Abrupt Scene Change Detection Using Spatiotemporal Regularity of Video Cube

Rupesh Kumar, Sonali Ray, Meenakshi Sharma and Basant Kumar

Abstract In this paper, we propose the detection method of abrupt scene change using spatial as well as spatiotemporal frames of video cube. Most of the methods use either intensity or motion of pixels for the scene change detection methodology. Unlike to the existing methods, both the intensity and flow vector of video frames are used simultaneously in this paper to propose a general abrupt scene change detection method. For a spatial frame, flow energy function is used for detection. Flow energy function, defined by the spatiotemporal regularity flow model, is the combinatorial form of intensity and flow vectors of the frames. In the spatio-temporal frames, abrupt scene change appears as a vertical line which is detected by the edge detection method. Combined results of spatial and the spatio-temporal frames provide the location of scene change. The proposed method detects almost all the locations of scene change with negligible false detection.

Keywords Regularity flow · Video cube · Flow vectors · Spatiotemporal · Boundary · Edge detection · Abrupt change

1 Introduction

To retrieve the semantic content in a video is a cumbersome task owing to the large size of video data. The retrieval problem in the video can be solved through indexing of video frames at the cost of high processing cost. Generally, scene change detection [2, 4, 6, 8–10] is a method of content detection for the application of annotation, scene analysis, fast searching, and indexing. The other use of scene change detection is in video compression where it is used to estimate the key frame. Manual indexing and annotation of large multimedia data are time-consuming tasks that encourage

R. Kumar (✉)
IIT, Kanpur, India
e-mail: rupeshkr@iitk.ac.in

S. Ray · M. Sharma · B. Kumar
MNNIT, Allahabad, India

D. Dutta et al. (eds.), *Advances in VLSI, Communication, and Signal Processing*,
Lecture Notes in Electrical Engineering 587,
https://doi.org/10.1007/978-981-32-9775-3_88

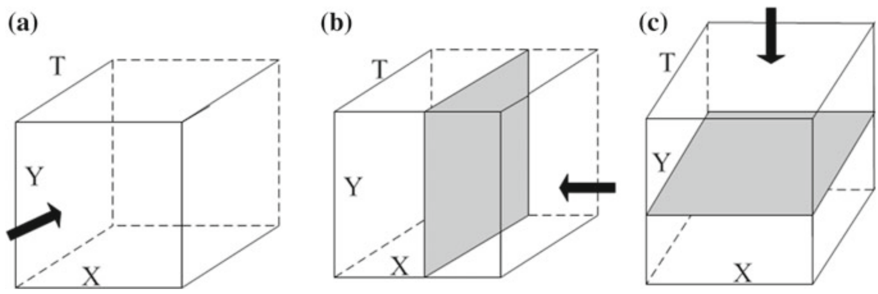


Fig. 1 Video cube. **a** V_tXY, **b** V_xTY, **c** V_yTX

the researcher to make an automatic scene change detection algorithm. Location of scene change in a video sequence appears as a boundary [13], because continuous scene appears as a flow of continuous action that depends on the foreground and background contents. Background and foreground of continuous scene possesses similar contents that make a shot [7, 12] or content of similarity. Scene change appears at the boundary of two shots or in between two scenes. Therefore, for the content retrieval purpose, the video is organized into a groups of shots. The aim of the scene change detection method is to partition the video sequences into the meaningful and manageable segments (shots) [10] for video indexing. The key frame is extracted from each segment of the scene that represents spatial and temporal features of that scene segment. A scene change [3] in a video stream can also be explained as a change of feature points or change of pixel intensity between two consecutive frames up to a remarkable limit. The term limit is interpreted as Thresholding [5, 14] which is widely used for the detection of scene change. Threshold may be fixed or dynamic in nature and the value of dynamic threshold [11, 18] is always updated according to the content of scene segment. Similarity measure between two consecutive frames is the basic idea of scene change detection technique and most of the prior work uses such methodology.

This paper is organized as follows. Section II explains the proposed scene change detection method. Experimental results are given in Section III, and Section IV concludes the paper. Notation of upper case capital X, Y, T is used for the respective axes and lower case x, y, t are for flow direction in the whole paper. V_tXY represents video cube with XY frame along t direction (Fig. 1). Similarly, V_xTY and V_yTX represent video cubes with frames TY and TX along x- and y- directions.

2 Proposed Detection Method

Proposed scene change detection method uses spatial frame XY. The novelty of the proposed techniques is to consider the video as a cube and process the entire cube simultaneously. Unlike the other existing techniques, the proposed method does not use frame-by-frame processing. The proposed detection method is a hybrid approach [16] which incorporates both the spatial frame and spatiotemporal frames (*TY* and *TX*). In the spatial frame, SPREF [1]-based frame energy is used to detect the abrupt scene change and the obtained result is fused with the method reported in [15] that uses both the spatiotemporal frames. In the next section, the proposed SPREF-based detection method has been explained.

2.1 SPREF-Based Detection Method

SPREF (spatiotemporal regularity flow) [1] is a general framework to model the video. Assumption of the video as a cube is one of the advantages of this model. SPREF (Spatiotemporal Regularity Flow) is a 3D vector field and it proposes a regular flow direction as a path in which the intensity of the pixel varies the least. If the scene is continuous, then intensity as well as flow vectors of frames vary regularly, but on the other hand, they show large deviation at the location of abrupt scene change. Using the SPREF model, we detect the deviation at the boundary of scene change with the help of flow energy function. In this paper, the translational-SPREF model is used and the flow energy is defined as

$$E\,(t) = \sum_{\Omega} |\left(I \star \frac{\partial H}{\partial x}\right) c_1'(t) + \left(I \star \frac{\partial H}{\partial y}\right) c_2'(t) + I \star \frac{\partial H}{\partial t}|^2 \tag{1}$$

where $[c_1'(t), c_2'(t)]$ are the flow vector components in x- and y directions of the video frame XY with flow direction t. H is defined as a Gaussian filter and intensity of the image is I. Temporal size of the video cube is Ω and term c is used for translational flow. Flow energy function (Eq. 1) is solved by using translated box spline functions $b(u)$ of the first degree. Due to smoothness of spline, it approximates the regular flow direction by minimizing the flow energy function. Flow vectors in terms of spline coefficients are explained as

$$c_m'(u) = \sum_{n} \alpha_n^m b(2^{-l}\, u - n) \tag{2}$$

where $m \in (1, 2)$ and $u \in (t)$. Term α_n is the nth spline coefficient. Length of temporal axis of video cube region Ω is = 2^k. Scaling factor of video cube is taken as l and its value has been taken as $l = 1, 2, \ldots, k$. Value of n is defined as $n = 2^{k-l}$. The spline function used here is defined as

$$b(z) = \begin{cases} 1 - |z| & \text{if } |z| < 1 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Since flow energy function is the combinatorial form of intensity and the flow vectors of the frame, it estimates the regularity of frame efficiently. If no scene change occurs in a sequence of frames, then all the frames are regular on the basis of their frame energy. Flow energy of video defined by SPREF model combines both the features and therefore, it models the regularity of frame contents effectively than either of the pixel or flow vectors of the frame. Abrupt scene change in *XY* frames creates large deviation in their flow energy and the location of deviation is by the proposed threshold value:

$$Threshold = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(E_t - \mu)} \tag{4}$$

$$where \quad \mu = \frac{1}{N}\sum_{t=1}^{N} E_t \tag{5}$$

where E_t represents the flow energy of tth frame and N is the total number of frames. Abrupt scene change is detected with the help of the following condition:

$$E_{detected} = \begin{cases} 1 & \text{if } E_t >= 5 \times Threshold \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

$E_{detected}$ gives only the location of abrupt scene change with value 1. It has been investigated that all the peaks of abrupt scene change are greater than the mean value of the flow energy. How much the maximum peak value of flow energy is deviated from the mean value is defined by the standard deviation and this is the reason to select standard deviation for thresholding. The next section explains the detection approach by using spatiotemporal frames.

2.2 Boundary Detection in Spatiotemporal Frames

Apart from the use of flow energy for scene change, spatiotemporal frames are also considered for the detection of abrupt scene change which has been proposed in [15]. In this method, both spatiotemporal frames *TY* and *TX* are considered. Abrupt scene change produces large pixel intensity variation between two *XY* frames, but in spatiotemporal frames such variation appears as a vertical line or vertical boundary. The aim of this work is to detects the location of scene change as a vertical line and to combine the result with the results obtained from the previous section. Spatiotemporal frame-based abrupt scene change detection method [15] is summarized as

– Select four sampled TY and TX frames:

$$S_{TY} = [TY_{s_1}, TY_{s_2}, TY_{s_3}, TY_{s_4}]$$
$$S_{TX} = [TX_{s_1}, TX_{s_2}, TX_{s_3}, TX_{s_4}] \quad (7)$$

where s_1 is the first frame and sampled interval for other three frames is taken as

$$s' = \lfloor N/4.5 \rfloor \quad (8)$$

N is the total number of frames along the flow direction.
– Canny edge detection method is used to detect the edges of all the sampled frames and it produces binary images. Binary image of sampled frames are represented as

$$S_{TYedge} = [TY'_{s_1}, TY'_{s_2}, TY'_{s_3}, TY'_{s_4}]$$
$$S_{TXedge} = [TX'_{s_1}, TX'_{s_2}, TX'_{s_3}, TX'_{s_4}] \quad (9)$$

– In a binary image, the pixel value of the detected edges is assigned $'1'$ and only vertical lines are considered because they are part of the scene change location. As discussed earlier, abrupt scene change appears as a vertical line which is occupied by the column in both TY and TX frames.
– Spatiotemporal frames are considered as noisy images and hence, sometimes, the boundary of scene change might be distorted. Therefore, the length of the boundary is defined as a scene change location when

$$Length = 40\% of\ (frame \quad height). \quad (10)$$

where frame height is defined as the height of TY or TX frames.
– Condition: $(number \quad of \quad 1's \quad in \quad boundary \quad line) >= Length.$
If the pixel value $('1')$ of any boundary or vertical line follows the above condition in both the TY and TX frames, then it is interpreted as the location of abrupt scene change.
– The above procedure is repeated for all the sampled frames.
– Now look up Table 1 that has been generated. In this table, all the detected locations obtained from all the frames $(TY, TX, andXY)$ are tabulated.
– Only those locations are considered that appeared at least twice among these spatial and spatiotemporal frames.

3 Experimental Results

Four natural test videos are taken from [17] for experiments and these are gstennis, anni002, anni003, and anni006. Video sequence gstennis (in Fig. 2) has 64 frames with one scene change and detection of scene change in spatiotemporal TY frame is

Table 1 Detected locations in spatial and spatiotemporal frames of anni006

Sampled frame	Detected Locations																					
TY'_{s_1}	–	**77**	–	**215**	**277**	**349**	**413**	**530**	577	**581**	**702**	–	**865**	**936**	–	**1029**	**1142**	**1318**	**1554**	**1656**	**1774**	**1890**	**1976**
TY'_{s_2}	–	**77**	–	**215**	**277**	**349**	**413**	**530**	–	**581**	**702**	–	**865**	**936**	–	**1029**	**1142**	**1318**	**1554**	**1656**	**1774**	**1890**	**1976**
TY'_{s_3}	–	**77**	–	**215**	–	**349**	**413**	**530**	577	**581**	**702**	–	**865**	**936**	–	**1029**	**1142**	**1318**	**1554**	**1656**	**1774**	**1890**	**1976**
TY'_{s_4}	30	**77**	–	**215**	–	**349**	**413**	**530**	–	**581**	**702**	–	**865**	**936**	–	**1029**	**1142**	**1318**	**1554**	**1656**	**1774**	**1890**	**1976**
TX'_{s_1}	–	**77**	–	**215**	**277**	**349**	**413**	**530**	–	**581**	**702**	–	**865**	**936**	–	**1029**	**1142**	**1318**	**1554**	–	**1774**	**1890**	**1976**
TX'_{s_2}	–	**77**	–	**215**	**277**	**349**	**413**	**530**	–	–	**702**	–	**865**	**936**	–	**1029**	**1142**	**1318**	**1554**	**1656**	**1774**	**1890**	**1976**
TX'_{s_3}	–	**77**	–	**215**	**277**	–	**413**	**530**	–	**581**	**702**	–	**865**	**936**	–	**1029**	**1142**	**1318**	**1554**	**1656**	**1774**	**1890**	**1976**
TX'_{s_4}	–	**77**	116	**215**	–	**349**	**413**	**530**	–	**581**	**702**	769	**865**	**936**	962	**1029**	**1142**	**1318**	**1554**	**1656**	**1774**	**1890**	**1976**
Frame XY	–	**77**			**277**	**349**	**413**	**530**	–	**581**	**702**	–	**865**	**936**	–	**1029**	**1142**	**1318**	**1554**	**1656**	**1774**	**1890**	**1976**

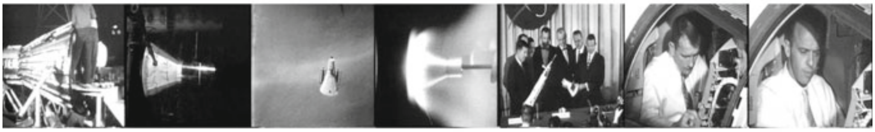
Fig. 2 Scenes of gstennis

(a) TY frame (b) TY edge (c) TY boundary

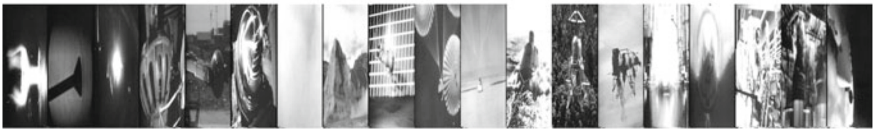
Fig. 3 Scene change detection in gstennis video

shown in Fig. 3. Since one scene change appears in gstennis video sequence, only one boundary or vertical line appeared in their binary image (Fig. 3b). Number of frames taken in videos anni002, anni006, and anni003 are 2048, 2048, and 1024, respectively, and all the scenes are represented in Fig. 4. The total number of abrupt scene changes in videos anni002, anni003, and anni006 are 12, 7, and 19, respectively. All the spatial and spatiotemporal frames have been processed and the obtained results by both the methods have been combined so as to obtain optimal results. Flow energy function of videos anni002 and anni003 are shown in Fig. 5 and 6, respectively. The vertical axis of the energy plot is the magnitude of flow energy and the horizontal axis represents the frame numbers. Flow energy of *XY* frames of video anni002 is shown in Fig. 5a and deviations in the plot show the location of scene change. Location of scene change obtained through threshold (6) is now represented by Fig. 5b and magnitude 1 is assigned to the detected location. Similarly, Fig. 6 represents the scene change detection of video anni003. Spatiotemporal frames *TY* and *TX* of video annoi006 have been shown in Figs. 7 and 8. In their binary images, vertical lines represent the location of scene change and all of them appear in the columns of the frame. Now these columns (in *TY* and *TX* frames) are directly converted into frame numbers (*XY*) with scene change locations. Flow energy of frames *XY* of video anni006 has been shown in Fig. 9a. Detected locations of abrupt scene change obtained from the proposed method are shown in Fig. 9b. Scene change location in video anni006

(a) Number of scenes in video anni002.

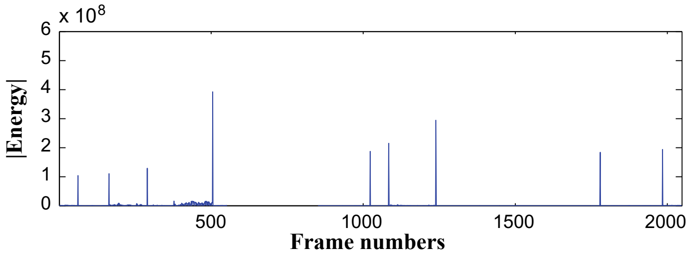


(b) Number of scenes in video anni003.

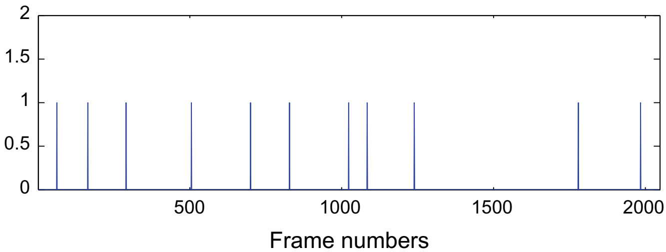


(c) Number of scenes in video anni006.

Fig. 4 Frames of different scenes



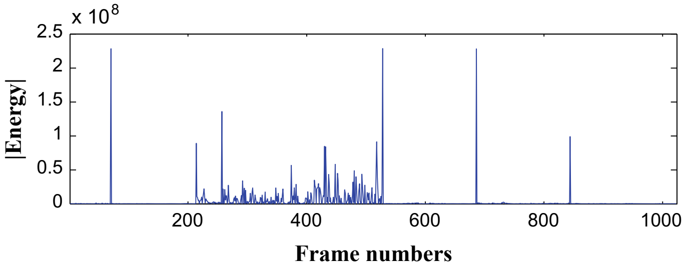
(a) Energy plot



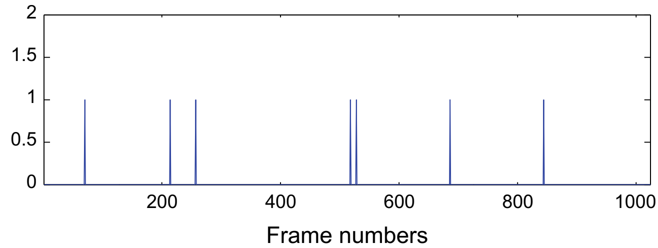
(b) Detected frames

Fig. 5 Detection of scene change in anni002

obtained by both the methods are combined together and tabulated in Table 1. There are three horizontal sections in Table 1 for the results obtained from frames TY, TX, and XY.



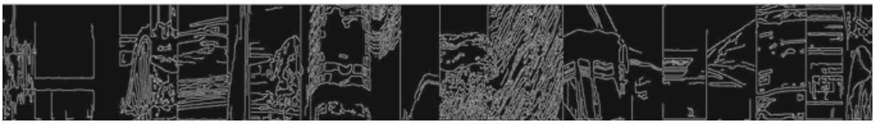
(a) Energy plot



(b) Detected frames

Fig. 6 Detection of scene change in anni003

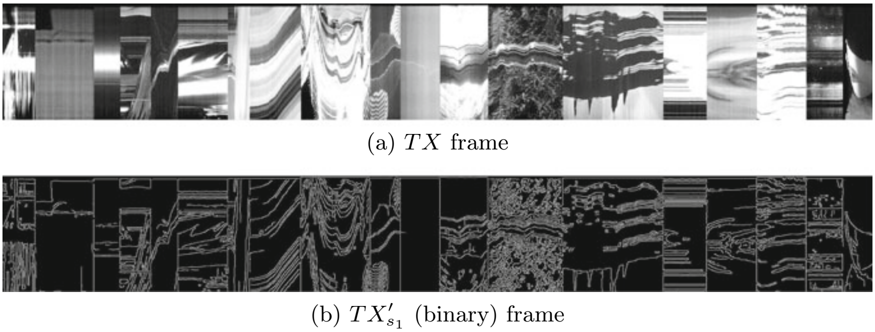
(a) TY frame



(b) TY'_{s_1} (binary) frame

Fig. 7 TY frame of video anni006

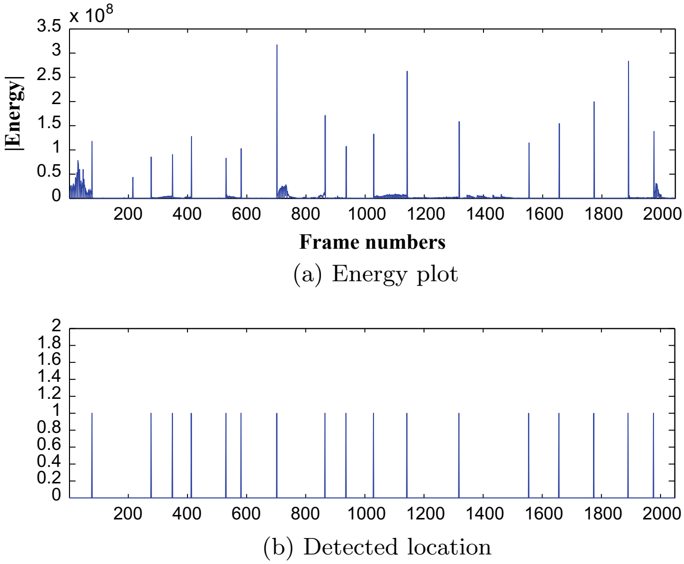
Among the three sections of spatial and spatiotemporal frames, location that appeared for at least two sections is considered as the approximated location of abrupt scene change. As shown in Table 1, locations obtained from the proposed method are [77, 215, 277, 349, 413, 530, 581, 702, 865, 936, 1029, 1142, 1318, 1554, 1774, 1890, 1976] and it is the same for the actual location. No missed or false detection are found and the proposed method detected all the abrupt scene changes. For the performance evaluation, the proposed method has been compared with the



(a) TX frame

(b) TX'_{s_1} (binary) frame

Fig. 8 *TX* frame of video anni006



(a) Energy plot

(b) Detected location

Fig. 9 Detection of scene change in anni006

method given in [2, 8] and the comparison result has been shown in Table 2. The accuracy of the proposed method has been evaluated with the help of precision, recall, and $F1$ score. $F1$ score is used for the evaluation of accuracy and it is defined as

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{11}$$

The proposed method detects all the scene changes and therefore the $F1$ score obtained is high as compared to the methods reported in [2, 8].

Table 2 Comparison

Test video	Methods	No. of scene change	F1
anni002	Method [8]	11	0.91
	Method [2]		0.83
	Proposed method		1
anni003	Method [8]	6	0.90
	Method [2]		0.86
	Proposed method		1
anni006	Method [8]	18	0.93
	Method [2]		0.82
	Proposed method		1

4 Conclusion

The proposed abrupt scene change detection method incorporates intensity as well as the flow energy of the frames. Using spatial and spatiotemporal frames reduces the false or missed detection. Edges of spatiotemporal frames and flow energy of spatial frames have been used for the detection. Detection accuracy of the proposed method is high with no false or missed detection as compared to the other methods.

References

1. Alatas, O., Yan, P., Shah, M.: Spatio–temporal regularity flow (SPREF): its estimation and applications. IEEE Trans. Circuits Syst. Video Technol. **17**(5), 584–589 (2007). https://doi.org/10.1109/TCSVT.2007.893832
2. Birinci, M., Kiranyaz, S.: A perceptual scheme for fully automatic video shot boundary detection. Signal Process.: Image Commun. **29**(3), 410–423 (2014)
3. Cyganek, B., Woźniak, M.: Tensor-based shot boundary detection in video streams. New Gener. Comput. **35**(4), 311–340 (2017)
4. Faernando, W., Canagarajah, C., Bull, D.: Scene change detection algorithms for content-based video indexing and retrieval. Electron. Commun. Eng. J. **13**(3), 117–126 (2001)
5. Hong, S., Cho, B., Choe, Y.: Adaptive thresholding for scene change detection. In: IEEE Third International Conference on Consumer Electronics, pp. 75–78. IEEE (2013)
6. Huang, C.L., Liao, B.Y.: A robust scene-change detection method for video segmentation. IEEE Trans. Circuits Syst. Video Technol. **11**(12), 1281–1288 (2001)
7. Jang, S.W., Byun, S.: Hough transform-based robust shot change detection in digital video images. Int. Inf. Inst. (Tokyo), Inf. **20**(2B), 1245 (2017)
8. Kang, S.J.: Adaptive luminance coding-based scene-change detection for frame rate up-conversion. IEEE Trans. Consum. Electron. **59**(2), 370–375 (2013)
9. Kang, S.J., Cho, S.I., Yoo, S., Kim, Y.H.: Scene change detection using multiple histograms for motion-compensated frame rate up-conversion. J. Disp. Technol. **8**(3), 121–126 (2012)
10. Koprinska, I., Carrato, S.: Temporal video segmentation: a survey. Signal Process.: Image Commun. **16**(5), 477–500 (2001)

11. Li, H., Liu, G., Zhang, Z., Li, Y.: Adaptive scene-detection algorithm for VBR video stream. IEEE Trans. Multimed. **6**(4), 624–633 (2004)
12. Majumdar, J., Aniketh, M., Abhishek, B., Hegde, N.: Video shot detection in transform domain. In: 2nd International Conference for Convergence in Technology (I2CT), pp. 161–168. IEEE (2017)
13. Prabavathy, A.K., Shree, J.D.: Histogram difference with fuzzy rule base modeling for gradual shot boundary detection in video cloud applications. Clust. Comput. 1–8 (2017)
14. Rosin, P.L., Ioannidis, E.: Evaluation of global image thresholding for change detection. Pattern Recognit. Lett. **24**(14), 2345–2356 (2003)
15. Rupesh, K., Gupta, S., Venkatesh, K.S.: Cut scene change detection using spatio temporal video frame. In: International Conference on Image Information Processing (ICIIP) (2015)
16. Shen, R.K., Lin, Y.N., Juang, T.T.Y., Shen, V.R., Lim, S.Y.: Automatic detection of video shot boundary in social media using a hybrid approach of HLFPN and keypoint matching. IEEE Trans. Comput. Soc. Syst. **5**(1), 210–219 (2018)
17. The open video project.: http://www.open-video.org/results.php?genre=Documentary
18. Youm, S., Kim, W.: Dynamic threshold method for scene change detection. In: International Conference on Multimedia and Expo, ICME'03, vol. 2, pp. II–337. IEEE (2003)