Circuits and Systems, 2016, 7, 2875-2883
Published Online August 2016 in SciRes. http://www.scirp.org/journal/cs
http://dx.doi.org/10.4236/cs.2016.710246

Video Shot Boundary Detection Using Normalized Periodogram Distance Metric

A. Sasithradevi, S. Mohamed Mansoor Roomi

Electronics and Communication Engineering, Thiagarajar College of Engineering, Madurai, India
Email: devisasithra@gmail.com

Received 14 May 2016; accepted 23 May 2016; published 5 August 2016

Abstract

Video shot boundary detection is the primary task for content based video management and retrieval system. This paper proposes a shot boundary detection strategy by exploiting the pros of Normalized Periodogram for efficiently representing the content of the video. A Normalized Periodogram based distance metric to detect the key frames using shot boundary, namely Distance-Left-Right (D_{LR}), is addressed, which is computed on a sliding sub-window basis. The D_{LR} sequence is used to detect the suspected shot boundary frames and a transition type detection procedure is adapted to these suspected frames for discriminating the abrupt and gradual transitions. The proposed shot boundary detection methodology yields Precision—95.02%, Recall—93.15% and F1 score—94.07% for cut, Precision—86.57%, Recall—86.67% and F1 score—86.61% for gradual, Precision—90.6%, Recall—90.02% and F1 score—90.3% for overall transitions. Experimental results show that the proposed approach is superior to the recently available shot boundary detection techniques because of its robustness and simplicity, and presents an effective distance metric to detect the shot boundary.

Keywords

Shot Boundary, Abrupt Transition, Gradual Transition, Normalized Periodogram

1. Introduction

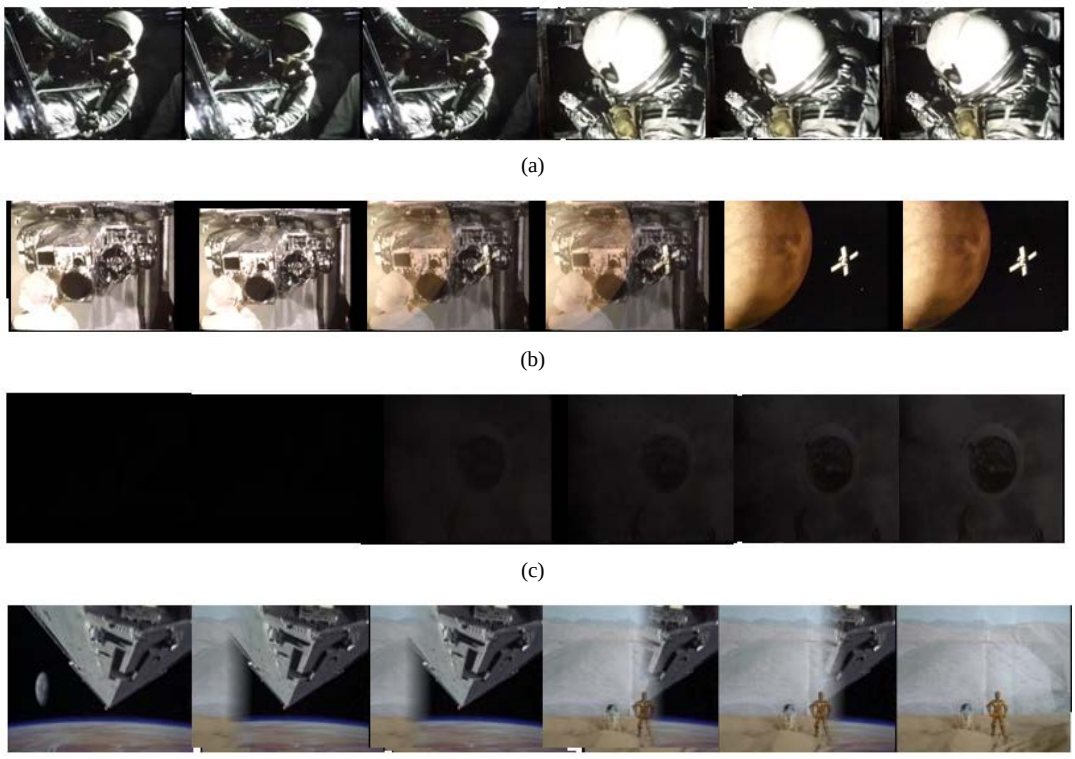
In this internet era, Digital Video plays a significant role in human's daily lives. Many practical applications like Video Retrieval, Video Surveillance, Video Content Analysis, Video Indexing, Video Skimming, etc., face trade-off between complexity and accuracy. The diverse content of video makes video management systems, a challenging task for multimedia researchers. Manual annotation of multimedia data is possible, but it is highly time consuming, which seeks the need for automatic vision algorithms for annotating the multimedia database over Internet. Video Shot Boundary Detection (VSBD) has been widely accepted as a solution to this trade-off

How to cite this paper: Sasithradevi, A. and Roomi, S.M.M. (2016) Video Shot Boundary Detection Using Normalized Periodogram Distance Metric. *Circuits and Systems*, **7**, 2875-2883. http://dx.doi.org/10.4236/cs.2016.710246

and structural analysis of video. Generally, frames extracted from the shot boundary are minimal compared to entire video content and represent the video effectively. A set of frames captured on a single camera is termed as shot. A shot can be categorized into cut or smooth based on the frames involved in transition as shown in **Figure 1**. Transition which involves sudden change from one frame to another is cut transition and smooth transition that involves sequence of frames due to several editing effects. This gradual transition can be dissolve, fade and wipe transitions. Dissolve involves very smooth disappearance of previous data and gradual appearance of new data in video. Wipe transition includes shapes like diamond, straight line, star or clock for frame transition. Fade of a frame occurs when multimedia information disappears onto a dark black screen. The process of temporally segmenting a video into shots includes three basic steps: 1) Frame Content Representation; 2) Similiarity/Dissimilarity evaluation between frame features and 3) Shot Boundary Detection [1].

Prior work on shot boundary detection mainly concentrates on the abrupt boundary detection and is very easy to detect the frames of sudden transition since the phenomena involve great discontinuity between adjacent frames. Most approaches involve a feature dissimilarity measure between the adjacent frames and predict the cut transition when the dissimilarity measure exceeds a threshold. Compared to abrupt SBD detection gradual SBD is complex as it does not involve great discontinuity between consecutive frames. Gradual SBD algorithms should be robust enough to issues like camera and object motion. The overall research work carried out in VSBD can be categorized, viz. pixel wise, global based, block based and motion activity based techniques. Various methodologies like [2] [3] use pixel difference as a common feature and these methods fail due to high false alarm rates raised due to fast camera operations in small area. To overcome these drawbacks global based approaches [4]-[6] have been proposed, which detect the boundary using measures like histogram difference, histogram intersection, weighted histogram difference, etc. Even though global approaches are robust to camera and object motion, spatial distribution changes between two different shots are not detected.

Block based approaches have been introduced to improve the SBD accuracy and reduce the computation time. All these approaches discussed so far involve features like moment invariants, local feature fusion, entropy, motion vector, Visual Bag of Words, Edge Change ratio, feature points, etc. Detecting the gradual transition by



(a)

(b)

(c)

(d)

Figure 1. Example of shot boundary transitions. (a) Abrupt Transition; (b) Dissolve Transition; (c) Fade Transition; (d) Wipe Transition.

predicting and training an appropriate model [7] for the corresponding transition has been reported. Mutual information and the joint entropy based cut/fade transitions [7], between consecutive frames, are also reported in literature. Using edge energy of DC coefficients [8], dissolve shot is detected based on U-shaped diagram search. Representing the video content directly by gradient and edge based features [9] is addressed for detecting shot boundary. In [9], the distribution of variance on the edge information is used to detect dissolve and fade transition. Edge based shot boundary detection algorithms suffer from poor performance due to object and camera motion.

As motion is continuous along a shot, motion is also used as a cue to detect shot changes. As the camera and object move graciously within a shot, the resulting motion field within a shot will be continuous. In [10] [11], a block matching algorithm which involves matching a block in the reference frame with all other blocks in the next frame for detecting shot boundary has been proposed. The performance of these methods depends only on the threshold procedure. To overcome this drawback, a multiple feature based cut and gradual detection with minimum number of threshold compared to [11] has been proposed. Though motion based algorithms are computationally expensive, cut transitions can be easily detected. One of the major drawbacks of motion based algorithms is that the algorithms can be easily fooled by varying illuminations.

Multiple features like pixel wise difference, color and edge histogram are extracted from the video frames and fed as input to the machine learning classifier, and support vector machine for transition classification [8] [12]. An accumulation histogram difference approach which can identify the dissolve and fade even under flash lights has been proposed [13]. In recent years, mutual information and joint entropy based transition detection algorithm [7], which can detect fade and cut, has also been proposed. A model based shot boundary detection algorithm based on frame transition parameter [14], a SVD based fast shot boundary detection algorithm [15] and Walsh Hadamard Transform (WHT) [16] based VSBD technique are the recent techniques available in literature.

One of the limitations of various algorithms proposed for VSBD phenomena is the lack of unified approach for detecting all types of transitions in various video streams like Video Lecture, News, Entertainment Shows, Sports and Movies. Many algorithms proposed for detecting all types of transitions include a tedious procedure and high computational cost. Most of the earlier SBD works are evaluated only on bench mark datasets and produce better results at high computational cost. Hence, the proposed methodology introduces a normalized periodogram distance based Left-Right (LR) ratio to detect the abrupt as well as gradual shot boundaries in video, which is efficient and effective in terms of accuracy and computational cost. The main contributions of this work are:

1) A normalized periodogram distance metric based LR ratio is introduced to detect the shots in a given video.
2) The proposed methodology is evaluated in unconstrained videos, including News, entertainment shows, Movies, Sports and TRECVID 2001 Dataset.

2. Proposed Video Shot Boundary Detection (VSBD) Methodology

This section elaborates the proposed normalized periodogram based D_{LR} metric for detecting both abrupt and gradual transition simultaneously. Given a video, sequence of frames obtained by partitioning the video is denoted as $V = \{f_1, f_2, \cdots, f_k\}$. For each frame f_k, the power spectrum is estimated using the classical non- parametric periodogram method. The periodogram of frames can be written as $Per = \{Per_1, Per_2, \cdots, Per_k\}$. Using suitable sub-window, the normalized periodogram based D_{LR} metric is computed for the feature frames and compared against the statistical threshold S_{th} chosen by trial and error method. The Frames with D_{LR} metric greater than S_{th} are suspected frames for shots. With the suspected frame as centre, the suitable suspected window is selected to decide the transition type as abrupt/gradual. The proposed flow graph for VSBD is shown in **Figure 2**.

2.1. Non-Parametric Power Spectrum Estimation

Periodogram is a non parametric technique for power spectrum estimation [17]. The periodogram of a random process is the Fourier transform of the autocorrelation of the random sequence. The autocorrelation of a matrix 'F' can be determined by,

$$R_f(k,l) \cong \sum_{i,j} f(i+k, j+l) f^*(i+k, j+l) \quad (1)$$

The periodogram can be written as,

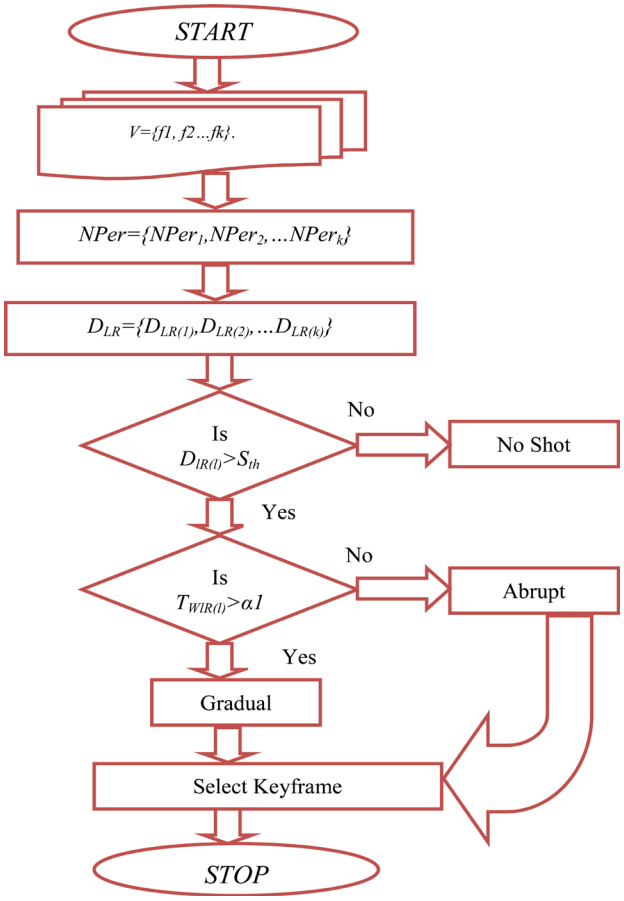


Figure 2. Proposed flow graph for key frame extraction.

$$per_f \cong \sum_{k,l} R_f(k,l)\mathrm{e}^{-j\omega(k+l)} \tag{2}$$

Even though the periodogram is represented using the autocorrelation function, it is necessary to represent periodogram in terms of the input frame/matrix '*f*'. Let $f_B(i, j)$ be the dot product of $f(i, j)$ and the box window filter $B(i, j)$,

$$f_B(i,j) = f(i,j)\cdot B(i,j) \tag{3}$$

The autocorrelation function of $F_B(i, j)$,

$$R_{f_B}(k,l) \cong \sum_{i,j} f_B(i+k, j+l)\, f_B^*(i+k, j+l) \tag{4}$$

Using Convolution Theorem of Fourier transform,

$$R_{f_B}(k,l) \cong f_B(k,l)\, f_B^*(k,l) \cong \left|F_B(k,l)\right|^2 \tag{5}$$

where $F_B(k, l)$ is the Fourier Transform of the frame $f_b(i, j)$ at pixel i, j of size $M \times N$.

Properties of Periodogram

Previous section clearly shows that the Periodogram is directly proportional to the squared magnitude of the Fourier Transform and is very simple to compute. This section gives a gist of the properties of periodogram as follows,

1) Bias of the Periodogram:

The expected value of the peridogram of $f(i, j)$ is the convolution of the power spectrum with the Fourier transform of Bartlett Window, Periodogram is a biased estimate.

$$E\left\{Per_f\left(k,l\right)\right\} = \frac{1}{2\pi} P_f\left(k,l\right) * W_B\left(k,l\right) \tag{6}$$

where $P_f(k, l)$ is the power spectrum of $f(i, j)$ and $W_B(k, l)$ is the Fourier Transform of the Bartlet window.

2) Variance of the periodogram:

Variance of the periodogram does not converges and the periodogram $Per_f(k, l)$ is not the consistent estimate of the power spectrum. The variance of the periodogram is proportional to the square of the power spectrum of $f(i, j)$

$$Var\left\{Per_f\left(k,l\right)\right\} = P_f^2\left(k,l\right) \tag{7}$$

2.2. D_{LR} Metric Computation and Statistical Threshold Selection

A normalized periodogram distance, a periodogram based metric for shot boundary classification is detailed in this section. Consider the Power spectral estimate of two frames as

$$Per_x\left(k,l\right) \cong \left| \sum_{\hat{i},\hat{j}} x\left(\hat{i},\hat{j}\right) \mathrm{e}^{-j\omega\left(k+l\right)} \right|^2 \tag{8}$$

$$Per_y\left(k,l\right) \cong \left| \sum_{\hat{i},\hat{j}} y\left(\hat{i},\hat{j}\right) \mathrm{e}^{-j\omega\left(k+l\right)} \right|^2 \tag{9}$$

The periodogram distance between frame x and y can be written as,

$$Dist_{per(x,y)} \approx \sqrt{\sum_{k,l} \left[Per_x\left(k,l\right) - per_y\left(k,l\right) \right]^2} \tag{10}$$

The main intention of using periodogram in this work is to visualize the correlation between frames, Hence normalized periodogram is sufficient for this objective and is given by $N_{per_x(k,l)} = Per_x\left(k,l\right)/\beta$, where β is the variance estimate of the frame x. The normalized periodogram distance is written as,

$$Dist_{N_{per(x,y)}} \approx \sqrt{\sum_{k,l} \left[N_{Per_x(k,l)} - N_{per_y(k,l)} \right]^2} \tag{11}$$

From the property 2 of periodogram it is evident that the variance of the periodogram is proportional to the spectral value and therefore it is meaningful to use logarithm of normalized periodogram. The normalized periodogram distance satisfies the basic properties of a metric:

Property 1: Symmetry property; $Dist\left(a,b\right) = Dist\left(b,a\right)$

Property 2: Non-negative property; $Dist\left(a,b\right) \neq 0,\ a \neq b$

Property 3: Triangle-inequality; $Dist\left(a,b\right) \leq Dist\left(a,c\right) + Dist\left(c,b\right)$

With the knowledge of Normalized Periodogram Distance (NPD) between consecutive frames, select a sub-window of size 2W + 1 for the D_{LR} metric computation, explained as follows:

Step 1: Select the left "W" frames as sample set "L" and right "W" frames as sample set "R".

Step 2: Compute the normalized periodogram distance between each sample in L and centre sample, D_{LC} = *median*(L_j-C), where j is the number of frames in left window, and C is the centre NPD frame in the sub-window "W". Similarly, calculate D_{RC} = *median*(R_j-C).

Step 3: Compute $D_{LR} = D_{LC}/D_{RC}$.

The same process is repeated for all k frames in the video. The obtained D_{LR} metric is compared against the statistical threshold given by,

$$S_{th} = \alpha * \mu \tag{12}$$

The frames with D_{LR} metric greater than S_{th} are termed as suspected frames. These suspected frames are given as input to the Transition Type Identification Procedure (TTIP) and is detailed in the following algorithm. The flow graph of the D_{LR} metric computation followed by TTIP is shown in **Figure 3**. After detecting shot boundary using proposed methodology, key frames are extracted based on [18].

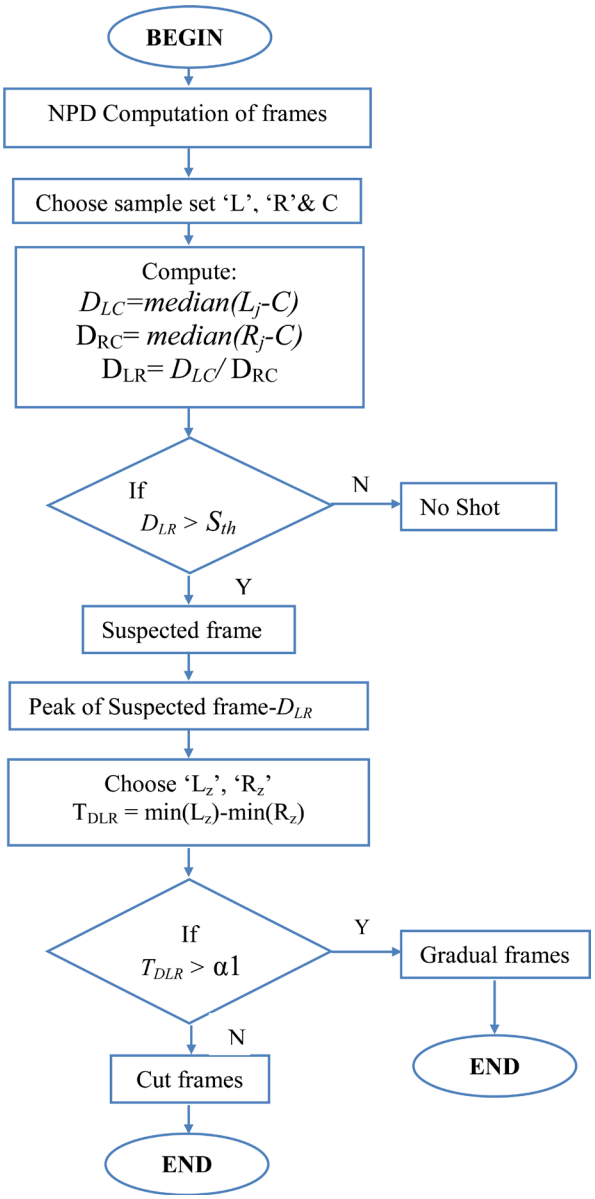


Figure 3. Proposed flow graph for D_{LR} and TTIP.

Algorithm 1: Transition Type Identification Procedure

Input: Peaks of D_{LR} metric of suspected frames

Output: cut/gradual

Step 1: For each peak of the D_{LR} metric choose a window of size 2Z + 1 with centre as peak value. Select the left "Z" D_{LR} metric as sample set L_Z and right "Z" D_{LR} metric as set R_Z.

Step 2: T_{DLR} = min(R_Z)-min(L_Z).

Step 3: If $T_{DLR} > \alpha 1$, suspected frame is gradual; Else suspected frame is cut.

3. Experimental Results and Performance Evaluation

This section presents the evaluation of proposed method over existing methodologies for shot boundary detection. Experimentation is carried out using Matlab 8.5 software on DELL i3 core system. Description of the test dataset, evaluation measures and performance of the proposed methodology over state of art methods are detailed below:

3.1. Description of Test Videos

To evaluate the performance of the proposed approach, various test videos from OPEN VIDEO [19] and Youtube are downloaded. The test videos include entertainment shows, Song, Movie, Sports and News videos. These videos include abrupt and gradual transitions. VID1-VID11 are the videos collected from Youtube and OPEN VIDEO with vast lighting effects and camera motion. The Benchmark dataset namely TRECVID2001 [20] which is widely used for VSBD purposes is also evaluated using proposed methodology. The details of the test video like number of frames, duration and number of shots are shown in **Table 1**.

3.2. Parameter Selection

The parameters need to be set in the proposed method are "W", "Z", "α", "α1". The sub window size "W" is varied from 5 to 25 in steps of 5 and experimented on the TRECVID data of 5000 frames as shown in **Figure 4**. For W = 5, though the precision measure is fair, more false hits occur. For W = 10, precision and recall measure shows little improvement, still false hits remain. Precision and recall value at W = 15, shows improved result than W = 10. Better performance is achieved at W = 20. The value of 'α' in (12) is set as 2 by trial and error

Table 1. Video test data description.

Video	Video details			
	Number of Frames	Duration(s)	Number of Shots	Characteristics
VID1	200	7	3	Varying lighting effects, Object motion
VID2	300	10	2	Varying lighting effects, Camera and Object motion
VID3	500	16	1	Camera motion
VID4	400	13	3	Varying illumination and Object motion
VID5	300	10	2	Varying illumination and Object motion
VID6	300	10	2	Varying illumination and Object motion
VID7	150	5	3	Special effects, Varying illumination and Object motion
VID8	900	30	6	Camera and Object motion
VID9	350	11	4	Varying lighting effects, Camera and Object motion
VID10	500	16	3	Varying lighting effects, Camera and Object motion
VID11	300	10	2	Varying lighting effects, Camera and Object motion
VID12	11,356	379	65	Varying lighting effects, Camera and Object motion
VID13	16,586	553	73	Varying lighting effects, Camera and Object motion
VID14	12,304	410	103	Varying lighting effects, Camera and Object motion
VID 15	31,389	1046	153	Varying lighting effects, Camera and Object motion

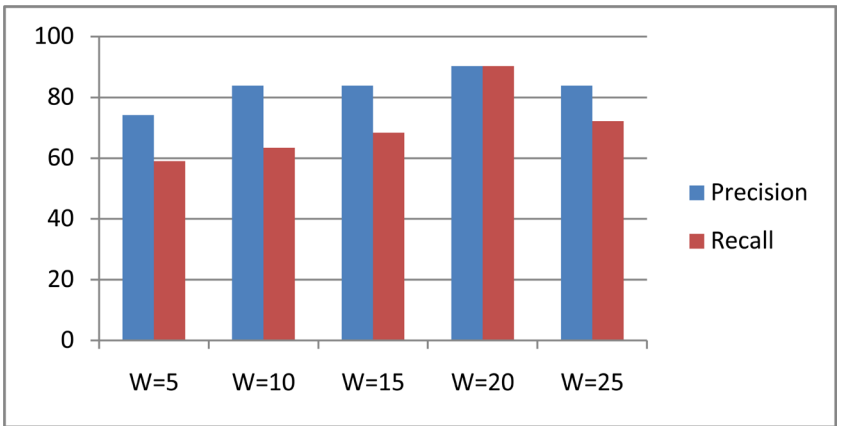


Figure 4. Window size vs. performance.

method and it varies for different video. The window size "Z", required to determine the type of shot, is chosen as 20, since minimum gradual transition duration involves 25 frames.

3.3. Performance Evaluation

For evaluating the proposed D_{LR} based methodology, benchmark video dataset TRECVID 2001 is used to select the key frames from the video. The evaluating metrics namely precision and recal1 are computed using,

$$\text{Precision} = \frac{\text{Number of shots correctly detected}}{\text{Number of shots in reference}} \tag{13}$$

$$\text{Recall} = \frac{\text{Number of shots correctly detected}}{\text{Number of shots in reference} + \text{false Hits}} \tag{14}$$

As illustrated in **Table 2**, the performance of the proposed method is compared with recent methodology namely Walsh Hadamard Transform (WHT) [16] based VSBD process. The proposed approach shows a better performance in VID12, VID13 and VID15. The poor performance of the proposed algorithm in VID14 is due to the drastic camera motion. With W = 20, Z = 20, α = 2, α1 = 20, the proposed approach is also verified with the user collected database including entertainment, news and sports videos. For VID1-VID8 the proposed approach produce astounding results, but VID9 include fast object movement, which is confused a shot by D_{LR} metric. Running the proposed approach on i3 core system, the time taken for processing the consecutive frames using [16] is six times greater than the proposed approach. **Table 3** depicts the performance of the performance of the proposed algorithm in VID1-VID9. Hence, the periodogram based technique is quite simple and efficient for detecting shot boundaries in any video.

4. Conclusion

A robust and efficient technique for detecting abrupt and gradual shots in a video is presented. The power spectrum is estimated for video frames and using suitable window size, D_{LR} metric is evaluated for the spectral features extracted from the frames. Suspected video frames are detected using statistical threshold approach on the

Table 2. Performance comparison with recent methodologies.

Video	VSBD using WHT [16]				VSBD using proposed method					
	Cut		Gradual		Cut		Gradual		Overall	
	P	R	P	R	P	R	P	R	P	R
VID12	85.4	97.6	90	87	**94.7**	**100**	**92**	**100**	**93.6**	100
VID13	86.5	82.1	88.7	85.9	**95.2**	**86.9**	**80.6**	**80.6**	89	84.4
VID14	**90.6**	**88.8**	**84.6**	**80**	92.3	87.8	82.8	76.8	86.4	80.9
VID15	93.5	95.6	88.3	88.5	**97.9**	**97.9**	**90.9**	**89.3**	**93.4**	94.8

Table 3. Performance of proposed D_{LR} metric.

Video	VSBD using proposed methodology	
	Manually annotated shots	Proposed automatic detection
VID1	3	3
VID2	2	2
VID3	1	1
VID4	3	3
VID5	2	2
VID6	2	2
VID7	3	3
VID8	6	6
VID9	4	5

computed D_{LR} metric and transition type detection procedure is used to classify the abrupt and gradual transitions. Thus the proposed periodogram based D_{LR} metric shows a promising performance in constrained and unconstrained video data for detecting shot boundaries. The proposed method fails under some drastic camera and object movement conditions, which can be improved by including motion feature.

References

[1] Yuan, J.H., Wang, H.Y., Xiao, W., Zheng, J., Li, J.M., Lin, F.Z. and Zhang, B. (2007) A Formal Study of Shot Boundary Detection. *IEEE Transactions on Circuits and Systems for Video Technology*, **17**, 168-186. http://dx.doi.org/10.1109/TCSVT.2006.888023

[2] Nagasaka, A., Tanaka, Y., Knuth, E. and Wegner, L. (1992) Automatic Video Indexing and Full Video Search for Object Appearances. In: Knuth, E. and Wegner, L., Eds., *Visual Database Systems II*, North Holland Publishing Co., Netherlands, 113-127.

[3] Zhang, C. and Wang, W. (2012) A Robust and Efficient Shot Boundary Detection Approach Based on Fisher Criterion. *Proceedings of* 20*th ACM International Conference on Multimedia*, Nara, 29 October-2 November 2012, 701-704.

[4] Patel, N.V. and Sethi, I.K. (1997) Video Shot Detection and Characterization for Video Databases. *Pattern Recognition*, **30**, 583-592. http://dx.doi.org/10.1016/S0031-3203(96)00114-8

[5] Tan, Y.P., Nagamani, J. and Lu, H. (2003) Modified Kolmogorov-Smirnov Metric for Shot Boundary Detection. *Electronic Letters*, **39**, 1313-1315.

[6] Dailianas, A., Allen, R.B. and England, P. (1996) Comparison of Automatic Video Segmentation Algorithms. *Proceedings of SPIE*, **2615**, 2-16. http://dx.doi.org/10.1117/12.229193

[7] Cernekova, Z., Pitas, I. and Nikou, C. (2006) Information Theory-Based Shot Cut/Fade Detection and Video Summarization. *IEEE Transactions on Circuits and Systems for Video Technology*, **16**, 82-91. http://dx.doi.org/10.1109/TCSVT.2005.856896

[8] Petersohn, C. (2004) Dissolve Shot Boundary Determination. In *Proceedings of European Workshop for the Integration of Knowledge, Semantics and Digital Media Technology*, London, 25-26 November 2004, 87-94.

[9] Yoo, H.W., Ryoo, H.J. and Jang, D.S. (2006) Gradual Shot Boundary Detection Using Localized Edge Blocks. *Multimedia Tools Appl*., **28**, 283-300. http://dx.doi.org/10.1007/s11042-006-7715-8

[10] Shahraray, B. (1995) Scene Change Detection and Content-Based Sampling of Video Sequences. *Proceedings of SPIE*, **2419**, 2-13. http://dx.doi.org/10.1117/12.206348

[11] Kawai, Y., Sumiyoshi, V. and Yagi, N. (2007) Shot Boundary Detection at TRECVID 2007. *Proceedings of TREC Video Retrieval Evaluation Online*, TRECVID 2007 Workshop, Gaithersburg, 16 January 2007.

[12] Xue, L., Li, C., Li, H. and Xiong, Z. (2008) A General Method for Shot Boundary Detection. *International Conference on Multimedia and Ubiquitous Engineering*, Busan, 24-26 April 2008, 394-397.

[13] Ji, Q.G., Feng, J.W., Zhao, J. and Lu, Z.M. (2010) Effective Dissolve Detection Based on Accumulating Histogram Difference and the Support Point. 1*st International Conference on Pervasive Computing Signal Processing and Applications*, Harbin, 17-19 September 2010, 273-276. http://dx.doi.org/10.1109/pcspa.2010.73

[14] Lu, Z.M. and Shi, Y. (2013) Fast Video Shot Boundary Detection Based on SVD and Pattern Matching. *IEEE Transactions on Image Processing*, **22**, 5136-5145. http://dx.doi.org/10.1109/TIP.2013.2282081

[15] Mohanta, P.P., Saha, S.K. and Chanda, B. (2012) A Model-Based Shot Boundary Detection Technique Using Frame Transition Parameters. *IEEE Transactions on Multimedia*, **14**, 223-233. http://dx.doi.org/10.1109/TMM.2011.2170963

[16] Lakshmi Priya, G.G. and Domnic, S. (2014) Walsh-Hadamard Transform Kernel-Based Feature Vector for Shot Boundary Detection. *IEEE Transactions on Image Processing*, **23**, 5187-5197. http://dx.doi.org/10.1109/TIP.2014.2362652

[17] Hayes, M.H. (2008) Statistical Signal Processing and Modelling, 393-407.

[18] Zhuang, Y.T., Rui, Y., Huang, T.S. and Mehrotra, S. (1998) Adaptive Key Frame Extraction Using Unsupervised Clustering. *Proceedings of IEEE International Conference on Image Processing*, Chicago, 4-7 October 1998, 866-870.

[19] Open Video Project. [Online]. http://www.open-video.org/

[20] TRECVID Dataset. [Online]. http://trecvid.nist.gov/