

Shot boundary detection using perceptual and semantic information

Dalton Meitei Thounaojam¹ · Vivek Singh Bhadouria² · Sudipta Roy³ ·
Kh. Manglem Singh⁴

Received: 21 January 2017 / Revised: 4 March 2017 / Accepted: 14 March 2017 / Published online: 23 March 2017
© Springer-Verlag London 2017

Abstract The paper proposes a shot boundary detection system using Gist and local descriptor. Gist provides the perceptual and conceptual information of a scene. The proposed system can be roughly divided into three steps. The first step consists of forming of groups of frames by calculating the correlation of the Gist features between consecutive frames of the video. Secondly, abrupt transitions are found out using the group (G), MSER and a threshold (for abrupt separately, th_{cut}). And lastly, gradual transitions of the video are found using triangular pattern matching. We have performed the experiment on TRECVID 2001 and 2007 dataset. The novel contribution of this paper is that the proposed system shows an activity-based shot boundary detection where only the possible transition regions of a video are considered for shot detection. This approach reduces the computational complexity by processing the transition regions only. We have achieved better results in terms of $F1$, *precision* and *recall*, when compared to previously published approaches.

Keywords Shot boundary detection · Abrupt · Gradual · MSER · Gist

1 Introduction

Lately, the Internet has become the most widely used system for exchanging and sharing information. With the advancement in the field of communication engineering, enormous amount of data has been shared or transferred between two points connected through the Internet. Social networking Web sites and other business systems also lead to the generation of huge amounts of data, namely text, audio, video and other graphics media, which is difficult to store and process because of their huge size. Generally, video signals require larger amount of space for storage and constitutes a major part in the overall Internet traffic [25]. In order to gain insights into any video data, it is essential to understand the content of the video. Usually, computer-based automatic approaches are used to discover the content present in the video, for instance scene recognition and video enhancement for consumer devices, to name a few [26]. For analysis, a video is divided into segments called shots and this process is commonly known as temporal video segmentation or shot boundary detection (SBD) in the field of video processing.

The motivation of shot boundary detection is to find out the boundary or transitions of the shots in any video. The transitions can be: (1) abrupt transition where a sudden change in between the two consecutive frames occurs or (2) gradual transition where the boundary of the change extends for a certain length of frames. A gradual transition is an editing effect which can be further divided into fade-in, fade-out, dissolve and wipes [10]. Fade-in, fade-out and dissolve are caused by lighting effect, and wipes is mainly an animated effect. In fade-in, the shot slowly appears from a monochromatic or blank frame which is usually a black frame, [7]. In fade-out, the shot slowly disappears to a monochromatic or blank frame (usually black frame) [24]. In dissolve transition, the current shot gradually changes into the next shot in the video.

✉ Dalton Meitei Thounaojam
dalton.meitei@gmail.com; dalton@cse.nits.ac.in

¹ National Institute of Technology Silchar, Silchar, Assam, India

² National Institute of Technology Agartala, West Tripura, Tripura, India

³ Assam University Silchar, Silchar, Assam, India

⁴ National Institute of Technology Manipur, Imphal, India

Usually it is a combination of fade-in and fade-out simultaneously excluding the occurrence of the monochromatic or blank frames.

In this paper, a shot boundary detection using Gist and local descriptor is proposed in which Gist feature is extracted from each frame in a video and the difference between consecutive frames in the video is calculated using Pearson's correlation coefficient. Abrupt transitions are detected by using the Gist and MSER features of the selected frames, which is given in Algorithm 1. The gradual transitions are detected using the MSER feature and the proposed pattern matching method. The experimental result shows a good recall, precision and $F1$ score of the proposed system.

The novel contributions of the paper are as follows:

1. We propose a MSER pattern-based approach for the detection of shot boundaries.
2. We proposed an activity-based shot boundary detection. Activity here refers to those transition regions where there is a possibility of the presence of shot boundaries. This approach reduces the computational complexity by processing the transition regions only.
3. We propose a dual confirmation stage which reduces the false positives leading to an increased performance.

The paper is organized as follows: Sect. 2 gives a short survey about SBD. Section 3 explains the theoretical background of the feature extraction techniques used in the system. A detailed explanation of the proposed shot boundary detection system is given in Sect. 4. Section 5 contains the experimental result and discussion which is followed by conclusion in Sect. 6.

2 Related works

Researchers have carried out many works on the temporal video segmentation or SBD and it is monitored by TRECVID [22]. SBD can be approached by two methods (1) working in the compressed domain and (2) working in the uncompressed domain. Features such as discrete cosine transform (DCT) coefficients [11] and motion vectors [5, 16, 21] are used for SBD in compressed domain.

In an uncompressed domain, the features can be broadly classified into local feature and global features; in local feature-based approach, some local features are extracted from the frame which can represent or approximate the whole frame, thereby reducing the computational requirements. Information such as color content and pixel intensity is a global feature and have been used for SBD [8]. Color histogram [6, 14, 23, 27] and intensity histogram [28] are the most widely used global features for SBD in which histogram for each frame is computed and the difference/similarity

between the corresponding bins of the histogram is calculated for consecutive frame histogram. Instead of taking the histogram, the pixel information in either color or intensity is directly employed for SBD in which the similarity between the corresponding pixels is considered [17, 28]. In [9], a hybrid system using Gist and color histogram is proposed for SBD. A frame transition parameters combining global feature and local feature is proposed in [18] for SBD in which a neural network scheme is used to classify the type of transitions.

In [12, 13], a frame skipping technique is proposed for SBD where a preprocessing technique is adopted using an adaptive thresholding. A triangle pattern matching method is deployed to detect the gradual transitions. In [4], a SBD using foveation technique is proposed in which a local attention consistency measure is calculated to find out the transitions. Foveation is an image feature where image resolution is higher at some fixation point than the surrounding region of the image. Other image feature such as texture feature is used for analyzing the content of the image. In [3], energy, moment and contrast are used to detect SBD in which the texture features are calculated using co-occurrence matrix. In [2], a SBD is proposed using entropy and SURF feature in which entropy of a frame is calculated by using the intensity histogram. In [1], a video retrieval system is presented using the feature of maximally stable extremal regions (MSER).

3 Feature extraction

In this section, a brief description of the feature extraction techniques used in the proposed system is discussed as follows.

3.1 Gist

Oliva et al. [20] provided Gist of a scene as the perceptual and semantic information of a scene perceived by an observer irrespective of the details of the scene. Gist of a scene can be of perception information (referred to as *Perceptual Gist*) which contains the image properties such as the spatial frequency, color and texture and the conceptual information (referred to as *Conceptual Gist*) which contains the semantic information of the scene. A detailed description of the Gist can be found in [20].

It represents the “Gist of a scene” by using a set of perceptual properties such as naturalness, openness, roughness, ruggedness and expansion, which are related to the shape of the scene space excluding the information of the local object and provides a holistic description of the real-world scenes which is termed as *Spatial Envelope*. As only the background/scene information is obtained in Gist feature, it

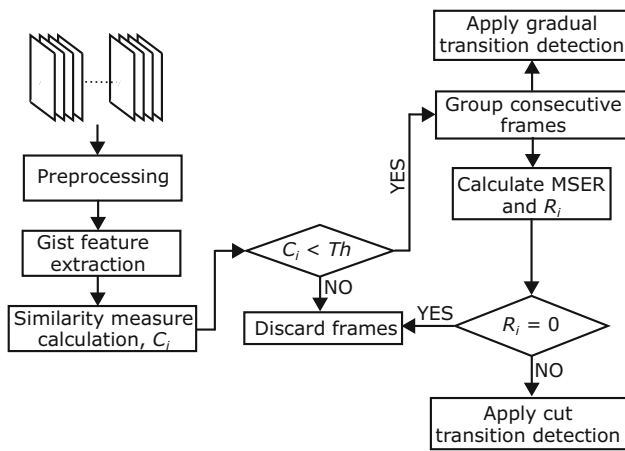


Fig. 1 Block diagram for proposed system

is convenient for reducing the risk against object motions which increases the fault in transition detection.

3.2 Maximally stable extremal region

Maximally stable extremal region (MSER) was proposed by Matas [15]. MSER is a widely used approach for finding the correspondence between the two images. Such maximally stable regions are determined by varying a threshold T (as given in Eq. 1) and then observing the corresponding variation in the size of the local regions in the thresholded image.

$$0 < T < I \tag{1}$$

where

$$I = \max(P) \tag{2}$$

and P is the pixel values in the image.

A region is considered to be maximally stable if the variation of T brings a change C , ($C < S$) where S is the maximum allowed variation. For a detailed description, readers are suggested to refer the work in [15].

4 Proposed system

This section discusses about the proposed shot boundary detection system. A figure illustrating the flow of the proposed approach is shown in Fig. 1.

The proposed system (see Fig. 1) starts with a preprocessing which includes:

1. Converting color image to gray level image.
2. Resizing each frame to $S \times S$, where $S = 256$ in our experiment.

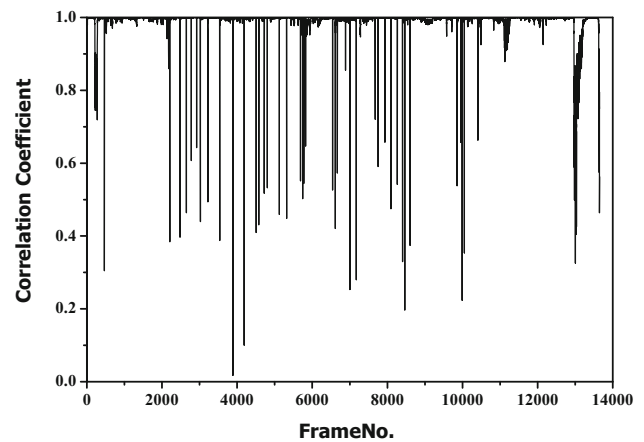


Fig. 2 Correlation coefficient for the video D6.mpg

The features of the preprocessed frames are extracted using Gist. Gist of an image gives the *Scene of the image*, so Gist is scale-invariant feature, thereby resizing the image will not change the Gist features. The perceptual and conceptual information can be extracted without considering the object present in the image [19] which reduces the risk of object motion effect in our proposed system. After feature extraction, the correlation between the consecutive frames are computed using Pearson correlation coefficient (C_i). Pearson correlation coefficient gives a value 1 if the consecutive frames are identical and it decreases to 0 as the dissimilarity between frames increases, as shown in Fig. 2.

A threshold Th , is applied to classify the possible transition frames and non-transition frames. The threshold (Th) is taken through careful observation and experimentation in different videos from TRECVID database. Empirically, we have found the optimal range of Th to be [0.95, 0.99]. For the remaining of our discussion, we have set $Th = 0.99$. To classify the image into transition and non-transition frame, we followed Eq. 3.

$$\text{Frame} = \begin{cases} \text{transition} & \text{if } C_i < Th \\ \text{non-transition} & \text{otherwise} \end{cases} \tag{3}$$

where C_i is the correlation coefficient between i th and $i + 1$ th frames.

After determining the similarity between the frames, the algorithm groups the consecutive frames such that one transition effect is captured in one group. It is to be noted here that every group will not contain gradual or cut transition; there is also a possibility of motion effect which may lead to a group. Therefore, it is essential to perform a post-processing operation on the groups in order to determine the correct nature of the transition. It is also taken in account the grouping of frames, which are not consecutive but missed by one or two frame(s). As an arbitrary example, consider frames 14

and 16 are present, but frame 15 is missing after the above processes; then, a group of (14, 15, 16) is included in the grouping process. This step is helpful in increasing the detection of possible transition frames. Each group contains the possible transitions of the whole video.

Separately, one of the main problems in the system is the blank frames, which behaves as an abrupt transition. So, for each frame, which satisfies Eq. 3, MSER is calculated and the number of MSER regions is recorded in R_i . The frames which satisfy Eq. 4 are discarded.

$$R_i = 0 \quad (4)$$

Above-explained steps ensure to remove the unwanted information from the video, thereby reducing any chance of false detection. After this we have clear groups of frames where we can apply abrupt and gradual transition detection algorithm that are described in the following text. It is to be noted that the blank frame is detected by using MSER count and it is only discarded in the abrupt transition detection only as it behaves as an abrupt transition. Blank frames are detected by using MSER feature. If the number of MSER regions extracted from a frame is 0 (zero), then the frame is discarded as blank frame. But for the gradual transition, the blank frames play a vital role in detecting the fade effect which is a part of gradual transition.

4.1 Abrupt transition detection

For abrupt transition detection, another threshold, Th_{cut} is considered. The selection of Th_{cut} is done in such a way that all the spikes as shown in Fig. 1 are included. The spikes are the possible abrupt transitions. Algorithm 1 gives the process of detecting abrupt transitions and the efficiency of it depends on the careful and proper selection of the Th_{cut} .

Algorithm 1 is applied to individual group G_i , where i is the number of groups. Logically, the frames whose correlation value are less than the threshold Th is completely omitted from the observation of the transition detection. The blank frames in a video give more error in the abrupt detection system. So, removing the blank frames prior in detection system increases the precision and $F1$ score (see Sect. 5).

4.2 Gradual transition detection

4.2.1 Analysis of MSER count in a video

Before applying the gradual transition algorithm, the TRECVID video test datasets are analyzed by using the MSER features. In this, the MSER features from a video frame is extracted and the number of MSER regions is counted and recorded (hereafter it is mentioned as *MSER count*). This step is repeated for all the remaining frames.

Algorithm 1 Abrupt transition detection

Input: group of frames, G_i where i is the number of group

Output: cut/abrupt transition frames.

```

1: procedure ABRUPT_TRANSITION( $G_i, C$ )           ▷ where  $C$  is the
   correlation value.
2:   for  $x := 1$  to length of  $G_i$  do
3:      $N \leftarrow G_i(x)$ 
4:      $I \leftarrow$  gray value of frame  $f_N$ 
5:      $M \leftarrow$  extract MSER feature of  $I$ 
6:      $R \leftarrow$  number of MSER regions in  $M$ 
7:     if  $R == 0$  then                               ▷ to reduce the false detection
8:       Discard the frame  $f_x$ 
9:     else
10:      if  $\{C_N < Th_{cut}\} \& \{C_{N-1} > Th_{cut}\} \& \{C_{N+1} > Th_{cut}\}$ 
   then
11:        Record  $N$  for cut transition
12:      end if
13:    end if
14:  end for
15:  return  $N$ 
16: end procedure

```

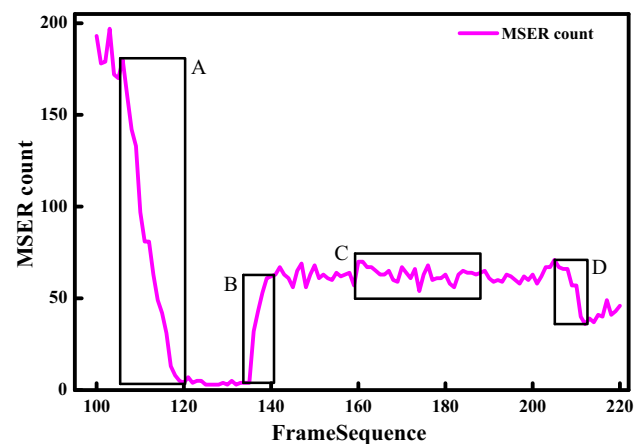
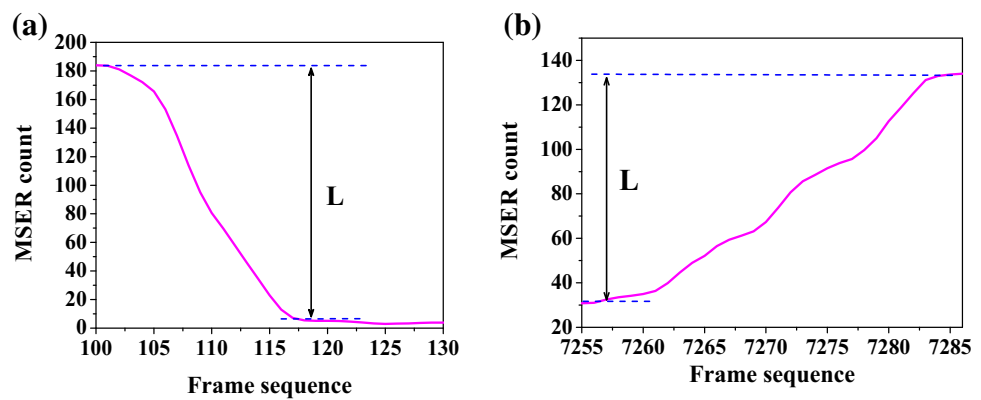


Fig. 3 MSER count value for the video *BG_37770.mpg* showing fade-out (A), fade-in (B), no transition (C) and dissolve (D)

Then, the transition part is analyzed by plotting the MSER count as shown in Fig. 3.

It is observed that the behavior for fade-in, fade-out and dissolve transitions are almost similar and can be represented by a single pattern (clearly a slope). It is also found after an analysis of the pattern that the fade-in, fade-out and dissolve transitions can be differentiated from each other in such a way that in fade-out and fade-in, the MSER count of the last frame (in case of fade-out) and first frame (in case of fade-in) in the transition always near to 0 (zero) level (see Fig. 3 A, B), whereas in dissolve, the MSER count either of the starting frame or the ending frame of the transition never approaches to zero level. The wipe transition is not considered in the proposed system. From the behavior analysis of the gradual transition frames using the TRECVID datasets, it is concluded that the gradual transitions share a common behavior.

Fig. 4 Gradual transitions **a** decreasing and **b** increasing patterns detected from the MSER count (L is the length of the gradual transitions)



4.2.2 Gradual detection approach

In this section, we will discuss the approach to determine the gradual transitions in a video sequence. Determining gradual transitions is relatively complex when compared to finding the cut transitions. This complexity arises because of the slow varying image features or characteristics spread over a number of frames. However, in a cut transition, the changes are abrupt and can be trivially determined by monitoring an appropriate metric between two frames.

A gradual transition can be seen as slow sweeping of a scene, S_i by another scene S_{i+1} . Such transition usually takes place over many frames. Therefore, it is essential to observe multiple frames in order to determine the gradual transition. In the present work, we have used MSER feature as a discriminating feature to find out the inter-frame transition.

A graphical visualization of MSER variation for a gradual transition and normal scene is illustrated in Fig. 3. Figure 3 shows that MSER varies in an increasing or a decreasing pattern for a gradual transition unlike for a normal scene variation. Such transition behavior has been observed for 300 gradual transitions (from TRECvid datasets). Therefore, we are proposing in this paper that if we observe increasing or decreasing pattern (i.e., region A and B in Fig. 3) in any video then we can conclude it to be a gradual transition.

Before declaring a gradual transition, all the detected gradual transitions must satisfy Eq. 5.

$$\text{match_feature}(f_{L_s-n}, f_{L_e+n}) \leq Th_{\text{verification}} \tag{5}$$

where f_{L_s-n} and f_{L_e+n} are the MSER feature of the previous n th frame of the starting and preceding n th frame of the end of the possible gradual transition of length L (shown in Fig. 4), respectively. L_s and L_e are the starting and ending of the detected gradual transition. $Th_{\text{verification}}$ is a threshold use for the verification of the gradual transition.

Equation 5 is used in our proposed system in order to reduce the number of false detections of the gradual tran-

Table 1 Description of the TRECvid 2001 and 2007 video dataset used for SBD

Video	Frame no.	Transition			Video sources
		Cut	Gradual	Total	
D2	16586	42	31	73	TRECvid 2001
D3	12304	39	64	103	
D4	31389	98	55	153	
D6	13648	40	45	85	
BG_3027	49813	127	1	128	TRECvid 2007
BG_3097	44987	91	0	91	
BG_3314	35800	42	0	42	
BG_16336	2462	20	0	20	
BG_28476	23236	176	2	178	
BG_36136	29426	88	12	100	
BG_37309	9639	11	8	19	
BG_37770	15836	8	29	37	

sition. If any gradual transition does not satisfy Eq. 5, then the possible gradual transition is not considered as gradual transition and is discarded.

5 Experimental results

For experimentation, TRECvid 2001 and 2007 video test dataset has been considered for shot boundary detection and the description of the video with the number of abrupt and gradual transitions (excluding the wipe transition) is given in Table 1. TRECvid 2001 video test data are openly available for the research and it can be downloaded from the *Open Video Project*. TRECvid 2007 video test data are obtained from *Netherlands Institute for Sound and Vision* under request for research.

The performance evaluation of the proposed system is done by using recall, precision and $F1$ score as in [13] which is calculated using Eqs. 6–8, respectively.

Table 2 $F1$ score of the abrupt transition using different Th_{cut} value in TRECVID 2001 video test dataset

Video	Abrupt threshold (Th_{cut})				
	0.95	0.90	0.85	0.80	0.75
D2	0.571	0.654	0.764	0.919	0.833
D3	0.505	0.638	0.767	0.809	0.836
D4	0.561	0.763	0.834	0.907	0.912
D6	0.850	0.963	0.987	0.964	0.987
Average	0.622	0.754	0.838	0.899	0.892

$$\text{Recall} = \frac{N_C}{N_C + N_M} \quad (6)$$

$$\text{Precision} = \frac{N_C}{N_C + N_F} \quad (7)$$

$$F1 \text{ score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (8)$$

where N_C is the number of correct transitions detected, N_M is the number of missed transitions and N_F is number of false transitions detected.

In our experiment, the threshold for cut Th_{cut} is selected by observing the graph of the correlation coefficient value (example shown in Fig. 2). Different Th_{cut} values have been selected ranging from 0.95 to 0.75, and the FI score of the abrupt transitions for TRECVID 2001 video test data is determined for testing the efficient value or range of Th_{cut} in Table 2.

Table 2 shows that the Th_{cut} value at 0.8 yields an overall good $F1$ score and this Th_{cut} value is used for our experiment.

For gradual transition, the increasing or decreasing pattern generated using local descriptor is observed from the group we have extracted using Gist. The formation of the group helps us to exclude most of the non-transitions frames from the processing of the gradual transition, thereby reducing the processing time. The verification threshold $Th_{verification}$ in Eq. 5 is found out experimentally. In our experiment, 4 is taken as the $Th_{verification}$ value and two frame, i.e., $n = 2$, previous and preceding of the starting and ending of the gradual transition of length L is taken. Some of the sample patterns observed for the consideration of gradual transition is shown in Fig. 4.

It is observed from our experiment that in the TRECVID videos, the length of the gradual transitions usually ranges from 6 to 32 frames as in [13]. In some cases, two or three gradual transitions exist in one group and subsequently the number of frames in that group is also more (eg. 200 frames in a group). So in this case, we have apply our technique for gradual transition for the range of 5–31 frames at a time in a group. Some examples of the detected gradual transitions are shown in Fig. 5.

In Table 3, the proposed system performance using TRECVID 2007 test data is shown. The proposed system is compared with SBD using color feature [27], SBD using SVD and pattern matching [13] and SBD using GA and fuzzy Logic [23] as shown in Tables 4, 5 and 6, respectively. Tables 4, 5 and 6 show that the proposed approach yields

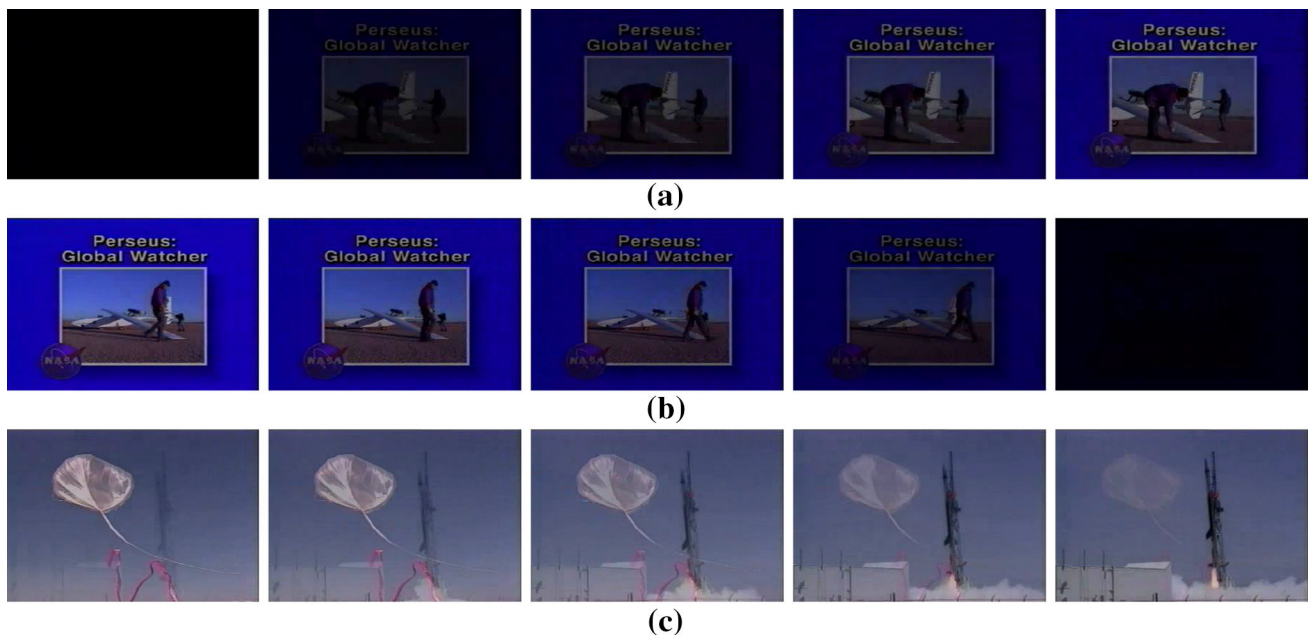
**Fig. 5** An illustration of **a** Fade-in **b** Fade-out and **c** gradual transitions detected from the video “D6.mpg”

Table 3 Proposed system results for TRECVID 2007 video test dataset

Video	Proposed system					
	Cut			Gradual		
	Recall	Pre	F1	Recall	Pre	F1
BG_3027	0.945	0.902	0.923	1.000	0.500	0.667
BG_3097	0.868	0.987	0.924	–	–	–
BG_3314	0.786	0.943	0.857	–	–	–
BG_16336	0.950	1.000	0.974	–	–	–
BG_28476	0.977	0.955	0.966	1.000	0.400	0.571
BG_36136	0.977	0.977	0.977	0.667	0.727	0.696
BG_37309	1.000	0.846	0.917	0.750	0.667	0.706
BG_37770	1.000	0.889	0.941	0.931	0.818	0.871
Average	0.938	0.937	0.935	0.869	0.622	0.702

Table 4 Comparison of the SBD using color feature [27] with the proposed system

Videos	SBD using color feature [27]			Proposed system		
	Recall	Pre	F1 score	Recall	Pre	F1 score
D2	0.928	0.951	0.939	0.952	0.889	0.919
D3	0.821	0.864	0.842	0.923	0.720	0.809
D4	0.826	0.900	0.861	0.949	0.869	0.907
D6	0.925	0.973	0.948	1.000	0.930	0.964
Average	0.875	0.947	0.898	0.956	0.852	0.900

Table 5 Comparisons between proposed system with SVD and pattern matching [13]

Video	SVD and pattern matching [13]						Proposed system					
	Cut			Gradual			Cut			Gradual		
	R	P	F1	R	P	F1	R	P	F1	R	P	F1
D2	0.905	0.905	0.905	0.935	0.725	0.817	0.952	0.889	0.919	0.870	0.794	0.830
D3	0.667	0.867	0.754	0.734	0.940	0.824	0.923	0.720	0.809	0.812	0.867	0.838
D4	0.888	0.897	0.892	0.727	0.741	0.734	0.949	0.869	0.907	0.836	0.676	0.747
D6	0.950	0.974	0.962	0.844	0.927	0.884	1.000	0.930	0.964	0.867	0.907	0.887
Average	0.853	0.911	0.878	0.810	0.833	0.815	0.956	0.852	0.900	0.846	0.811	0.825

Table 6 Comparison of the SBD using GA and fuzzy logic [23] with the proposed system

Video	GA and fuzzy logic [23]						Proposed system					
	Cut			Gradual			Cut			Gradual		
	R	P	F1	R	P	F1	R	P	F1	R	P	F1
D2	0.952	0.889	0.919	0.806	0.833	0.819	0.952	0.889	0.919	0.870	0.794	0.830
D3	0.846	0.805	0.825	0.764	0.942	0.844	0.923	0.720	0.809	0.812	0.867	0.838
D4	0.878	0.935	0.906	0.727	0.816	0.769	0.949	0.869	0.907	0.836	0.676	0.747
D6	1.000	0.889	0.941	0.844	0.864	0.854	1.000	0.930	0.964	0.867	0.907	0.887
Average	0.919	0.879	0.896	0.785	0.864	0.821	0.956	0.852	0.900	0.846	0.811	0.825

better *F1* score for cut and gradual transition, compared to [13,23,27], hence proving the superiority of the proposed approach. In Tables 4, 5 and 6, *R*, *P* and *F1* represents *recall*, *precision* and *F1* score, respectively. In [9], a shot boundary detection system using Gist and color histogram is proposed with an overall approximate performance of 84 and 91% of precision and recall, respectively. However, our proposed system yields an overall approximate performance of 89 and 94% of precision and recall, respectively.

6 Conclusion

The paper proposed a shot boundary detection algorithm using Gist and MSER. In this, Gist is extracted from each frame in a video and the similarity between consecutive frames is calculated using Pearson’s correlation coefficient. A threshold is applied, which selects the frames whose correlation is less than the threshold and these frames have a high chance of a transition (either abrupt or gradual) to occur. Consecutive frames are taken to form groups and each group is observed for the occurrence of transitions. For abrupt transition, Gist and a cut threshold are used and the local descriptor is used to reduce the false detection. For gradual transition, a pattern matching approach is proposed where increasing or

decreasing pattern of MSER count confirms the existence of gradual transition. The novel contribution of this paper is the most of the non-transition frames are skipped in the further processing for detecting the transitions. The proposed system yields a better result as shown in Tables 3, 4, 5 and 6.

Acknowledgements Sound and Vision video is copyrighted. The Sound and Vision video used in this work is provided solely for research purposes through the TREC Video Information Retrieval Evaluation Project Collection.

References

- Anjulan A, Canagarajah N (2007) Object based video retrieval with local region tracking. *Signal Process Image Commun* 22(7–8):607–621
- Baber J, Afzulpurkar N, Dailey M, Bakhtyar M (2011) Shot boundary detection from videos using entropy and local descriptor. In: 17th International conference on digital signal processing (DSP), pp 1–6
- Bhowmick B, Chattopadhyay D (2009) Shot boundary detection using texture feature based on co-occurrence matrices. In: International multimedia, signal processing and communication technologies, pp 165–168
- Boccignone G, Chianese A, Moscato V, Picariello A (2005) Foveated shot detection for video segmentation. *IEEE Trans Circuits Syst Video Technol* 15(3):365–377
- Du Q, Zheng H, Zhang S (2009) Motion estimation of projected spatio-temporal slice for shot boundary detection. In: 7th International conference on information, communications and signal processing, pp 1–5
- Hameed A (2009) A novel framework of shot boundary detection for uncompressed videos. In: International conference on emerging technologies, pp 274–279
- Hu W, Xie N, Li L, Zeng X, Maybank S (2011) A survey on visual content-based video indexing and retrieval. *IEEE Trans Syst Man Cybern Part C Appl Rev* 41(6):797–819
- Jadon R, Chaudhury S, Biswas K (2001) A fuzzy theoretic approach for video segmentation using syntactic features. *Pattern Recognit Lett* 22(13):1359–1369
- Jhuang H, Chikkeru S (2006) Video shot boundary detection using gist
- Koprinska I, Carrato S (2001) Temporal video segmentation: a survey. *Signal Process Image Commun* 16(5):477–500
- Lee S-W, Kim Y-M, Choi SW (2000) Fast scene change detection using direct feature extraction from mpeg compressed videos. *IEEE Trans Multimed* 2(4):240–254
- Li Y-N, Lu Z-M, Niu X-M (2009) Fast video shot boundary detection framework employing pre-processing techniques. *IET Image Process* 3(3):121–134
- Lu Z-M, Shi Y (2013) Fast video shot boundary detection based on svd and pattern matching. *IEEE Trans Image Process* 22(12):5136–5145
- Mas J, Fernandez G (2003) Video shot boundary detection based on color histogram. In: Notebook Papers TRECVID2003
- Matas J, Chum O, Urban M, Pajdla T (2004) Robust wide-baseline stereo from maximally stable extremal regions. *Image Vis Comput* 22(10):761–767 *British Machine Vision Computing*
- Meng J, Juan Y, Chang S-F (1995) Scene change detection in a mpeg compressed video sequence. In: Proceedings of the SPIE in digital video compression: algorithms and technologies 2419:14–25
- Luo M, DeMenthon D, Doermann D (2004) Shot boundary detection using pixel-to-neighbour image differences in video. In: TRECVID 2004Workshop Notebook Papers
- Mohanta P, Saha S, Chanda B (2012) A model-based shot boundary detection technique using frame transition parameters. *IEEE Trans Multimed* 14(1):223–233
- Oliva A (2005) Gist of the scene. In: Itti L, Rees G, Tsotsos JK (eds) *The encyclopedia of neurobiology of attention*. Elsevier, Amsterdam, pp 251–256
- Oliva A, Torralba A (2001) Modeling the shape of the scene: a holistic representation of the spatial envelope. *Int J Comput Vis* 42(3):145–175
- Panchal P, Merchant S, Patel N (2012) Scene detection and retrieval of video using motion vector and occurrence rate of shot boundaries. In: 2012 Nirma University International conference on engineering (NUiCONE), pp 1–6
- Smeaton AF, Over P, Doherty AR (2010) Video shot boundary detection: seven years of TRECVID activity. *Comput Vis Image Underst* 114(4):411–418
- Thounaojam DM, Khelchandra T, Singh KM, Roy S (2016) A genetic algorithm and fuzzy logic approach for video shot boundary detection. *Computational Intelligence and Neuroscience*, vol 2016, Article ID 8469428
- Thounaojam DM, Trivedi A, Manglem Singh K, Roy S (2014) A survey on video segmentation. In: Mohapatra DP, Patnaik S (eds) *Intelligent computing, networking, and informatics*, vol 243., *Advances in intelligent systems and computing* Springer, India, pp 903–912
- Wen Y, Li Z, Chen J, Zhao H (2013) A prediction algorithm for real-time video traffic based on wavelet packet. In: Su J, Zhao B, Sun Z, Wang X, Wang F, Xu K (eds) *Frontiers in internet technologies*, vol 401., *Communications in computer and information science* Springer, Berlin, pp 1–8
- Yunbo R, Leiting C (2012) A survey of video enhancement techniques. *J Inf Hiding Multimed Signal Process* 3(1):71–99
- Zhang H, Hu R, Song L (2011) A shot boundary detection method based on color feature. In: International conference on computer science and network technology (ICCSNT), 2011, vol 4, pp 2541–2544
- Zhang H, Kankanhalli A, Smoliar S (1993) Automatic partitioning of full-motion video. *Multimed Syst* 1(1):10–28