and texture differences between frames of decoded video sequence for abrupt transition detection. These methods are robust for camera/object motion, but require significant computational power.

Local features extracted from the video frames can be employed to discover shot boundaries. Santos and Pedrini [13] introduced a method for detecting shots boundary by using the weber local descriptor with an adaptive threshold. A shot detection method designed by Hannane *et al.* [14] combined local and global features by using the distribution histogram of SIFT points for consecutive frames with adaptive threshold to detect the abrupt and gradual transitions. Other authors [15, 16] proposed a shot boundary detection method based on fuzzy logic as an important tool for computer research.

Several works in the literatures employed machine learning techniques. For example, Mondal *et al.* [17] introduced a Least Squares Support Vector Machine (LS-SVM) as a modified version of SVM. Features were extracted using the multiscale geometric analysis of non-subsampled contourlet transform and reduced using principal component analysis. The frames were classified using LS-SVM and features vectors into non transition, abrupt transition or gradual transition classes. Convolutional Neural Networks (CNNs) were utilized by Xu *et al.* [18]. Shot boundary positions were determined using a candidate segment selection method and adaptive thresholds. Then, CNN was executed to extract representative features of the frames in the candidate segments. Finally, shot transitions were obtained by using pattern matching method. However, the performance of these methods is effective and achieves high accuracy, but the execution speed is slow.

Researchers have tried to achieve greater accuracy and high execution speed. Lu and Shi [19] achieved an important step forward in reducing the execution time of shot boundary detection techniques. They adopted Singular Value Decomposition (SVD) and selection of a candidate segment to reduce the time execution of shot boundary detection. First, the candidate segments that may contain the shot boundaries were identified. Next, color histograms were extracted as features matrix of candidate segment frames. To reduce the feature dimension, SVD was then implemented. Finally, shot transitions were determined using pattern matching method. Dhiman *et al.* [20] used pixel-based technique with selection of candidate segment to accelerate detection of shot boundaries. Hato and Abdulmunem [21] provided a fast method for shot boundary. First, the features of SURF were extracted from the half number of the video frames. Then, the features matching were performed using the distance function. Finally, the similarities were calculated and compared to the global threshold.

One can conclude after reviewing the techniques of shot boundary detection, that the main challenges of detection processing are fast object movement, large camera motion and cost effectiveness. Some techniques achieved high accuracy at the cost of execution time. Hence, shot boundary detection techniques still need to provide accurate results with faster execution speed. Therefore, this proposed method aims to achieve high detection accuracy and maintain low computation costs.

### 3. Projections

The projections of an image are a one-dimensional representation of its contents, which are calculated on the coordinate axes. Horizontal projection is the sum of the pixel values in the image row, while vertical projection is the sum of the pixel values in the image column. For the image  $I(r, c)$  of size  $M \times N$ , the horizontal projection HP and vertical projection VP are defined as [22]:

$$HP(r) = \sum_{c=0}^{N-1} I(r, c) \quad \text{for } 0 < r < M \quad (1)$$

$$VP(c) = \sum_{r=0}^{M-1} I(r, c) \quad \text{for } 0 < c < N \quad (2)$$

The horizontal projection has an  $M$  length and the vertical projection has an  $N$  length that corresponds to the height and width of the image, respectively. Projections are oftentimes used to quickly analyze the image structure and isolate its component parts. Additionally, projections provide shape information that is useful in applications such as text detection and character recognition, where the objects of interest can be normalized with respect to size [22].

### 4. Optical Flow Estimation

The optical flow can be defined as the velocity distribution of motions for captured images data. These motions are estimated using the comparison of two images (or two frames in the case of video), which are captured on two different times or exactly at the same time but using two cameras. In general, optical flow involves two-dimensional vectors, where each vector is representing the displacement pixel from the first frame to the next frame [23].

Consider a pixel  $I(x, y, t)$  in the first frame, which moves to the next frame by taking time  $\Delta t$ . The  $(\Delta x, \Delta y)$  is the pixel displacement from the first frame to the next frame, so that the pixel in the next frame will be [24]:

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t) \quad (3)$$

The differential method (partial derivatives) is used to find pixel displacement with respect to the temporal and spatial coordinates, as follows:

$$I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x, y, t) + \frac{\partial I}{\partial x} \Delta x + \frac{\partial I}{\partial y} \Delta y + \frac{\partial I}{\partial t} \Delta t + c \quad (4)$$

where  $I(x, y, t)$  is the pixel at location  $(x, y, t)$ ,  $c$  is a real constant valued, and  $\Delta x, \Delta y, \Delta t$  are the movements between the two frames. The equation (5) is obtained from the equation (4):

$$\frac{\partial I}{\partial x} \Delta x + \frac{\partial I}{\partial y} \Delta y + \frac{\partial I}{\partial t} \Delta t = 0 \quad (5)$$

This results in:

$$\frac{\partial I}{\partial x} V_x + \frac{\partial I}{\partial y} V_y + \frac{\partial I}{\partial t} \Delta t = 0 \quad (6)$$

where  $V_x = \frac{dx}{dt}$  and  $V_y = \frac{dy}{dt}$  are the velocity or optical flow of  $I(x, y, t)$  and  $\frac{\partial I}{\partial x}$ ,  $\frac{\partial I}{\partial y}$  and  $\frac{\partial I}{\partial t}$  are frame gradients at  $(x, y, t)$ .

Equation (6) cannot be solved with two unknown variables  $V_x$  and  $V_y$ . Therefore, many differential methods, such as Farneback and Lukas-Kanade, are provided to solve this problem [24]. The Farneback algorithm is a method for estimating the motion of two frames, which is based on a polynomial expansion to approximate a neighborhood of each pixel. More details about this method can be found elsewhere [23].

## 5. Proposed Video Segmentation Method

A simple and effective method is proposed for detecting abrupt shots between the video frames. Candidate's transitions are first detected based on an analysis of the changing pattern of the projection features. Motion feature, estimated by optical flow, is utilized to filter out the candidate's transitions. The proposed method is composed of two stages: candidate's transition detection and abrupt transition specifying.

### 5.1. Candidates Transition Detection

An abrupt transition can be detected by looking for two completely different sequential frames. In other words, let  $(F_1, F_2, F_3, \dots, F_N)$  be the video frames,  $F_i$  is considered an abrupt transition if the dissimilarity of  $F_i$  and  $F_{i+1}$  is high. For abrupt shot detection, extracting the optical flow field from all frames in the video and performing feature matching is computationally very expensive. Therefore, candidate frames are first detected from the video based on projection features due to their ability to reduce the information redundancy of video frames. Horizontal Projection (HP) is computed to all video frames for introductory detection of abrupt transition, so each video frame is represented by HP vector of  $M$  length.

Higher - resolution images produce a large number of features, which are recommended for image classification and retrieval. But in case of shot boundary detection, it is not mandatory to use high - resolution images; a fewer numbers of features can decide whether two consecutive frames are different or similar. All frames are pre-processed by resizing them to  $256 \times 256$  and converting them into grayscale images before extracting the projection features. The features extracted remain unchanged in low resolution and high resolution frames as well as preserve a low computation time as shown later in the experimental results.

Detection of candidate's transitions focuses on the differences between consecutive frames and the similarities between these differences' values in the nearest neighborhood. The Mean Absolute Differences (MAD) measure is used to calculate the difference value between two consecutive HP vectors. Whenever the difference between two successive HP vectors is higher than the predefined threshold TH1, candidate boundaries are recorded.

The TH1 value can be adjusted experimentally to give high precision or high recall. However, while the lower value of TH1 gives high recall, it also results in lower precision due to the change in video content. On the other hand, high value of TH1 gives high precision but low recall. To overcome the above problem, the TH1 is set to a relatively low value in order to detect all abrupt boundaries, but with false detection being removed by the second stage of the proposed method. The condition for detecting the candidate transition  $C_j$  is defined as:

$$C_j = \begin{cases} \text{Record } i^{\text{th}} \text{ frame , } \text{MAD} (i) > \text{TH1} \\ \text{Discard , } & \text{Otherwise} \end{cases} \quad (7)$$

Where  $\text{MAD} (i)$  is:

$$\text{MAD}(i) = \frac{1}{M} \sum_{r=1}^M |\text{HP}_{i-1}(r) - \text{HP}_i(r)| \quad 1 \leq i < \text{FNum} \quad (8)$$

$C_j$ : represents  $i$ th frame index of video.

$j$ : represents  $j$ th candidate transition.

$i$ : represents  $i$ th frame of video.

FNum: represents the total frames number inside the video.

HP: represents horizontal projection vector of length  $M$  for video frame  $F_i$ .

$M$ : represents number of rows in the video frame (HP length).

$r$ : represents summation of  $r$ th row in HP vector.

## 5.2. Abrupt Transition Specifying

Two differently structured frames may have the same projection. The motion feature is used to increase the discrimination ability of the proposed method and eliminate false detection. The second stage of the proposed method involves applying optical flow estimation to produce the optical flow field. The Farneback algorithm is used as an optical flow computation algorithm that provides a dense and more stable estimation with a reasonable computational cost.

The optical flow algorithm estimates the direction and speed of the moving object in the video. It produces the orientation and the magnitude matrices which are computed from the velocity vector. Discontinuities in the optical flow can help in capturing the temporal variation between the successive frames.

A number of video files were analyzed using the standard division that was computed to magnitude matrices of estimated optical flow fields in order to detect the pattern of abrupt transition. It is concluded that abrupt transition has a relatively small value of standard division. On the other hand, the normal transition has large value of standard division.

Optical flow field is employed to determine discontinuity between the successive frames. Let  $CT = (c_1, c_2, c_3, \dots, c_j)$  represents the indexes of candidates transitions frames, optical flow estimation between the frames from  $F_{c_j-n}$  to  $F_{c_j+n}$  is applied where  $n$  signifies the frames index before and after  $c_j$  and  $n$  is taken as two for experimentation. Then the standard division  $SD$  for each magnitude matrix of optical flow field is computed, which will be  $SD = (sd_{i-2}, sd_{i-1}, sd_i, sd_{i+1}, sd_{i+2})$ .

The determination of abrupt transition is based on comparing  $SD$  values; if the frame  $F_{c_j}$  is abrupt transition then it should have  $sd_i$  value larger than those  $sd$  values of nearest neighbor frames. It was observed from the results that most of the motion and lighting effects were decreased, but there is a possibility that some of the motion affected could be classified as a candidate abrupt transition. Hence,  $sd_i$  value must not only be the largest value but its value also must exceed the experimentally determined threshold condition  $TH2$ , to reduce the detection of false boundaries caused by high activity of video components.

Equation (9) must be satisfied to confirm the candidate's abrupt transition; otherwise the potential transition will be ignored:

$$AT_k = \begin{cases} \text{Record } c_j^{\text{th}} \text{ frame,} & sd_i == \text{Max}(SD) \ \& \ sd_i > TH2 \\ \text{Discard ,} & \text{Otherwise} \end{cases} \quad (9)$$

$AT_k$ : represents  $i$ th frame index of video.

$k$ : represents  $k$ th abrupt transition.

$\text{Max}(SD)$ : represents maximum value of  $SD$ .

## 6. Experimental Results

A collection of videos are used to evaluate the proposed method's performance according to time, cost, and ability to detect abrupt shots. The news video files were downloaded from the BBC archive with the resolution of  $720 \times 1280$  pixels, available at <https://www.bbc.co.uk>. In addition, documentary video files were obtained from the TREC Video Retrieval Evaluation (TRECVID) 2001 test data released on the open video project with the resolution of  $320 \times 240$

pixels, available at <https://open-video.org>. The selected video files reflect a wide variety of content. Detailed description of the ground truth for all information of each video file is provided in Table 1.

Performance was evaluated on the basis of Recall (R), Precision (P) and F-Score (F) measures. These measures can be defined as follows [25]:

$$\text{Recall(R)} = \frac{\text{True}}{\text{True} + \text{Miss}} \quad (10)$$

$$\text{Precision(P)} = \frac{\text{True}}{\text{True} + \text{False}} \quad (11)$$

$$\text{F - Score(F)} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (12)$$

where True represents the correct detection of transition, False represents the false detection of transition, and Miss represents the missed detection of transition. Execution time was also used to evaluate performance; it is a measure of how quickly the proposed method can produce the results.

A video file was provided to the proposed method as an input. Identified transition frames were compared with the corresponding ground truth to identify the True, Miss and False boundaries. The MATLAB R2018a was used to implement the code on hp laptop of Intel CORE i7 and 5500U clocked at 4.40 GHz. TH1 and TH2 were set to 1.8 and 10, respectively. Table 2 summarizes the performance of the proposed method according to R, P, F measures and execution time.

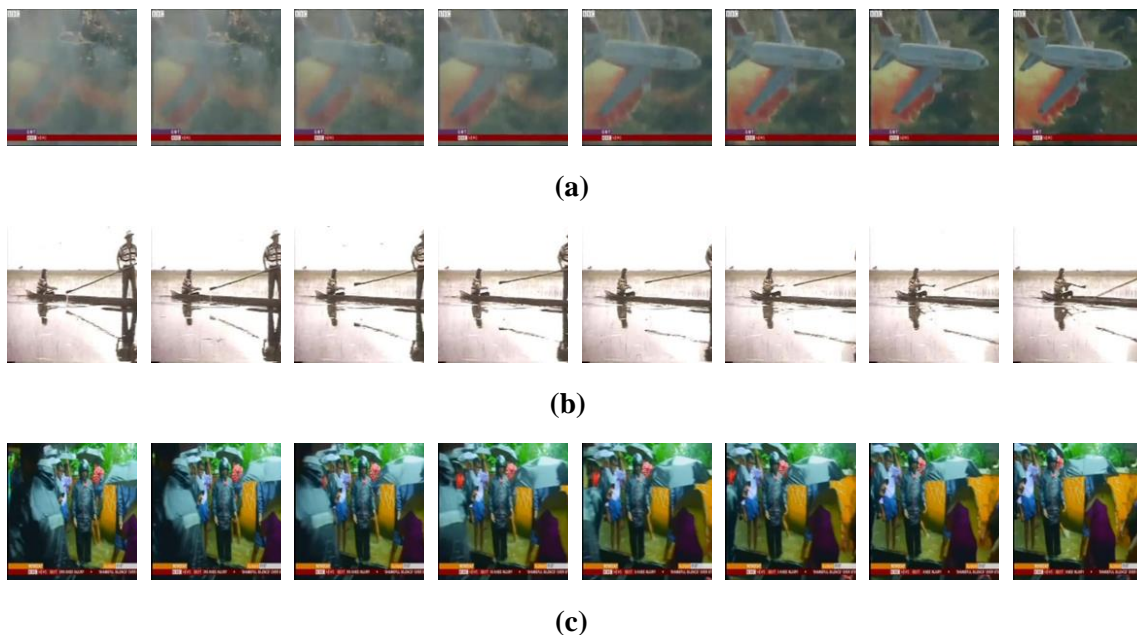
**Table 1-** The ground truth of the videos files.

Video s	Transitio ns Abrupt No.	Frames Number	Time Duration in Sec.	Video Description	Source
VF1	1	1002	33	Documentary video Hidden fury, segment 10	<a href="https://open-video.org">https://open-video.org</a>
VF2	4	707	23	Documentary video New Indians, segment 08	
VF3	5	1594	53	Documentary video Apollo, segment 3002	
VF4	8	1665	55	Documentary video Hoover dam construction, segment 10	
VF5	7	1814	60	Documentary video New Indians, segment 114	
VF6	4	1882	62	Documentary video The future of energy gases,segment09	
VF7	9	1723	57	Documentary video The voyage of the lee, segment 13	
VF8	4	1780	59	Documentary video Winning aerospace, segment 01	
VF9	2	1092	42	News video	
VF10	21	3398	136	News video	
VF11	29	3358	134	News video	
VF12	9	4081	136	News video	
VF13	13	5354	178	News video	
VF14	26	3836	226	News video	
VF15	10	6882	228	News video	
VF16	24	8477	282	News video	

**Table 2-** Performance results of the proposed method.

Videos	R	P	F	Time Cost in Sec.
VF1	1.00	1.00	1.00	1.58
VF2	1.00	1.00	1.00	1.42
VF3	1.00	1.00	1.00	2.72
VF4	0.62	1.00	0.77	4.52
VF5	1.00	1.00	1.00	4.19
VF6	1.00	1.00	1.00	5.73
VF7	1.00	1.00	1.00	3.76
VF8	1.00	1.00	1.00	3.46
VF9	1.00	1.00	1.00	3.62
VF10	1.00	1.00	1.00	9.36
VF11	1.00	1.00	1.00	14.51
VF12	1.00	1.00	1.00	7.66
VF13	1.00	1.00	1.00	11.12
VF14	1.00	1.00	1.00	18.84
VF15	1.00	1.00	1.00	12.88
VF16	1.00	1.00	1.00	20.86
<b>Average</b>	0.97	1.00	0.98	7.88

It is clear from Table 2 that the R, P and F measures are very high, indicating the high performance of the proposed method. In addition, the proposed method has a low execution time. The abrupt boundaries between shots were recognized in the presence of illumination variations and camera/object motion that represented the main sources of misleading abrupt boundaries. For more illustration, Figure 2 shows various samples of frame sequences with disturbance factors that include illumination variations, fast object motion and camera motion, which are correctly detected as non-abrupt transitions. This indicates that the proposed algorithm had robustness for illumination changes and camera/object motion.

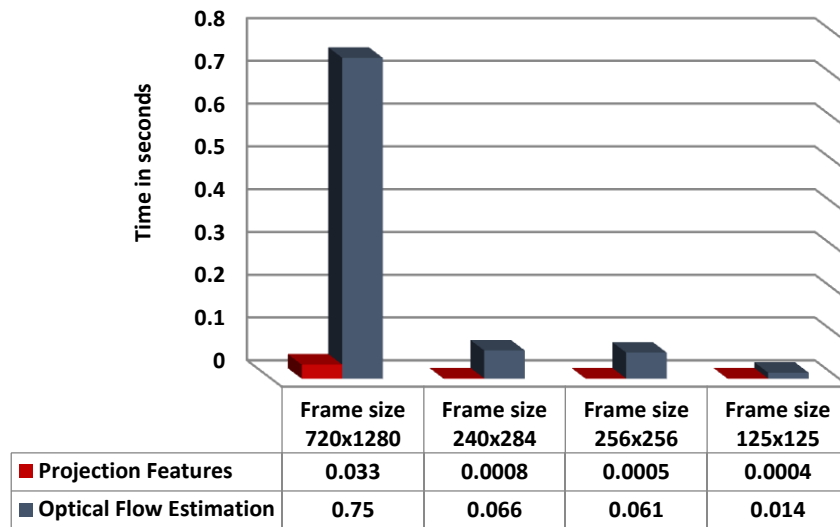


**Figure 2-** Examples of correct detection (a) Variations of illumination. (b) Motion of fast object. (c) Movement of fast camera.

As mentioned earlier, the video frames resolution was reduced to  $256 \times 256$ . The time required to compute optical flow field from original size frame is greater than the total time required for resizing frame and extracting optical flow field from lower resolution frames. In addition, the projection features is also faster to extract from lower resolution frames. A clear picture was given by visualizing the time required for optical flow estimation and projection features



extraction before and after resizing of the selected frames used as test materials, as illustrated in Figure 3.



**Figure 3-** The execution time comparison.

It is observed from the calculation of time that the 125×125 frame size consumes less time compared to the other frame sizes. In order to demonstrate the computational gain while maintaining efficiency by resizing frames, short videos (VF1 and VF9) were taken as illustrative examples. The proposed method was applied before and after resizing video frames. The results are summarized in Table 3. Although the execution time was reduced when the frame size was 125×125, the accuracy was minimal, as shown in Table 3 through the F values. Therefore the frame size was set to be 256×256 as the execution time was obviously reduced without affecting the performance of proposed method.

**Table 3-** Effects of resizing the video frame

Videos	Frame size 720x1280		Frame size 240x284		Frame size 256x256		Frame size 125x125	
	Time Cost in Sec.	F	Time Cost in Sec.	F	Time Cost in Sec.	F	Time Cost in Sec.	F
VF1	-	-	1.65	1	1.58	1	1.09	0.66
VF9	18.04	1	-	-	3.62	1	3.20	0.66

As explained above, an important aspect of effective shot detection methods is computational complexity. Detecting candidate transitions reduces the computational cost by avoiding processing all frames by optical flow estimation. Performing the proposed method without processing of candidate transitions would be really exhaustive. To offer a clear overall indication of the proposed method efficiency, the performance of the proposed method without candidate's transition detection was evaluated. Besides that, performance of the proposed method with candidate's transition detection only was evaluated, where only projection feature differences between frames were used. The performance evaluation according to R, P, F and time cost are reported in Table 4.

**Table 4-** Efficiency of the proposed method.

Videos	Candidates Transition Detection Only				Abrupt transition specifying Only			
	R	P	F	Time Cost in Sec.	R	P	F	Time Cost in Sec.
VF1	1.00	0.33	0.49	1.12	1.00	1.00	1.00	33.31
VF2	1.00	1.00	1.00	0.77	1.00	1.00	1.00	25.61
VF3	1.00	1.00	1.00	1.89	1.00	1.00	1.00	52.19
VF4	1.00	0.44	0.61	1.93	0.75	1.00	0.85	50.29
VF5	1.00	0.70	0.82	2.02	1.00	0.87	0.93	76.33
VF6	1.00	0.18	0.30	2.05	1.00	0.5	0.66	56.69
VF7	1.00	0.90	0.94	2.01	1.00	0.81	0.89	51.86
VF8	1.00	0.50	0.66	2.11	1.00	0.57	0.72	57.38
VF9	1.00	0.66	0.79	3.62	1.00	1.00	1.00	36.13
VF10	1.00	0.65	0.78	5.55	1.00	1.00	1.00	108.35
VF11	1.00	0.90	0.94	9.77	1.00	0.85	0.91	115.95
VF12	1.00	1.00	1.00	6.25	1.00	0.81	0.89	132.56
VF13	1.00	1.00	1.00	9.44	1.00	0.56	0.71	187.95
VF14	1.00	0.54	0.70	11.86	1.00	0.92	0.95	264.87
VF15	1.00	1.00	1.00	11.20	1.00	1.00	1.00	247.01
VF16	1.00	1.00	1.00	20.86	1.00	1.00	1.00	308.49
Average	1.00	0.72	0.814	5.77	0.98	0.86	0.90	112.81

With reference to Table 4, several important observations can be clarified. For example, the proposed method performance using the candidates transition detection only is superior, based on execution time only, to its performance using the abrupt transition specifying ; however, the performance decreases around 10% according to F measure. On the other hand, the proposed method performance took significant execution time only when using the abrupt transition specifying. This variation of time cost is mainly due to the difference in the time required to extract projection features and optical flow field. When comparing the results of Table 2 and Table 3, the two-stage integration of the proposed method is demonstrated to have the most efficient performance in terms of R, P, F and time cost.

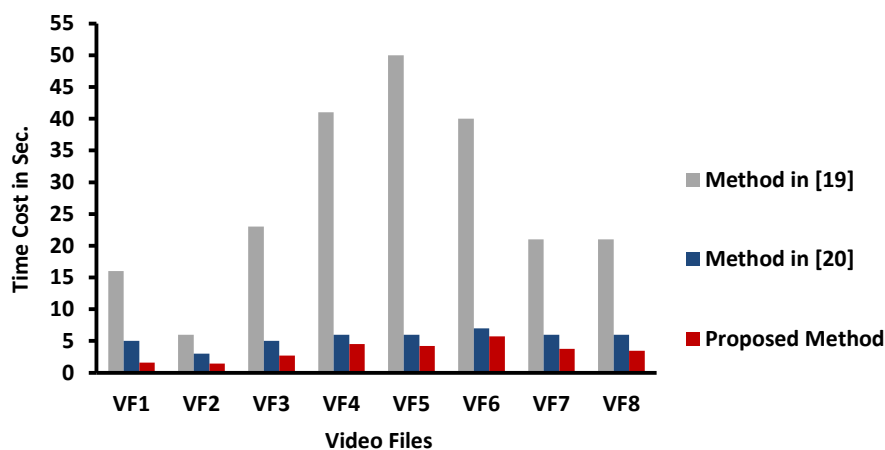
A comparison between the proposed method and other existing methods was performed to reinforce the capability of the proposed method. The comparison was made with methods presented elsewhere [19, 20]. Video files VF1 to VF8 were used for comparison in accordance a previously described method [20], where Dhiman *et al.* presented a quantitative comparison with another method [19] according to R, P, F and cost of time. This approach was exploited by our proposed method. The performance comparisons of the proposed method with the existing methods in terms of time cost, R, P and F are tabulated in Table 5 and Table 6. For better comparison, bar graph representations of the obtained results are also used, as illustrated in Figure 4 to Figure 7.

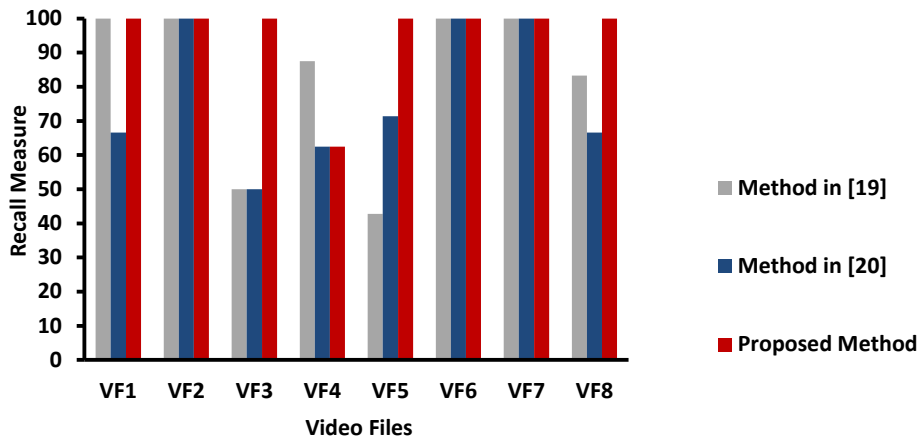
**Table 5-** Comparison of time cost of proposed method with other methods.

Videos	Time Cost in Sec.		
	Method in [19]	Method in [20]	Proposed Method
VF1	16	5	1.58
VF2	6	3	1.42
VF3	23	5	2.72
VF4	41	6	4.52
VF5	50	6	4.19
VF6	40	7	5.73
VF7	21	6	3.76
VF8	21	6	3.46

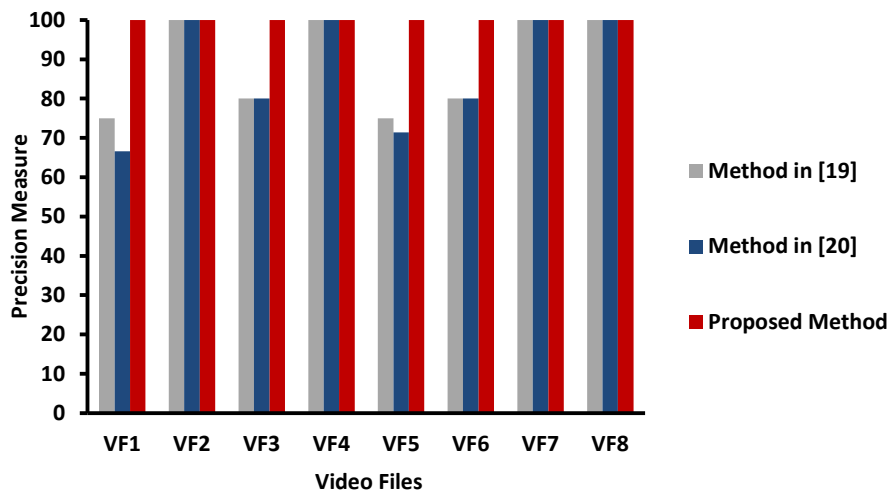
**Table 6-** Comparison of performance of proposed method with other methods.

Videos	Method in [19]			Method in [20]			Proposed Method		
	R	P	F	R	P	F	R	P	F
VF1	100	75	85.7	66.6	66.6	66.6	100	100	100
VF2	100	100	100	100	100	100	100	100	100
VF3	50	80	61.5	50	80	61.5	100	100	100
VF4	87.5	100	93.3	62.5	100	77	62.5	100	77
VF5	42.8	75	61.9	71.4	71.4	71.4	100	100	100
VF6	100	80	88.8	100	80	88.8	100	100	100
VF7	100	100	100	100	100	100	100	100	100
VF8	83.3	100	90.3	66.6	100	80	100	100	100

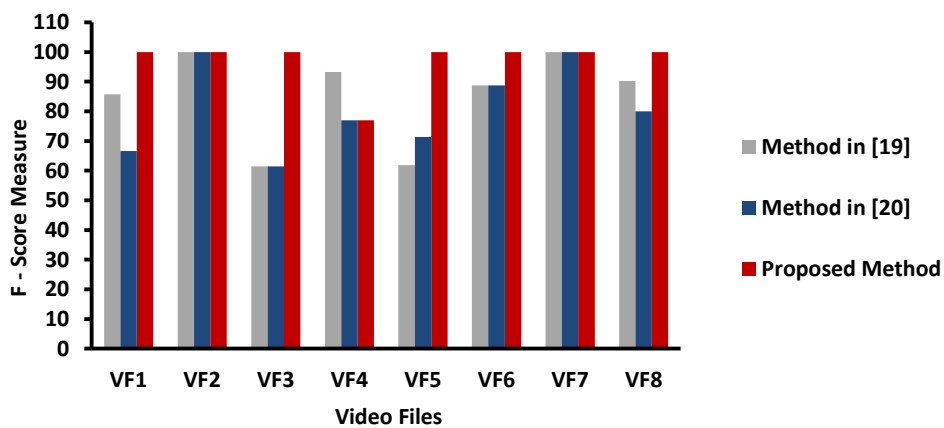
**Figure 4-** Comparison of time cost between the proposed method and the other methods.



**Figure 5-** Comparison of Recall measure between the proposed method and the other methods.



**Figure 6-** Comparison of Precision measure between the proposed method and the other methods.



**Figure 7-** Comparison of F-Score measure between the proposed method and the other methods.

It is noticeable that the proposed method had a lower execution time than the other methods and could still detect the abrupt shots at high performance. The results of the comparative analysis showed the superior performance of the proposed method versus the other current methods.

## 7. Conclusions

A new method is presented in this paper to handle the problem of balancing speed and accuracy of shot boundary detection. The proposed method is based on candidate transitions selection to discard most non-boundary frames before abrupt detection, helping to reduce the execution time and effectively detect abrupt shots. First, projection features is employed to select the candidate transitions that may contain abrupt boundaries. Then optical flow field is estimated for candidate transitions with their adjacent frames to filtrate the candidate transitions and determine the abrupt boundaries by recognizing the discontinuity of optical flow field. Experimental results showed the effectiveness of the proposed method, which could achieve excellent accuracy in accordance with Recall, Precision and F-measure and increase the detection speed obviously. The proposed method performance is satisfactory for recognizing transitions of abrupt shot in the presence of motion and illumination change of video sequences. Furthermore, the presented method outperforms the different latest methods.

## Acknowledgements

The work presented in this paper was supported by Al-Mustansiriyah University (www.uomustansiriyah.edu.iq), which is gratefully acknowledged.

## References

- [1] B.S. Rashmi, and H. S. Nagendraswamy, "Video shot boundary detection using block based cumulative approach," *Multimedia Tools and Applications*, vol. 80, no. 1 pp. 641-664, 2021.
- [2] S.H. Abdulhussain, A. R. Ramli, B. M. Mahmmud, M. I. Saripan, S. A. R. Al-Haddad and W. A. Jassim, "Shot Boundary Detection Based on Orthogonal Polynomial," *Multimedia Tools and Applications*, vol. 78, pp. 20361-20382, 2019. [online]. doi:10.1007/s11042-019-7364-3.
- [3] M. E. Abdulmunem, and E. Hato "Semantic Based Video Retrieval System: Survey," *Iraqi Journal of Science*, vol.59, no. 2A, pp. 739-753, 2018.
- [4] S. Chakraborty, and D.M. Thounaojam, "SBD-Duo: a dual stage shot boundary detection technique robust to motion and illumination effect," *Multimedia Tools and Applications*, vol. 80, no. 2, pp.3071-3087, 2021.
- [5] Singh, D. M. Thounaojam and S. Chakraborty, "A Novel Automatic Shot Boundary Detection Algorithm: Robust to Illumination and Motion Effect," *Signal, Image and Video Processing*, vol. 14, pp. 645-653, 2019. [online]. doi:10.1007/s11760-019-01593-3.
- [6] Javed, K. M. Malik, A. Irtaza and H. Malik, "A decision tree framework for shot classification of field sports videos," *The Journal of Supercomputing*, vol. 26, pp. 7242-7267, 2020. [online]. doi:10.1007/s11227-020-03155-8.
- [7] M. Haroon, J. Baber, I. Ullah, S. M. Daudpota, M. Bakhtyar and V. Devi, "Video Scene Detection Using Compact Bag of Visual Word Models," *Advances in Multimedia*, vol. 2018, no.2564963, pp. 1-9, 2018.
- [8] X. Liu, and J. Dai, "A Method of Video Shot-Boundary Detection Based on Grey Modeling for Histogram Sequence," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 9, no. 4, pp. 265-280, 2016.
- [9] H. Shao, Y. Qu and W. Cui, "Shot Boundary Detection Algorithm Based on HSV Histogram and HOG Feature," in *Proceedings of the 5th International Advanced Engineering Materials and Technology*, Atlantis press, pp. 951-957, 2015.
- [10] T. Kar, and P. Kanungo, "A Motion and Illumination Resilient Framework for Automatic Shot Boundary Detection," *Signal, Image and Video Processing*, vol. 11, no. 7, pp. 1237-1244, 2017.
- [11] D. M. Thounaojam, S. Roy and Kh. M. Singh, "Video Shot Boundary Detection using Gray Level Cooccurrence Matrix," *Indian Journal of Science and Technology*, vol. 9, no. 7, pp. 1-5, 2016.

- [12] G. Zajic, A. Gavrovska, I. Reljin and B. Reljin, "A Video Hard Cut Detection Using Multifractal Features," *Multimedia Tools and Applications*, vol.78 , no. 5, pp. 6233-6252, 2019.
- [13] C. S. e Santos, and H. Pedrini, "Shot Boundary Detection for Video Temporal Segmentation Based on the Weber Local Descriptor," in *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE*, pp. 1310-1315, 2017.
- [14] R. Hannane, A. Elboushaki, K. Afdel, P. Naghabhushan, M. Javed, "An Efficient Method for Video Shot Boundary Detection and Keyframe Extraction Using SIFT-Point Distribution Histogram," *International Journal of Multimedia Information Retrieval*, vol. 5, no. 2, pp. 89-104, 2016.
- [15] D. M. Thounaojam, T. Khelchandra, Kh. M. Singh and S. Roy, "A Genetic Algorithm and Fuzzy Logic Approach for Video Shot Boundary Detection," *Computational Intelligence and Neuroscience*, vol. 2016, no. 8469428, pp. 1-11, 2016.
- [16] J. Fan, S. Zhou and M. A. Siddique, "Fuzzy Color Distribution Chart -Based Shot Boundary Detection," *Multimedia Tools and Applications*, vol. 76, pp. 10169–10190, 2017. [online]. doi: 10.1007/s11042-016-3604-y.
- [17] J. Mondal, M.K. Kundu, S. Das and M. Chowdhury, "Video Shot Boundary Detection Using Multiscale Geometric Analysis of Nscst and Least Squares Support Vector Machine," *Multimedia Tools and Applications*, vol. 77, no.7, pp. 8139-8161, 2018.
- [18] J. Xu, L. Song, and R. Xie, "Shot Boundary Detection Using Convolutional Neural Networks," *Proceedings of Visual Communications and Image Processing (VCIP), IEEE*, pp. 1-4, 2016.
- [19] Z. M Lu, and Y. Shi, "Fast Video Shot Boundary Detection Based on SVD and Pattern Matching," *IEEE Transactions on Image processing*, vol. 22, no. 12, pp.5136-5145, 2013.
- [20] S. Dhiman, R. Chawla, S. Gupta, "A Novel Video Shot Boundary Detection Framework Employing DCT and Pattern Matching," *Multimedia Tools and Applications*, vol. 78, no.24, pp. 34707-34723, 2019.
- [21] E. Hato, and M. E. Abdulmunem, "Fast Algorithm for Video Shot Boundary Detection Using SURF Features," in *Proceedings of IEEE second Scientific Conference of Computer Sciences (SCCS), IEEE*, pp. 81-86, 2019.
- [22] W. Burger, and M. J. Burge, "*Digital Image processing: an Algorithmic Introduction Using Java*," 2<sup>nd</sup> ed, Springer, 2016, [E-book] Available: <https://www.springer.com/gp/book/9781447166832>.
- [23] Edison, and C. V. Jiji "Automated Video Analysis for Action Recognition Using Descriptors Derived From Optical Acceleration," *Signal, Image and Video Processing*, vol. 13, no.5, pp. 915-922, 2019.
- [24] R.J. Suji, S. S. Bhadouria, J. Dhar, and W. W. Godfrey, "Optical Flow Methods for Lung Nodule Segmentation on LIDC-IDRI Images," *Journal of Digital Imaging*, vol.33, no. 5, pp.1306-1324, 2020.
- [25] T.B. Thurayaa, S. M. Hameed and B. A. Attea, "A Decision Tree-Aware Genetic Algorithm for Botnet Detection," *Iraqi Journal of Science*, vol.62, no.7, pp.2454-2462, 2021.