Multilevel Wavelet Based Feature for detecting Gradual Transitions in High Quality Videos

Kamal S. Chandwani
PhD Scholar, Sarvepalli Radhakrishnan University, Bhopal, Madhya Pradesh, India
chandwani1@rediffmail.com

Varsha Namdeo
Professor, Department of Computer Science & Engineering, Sarvepalli Radhakrishnan University
Bhopal, Madhya Pradesh, India, Varsha_namdeo@yahoo.com

Narendra P. Giradkar
Associate Professor, Department of Electronics & Telecommunication Engineering
Smt. Radhikatai Pandav College of Engineering , Nagpur, Maharashtra, India
giradkarnaresh@gmail.com

Prashant R. Patil
Professor, Department of Management Studies , Smt. Radhikatai Pandav College of Engineering ,
Nagpur, Maharashtra, India, patilnagpur@gmail.com

Abstract

Gradual transitions are highly confused in high quality videos with camera motion and illumination effects and most of the time are missed or falsely detected. Today's multimedia technology have immensely advanced that video creators use various eye straightening techniques to attract audience thus making difficult to isolate gradual transitions from such videos. The special effects are confused with actual transitions and a highly robust system would possibly fail to locate the actual endpoints of gradual transitions. We propose a simple and efficient method to detect gradual transitions especially fade in and fade out in videos using mean of features obtained from Multilevel Wavelet transform of the grayscale frames. These values give better correlation between successive frames and the significant decreasing and increasing slopes obtained when properly analysed using global minima and local maxima's can accurately find the locations of fade out and fade in respectively. An experimentally found threshold able to lower the computational complexity and conserve computation time. We tested our results with manually found transitions extremes points and found that our method worked better than manual approach and possess the ability even the transitions occurred under worst illumination.

Keywords – *Gradual transitions, fade in, fade out, multilevel wavelet transform, global minima, and local maxima.*

DOI Number: 10.14704/nq.2022.20.13.NQ88349 **Neuro Quantology 2022; 20(13):2805-2812**

Introduction

Rapid advancements in Multimedia Technology has fascinated elementary units of video and contributed in the development of enhanced videos across the world. The concatenation of several such transitions based clips had become the source of a video as a whole. A video sequence comprising an abrupt or gradual transition to constitute a single shot is a continuous sequence of frames acquired using a single acquisition system like camera or a cell phone. A hard cut or an abrupt transition and the gradual transition is used to distinguish between two scenes in a way that the boundary between two such scenes can only be distinguished with soft computation. The hard cuts are due to sudden change of one shot to another and can be easily identified whereas gradual transitions including fade in, fade out, wipes and dissolves are difficult to be recognised with the assistance of simple algorithms. The difficulty of identifying scene change are due to object motion, lightening effects and other special effects to improve the perceptual quality when a user enjoys during watching. These appearances sometimes are confused with the gradual changes and any sophisticated applications designed for detecting the scene cuts are then confused and leads to false detection.

Content Based Video Retrieval (CBVR) is and effective solution to facilitate video indexing, editing, summarization and retrieval. Past literature focuses on different techniques with reliable performances especially for key frames extraction and summarization. Static and local video summarization includes selection of representative frames from each shot to form a storyboard. A simple approach of comparing two consecutive frames based on pixel intensities/density or color components were the source of shot boundary detection (SBD) algorithms which were suited to videos not prone to camera and object motions and inhomogeneous lightning conditions [1]. These simple approaches were then enhanced by filters of variable sizes to smoothen the image frames and then subjected the frequency of pixels to some predefined threshold which was experimentally found from the videos under test [2]. Information regarding to changes in Edge feature were used for block based comparison methods to improve the robustness against the camera and object motions [3-4]. Since other features of the frames were neglected, blurred edges were not sensitive to detect the SBD. Work presented in [5] used inter-frame differences based on color histogram for SBD with higher computation speed and precision. Finer version of histogram including leverage x^2 histogram [6] and block histograms [7] were used to include fine details from the frames. Illumination variations are compensated in [8] for SBD whereas appropriate distance metrics to compare frames was put forthwith by [9]. The degree of similarity between successive frames based on mutual information between them was suggested by [10]. The degree of similarity is higher when the mutual information between two frames is higher.

Although all the above mentioned methods were successful in detecting hard or abrupt cuts but faced difficulties in finding the gradual transitions since the intersection between frames at the gradual transition is not apparent and the difference between consecutive frames at the transition are closely correlated. The work presented in [11] used two thresholds but was unable to find the starting or initial frame of gradual transition which was then corrected using principle of optical flow by [12-13] on the basis that there was no optical flow at such transition. Due to complexity of the work mentioned in [12-13], a three stage work proposed by [14] used segment selection, singular value decomposition and pattern matching. The position of boundary and length of the transition was calculated using an adaptive threshold discarding the frames not involved in the transition. Support vector machine was used to classify different types of shots using extracted features in [15].

Related Work

Datasets provided by TRECVid which was introduced by National Standards and Technology (NIST) are considered to be benchmark for CBVD processes. TRECVid provides multiple datasets for SBD and also conducted workshop where large numbers of researchers across the globe had actively participated and contributed in the field [16-18]. A detailed study of various SBD techniques is listed in the work outlined by Abdulhussian et al. [19]. It includes methods involving SBD with simple or local and global features based difference, statistical features to features extracted with convolutional networks. Work proposed in [20] highlighted comparison based on colour histogram, temporal relationship was gained through the use of local binary pattern (LBP) to extract more detailed fine features with cosine distance metric in [21], histogram in HSV color space with single level difference and level 5 difference were used in [22] for abrupt and gradual transitions respectively to mitigate the illumination effects in the frames. Further, in order to increase the depth of features, multilevel features based on SURF histograms, colour histograms and gradient histograms were efficiently used by Tippayya et al. [23] and Lin et al. [24].

More sophisticated SBD algorithm to detect especially gradual changes are the twin comparison method [25] where two thresholds viz. higher and the lower are used to detect abrupt and the gradual cuts. After applying the higher threshold to detect high discontinuity values and isolating the abrupt cuts, another threshold is applied to the rest of the discontinuity values. The initial point of gradual transition is detected when the discontinuity value is higher than the next threshold. The addition of consecutive discontinuity values in monitored until the cumulative sum exceeds the first threshold. The last discontinuity value represents the end of gradual transition. The methods adversely depends on two threshold and increases the probability of false positives. The author in [26] made use of width and depth of the parabolic curve obtained using variance of successive frames under compressed domain but fails to identify when the frames are subjected to noise and motions in video. Plateau detection scheme introduced in [27] used difference of current and the k^{th} frame. A significant plateau which is higher than former and later values is characterized by a sequence of similar values. The method stood successful for linear and nonlinear transitions and relied on rises and falls of the plateau extremes. A simpler concept to detect gradual transitions was adopted in [28] where the transition was considered to be two piece wise linear function of time (increasing and decreasing) and no such linearity outside the transition area. The author investigated the first and the second derivative of the slope in time. A no transition region is identified when the second derivative is smaller than threshold part of first derivative. Troung et al. in [29] used an adaptive threshold calculated from intensity histogram difference curve from a local window.

Advance methods used for SBD includes machine learning algorithms such as SVM and Neural Networks where the classifiers are trained with some extracted features. The performance of such modern classifiers depends solely on the frames used for training the classifier which needs to be well balanced because any outlier test sample will decrease the SBD accuracy. Work proposed in [30] used bagged tree classifiers with features extracted using pyramid level opponent colour space. Features extracted using the non-subsampled contourlet transform were further reduced using principal component analysis and then classified using least squared SVM in [31] and minimized the effect of class imbalance. K-means clustering was used to identify initial cut and extract features using the SVM classifier in [32] whereas deep learning convolutional neural was used for detecting gradual transitions in [33].

Our Contribution

We propose a novel method to detect fade out and fade in with higher accuracy and use only one threshold T which is determined through experimental analysis. The threshold T so considered in optimum when applied to number of distinct videos subjected to object motion and inhomogeneous illumination. The novel approach is simple with low computational and time complexity. The approach is based on finding suitable global minima and local maxima in the vector obtained by averaging the frame features. The grayscale frames are subjected to multilevel wavelet transform and the average value of Approximation coefficients are stored in a vector. The vector is then analyzed with respect to global minima and local maxima to find the starting and ending points of fade in's and fade out's. The process considers 8 frames for fade in and fade out as minimum numbers of frames to exist. We consider a block of 75 frames at any instance to determine whether the block contains any of the gradual transition with an offset of 10 frames every time the frame block is changes until the last block is scanned.

Proposed Methodology

The frames in the video are converted to grayscale frames and stored to reduce the computational complexity and do not take into account the color components as the minima and the maxima's are related to the intensity values of the pixels. The first block under consideration starts from first frame and ends at 75^{th} frame in the video. The next successive block will start from 11^{th} frame and ends at 85^{th} frame. Thus an offset of 10 frames is considered while considering each block of frames. We have split longer videos in a time frame of 20 seconds minimum.

All the frames in a block are subjected to wavelet decomposition to level 5 using debauchee's mother wavelet to find the features of the frames. The Approximation coefficient at level 5 which is exact replica of the original frame in compressed form is then averaged over the rows and columns to calculate a single feature value of the frame for finding the correlation between its immediate neighbors. The mean value representing the frame depends on the coefficient at level 5 and is the function of objects present in the frame and its illumination. Therefore to bound all the mean values corresponding to 75 frames in a fixed interval of [0 1], the values are normalized by dividing each of the value by the maximum value found for any frame in the block.

Let A denotes the array consisting of all grayscale frames from a video and B to represent the block of 75 frames from A. The level 5 decomposition of frames in B can be evaluated using the following figure 1. All other components at all levels including horizontal (Hx), vertical (Vx) and diagonal (Dx) are ignored, where x = 1 to 5. The significant approximation coefficient A5 is used for feature whose average of mean is computed.

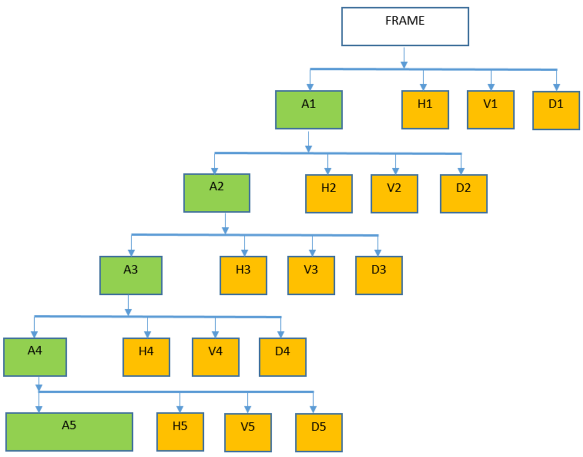

Figure 1 –Wavelet decomposition to level 5 for the grayscale frame

The average or the mean value of A5 component is computed using the following expression (1):

$$I = \sum_{c=1}^{csize} \sum_{r=1}^{rsize} A5 \qquad (1)$$

Where, csize and rsize are the number of columns and rows in the coefficient matrix A5. Let I(n) represents the vector consisting of all such mean values of approximation component A5 and 'n' represent number of frames in the block that is, n=1 to 75. Then the maximum value in I(n) is found by the following equation (2):

$$I_{max} = \max_{n} I \qquad (2)$$

Then the normalized vector I_{norm} is given by the following expression (3).

$$I_{norm} = \frac{I}{I_{max}} \qquad (3)$$

At this stage, it is to confirm whether the frame block under consideration does really contain (either fade in or fade out) the gradual transition. Our method is able to tell at this stage only whether the block feature vector is to be processed or not or just move to new block. Here we find the minimum value I_{min} given by expression (4) in the mean vector I and check the whether this minimum value I_{min} is less than the threshold T = 0.1. Expression (5) decides the significance G that is either the gradual transition exists or not.

$$I_{min} = \min_{n} I \qquad (4)$$

$$G = \begin{cases} Gradual\ Transition & I_{min} < T \\ No\ transition\ Block & Otherwise \end{cases} \qquad (5)$$

Now find the frame index (peakpt_idx) in vector I for the minimum point. And then the frame index (Imax_idx) whose value approaches value 1 in vector I starting from frame 1 in block to the frame index at minimum point (peakpt_idx). If the frame index (Imax_idx) is empty, fade out is ruled out. Therefore, we have to find the existence of fade in. The value of frame index (Imax_idx) is now found from vector I starting from frame index at (Imax_idx) to end of the block.

When Imax_idx is not empty, the following procedure is adopted to find the starting and ending frames constituting the fade out.

a. We find all the crests in frames (pkpt) starting at index 2 to frame index ending at Imax_idx.
b. Set end index of the fade out at fade_e = peakpt_idx(1)
c. If pkpt is not empty,
i. Set start index of fade out fade_s = pkpt(end)
ii. If, (fade_e – fade_s) > 8 (no_frames)
iii. Store the starting and ending frame index for fade transition.
iv. Increment the fade out count variable.

A similar approach is adopted for fade in except the step (a) in modified with the loop parameters. Here we find the crests in frames (pkptafter) starting at index peakpt_idx(1) and ends at last frame index minus 1.

When all the starting and ending frame indexes are collected from the video, the redundant transitions indexes are eliminated out. This is due to the offset of 10 frames considered while choosing the blocks. Figure 2 shows the selection of blocks from the video.

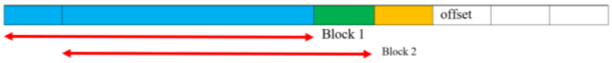

Figure 2 – Block selection with offset of 10 frames

Results and Conclusion

The proposed work is efficient in detecting not only fade out and fade in when they occur individually across the frames but also when both of them are immediate in a given block. We had manually found the starting and ending frame indexes when transition is present the videos and tested with results obtained with our method. We found that our method was able to find accurately the frame indexes than with manual approach since the transitions are not clearly perceptual at the edges. There is possibility of having multiple minimum points while estimating peakpt_idx points therefore we considered the first occurrence for such cases. The following figure 3 shows the detection of fade out in a video block from frames 528 to 556. Figure 4 shows the detection of fade in existing from frame 556 to 591. We have shown both the graph separately in figure 3 and figure 4 but as seen from the frame numbers these are successive. That is, fade out is followed immediately by fade in and depicted in figure 5.

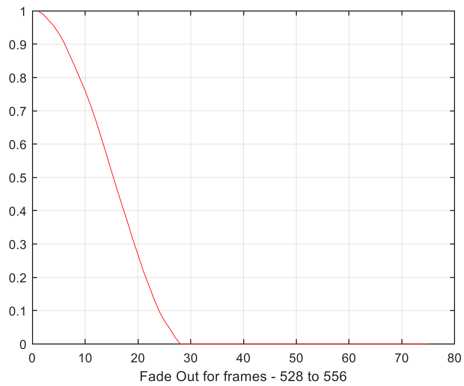

Figure 3 – Detected Fade out transition from frame 528 to 556

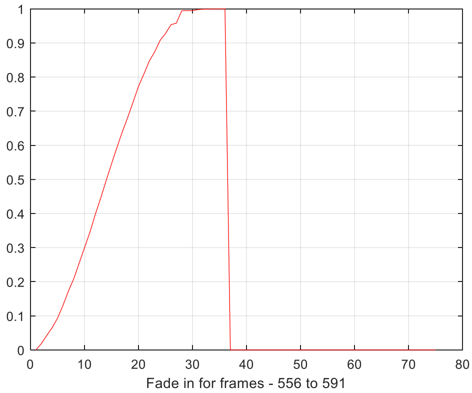

Figure 4 – Detected Fade in transition from frame 556 to 591

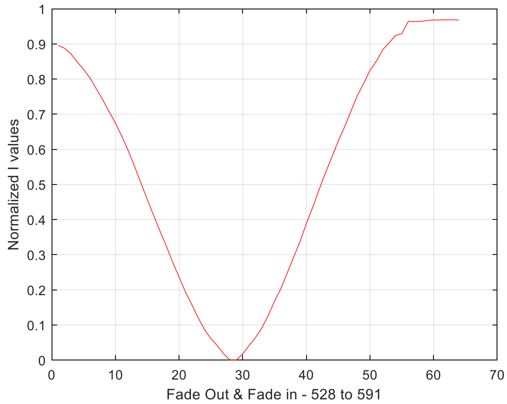

Figure 5 – Fade out and fade in in frames 528 to 591

The normalized values of I for all frames is plotted in figure 6 where the gradual transition is clearly seen as per detected in the frames by our proposed technique. The system is designed and evaluated using MATLAB R2021b with i5 processor, 16 GB RAM with 2.70 GHz. A single threshold T is experimentally found to know the existence of any gradual transition in the test video block. This drastically reduces the computation time and speed up the detection process processing only those block carrying such gradual transitions.

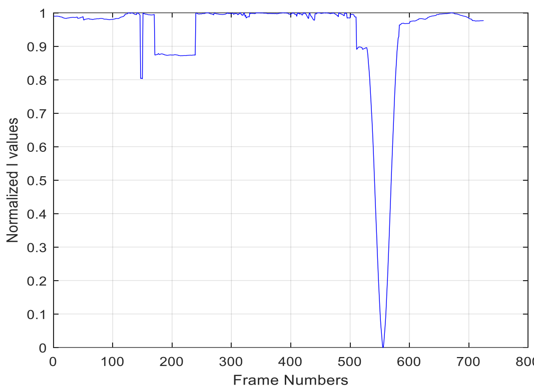

Figure 6 – Normalized values of mean values for all frame in the video

We conclude that the proposed method of finding features using multilevel wavelet transform and then finding the mean of the approximation coefficient at level 5 can successfully detect gradual transitions in video blocks. We have carefully examined all aspects regarding finding minima and maxima so that no transition is missed using the proposed technique. All parameters are properly worked out and tested over number of videos. The threshold of T = 0.1 is universal and is compatible with all sorts of videos. The technique suggested in this work eliminates the effect of any camera motion or illumination effects. We have not compared the results to state of art techniques available in the literature but definitely the same can be done using videos from standard dataset like TRECVID. The work will be focussed on detecting dissolve transitions in future. Other results are depicted in tabular format for video clips extracted from Hindi movie "DIL HI TO HAI". The reason to choose clips from its songs is for the number of gradual transitions and the complexity under object motion. Table I shows the detected transitions in various clips.

Table I – Result of proposed work over video clips from movie "DIL HI TO HAI"

CLIP	Fade out	Fade in	Both fade out and fade in
1	None	Frame 11 - 39	None
2	Frame 377 - 405	Frame 405 - 425	Frame 377 - 425
3	Frame 73 - 94	Frame 94 - 118	Frame 73 - 118
4	Frame 166 - 315	Frame 191 - 346	None
5	Frame 394 - 415	Frame 420 - 446	Frame 394 - 446

References

[1] J. Z. Wang, J. Li, R. Gray and G. Wiederholdm, "Unsupervised multiresolution segmentation for images with low depth of field,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 1, pp. 85-90, Jan. 2001.

[2] H. Zhang, A. Kankanhalli and S. W. Smoliar, "Automatic partitioning of full-motion video,'' Multimedia Syst., vol. 1, no. 1, pp. 10-28, Jan. 1993.

[3] S. Lefèvre and N. Vincent, ``Efficient and robust shot change detection,'' J. Real-Time Image Process., 2007, vol. 2, no. 1, pp. 23-34.

[4] R. Zabih, J. Miller and K. Mai, "A feature-based algorithm for detecting and classifying scene breaks,'' in Proc. 3rd ACM Int. Conf. Multimedia, 1995, vol. 7, no. 2, pp. 189-200.

[5] F. Hong-Cai, Y. Xiao-Juan, M. Wei and Y. Cao, "A shot boundary detection method based on color space,'' in Proc. IEEE Int. Conf. E-Bus. E-Government, May 2010, pp. 1647-1650.

[6] A. Nagasaka and Y. Tanaka, "Automatic video indexing and full-video search for object appearances,'' IPSJ J., January 1991, vol. 33, pp. 113-127.

[7] Z. Yihua, "A quadratic difference-based method for abrupt shot change detection,'' Comput. Eng. Appl., 2005, vol. 41, no. 6, pp. 22-25.

[8] Y. Yong and H. Hou, "Adaptive shot segmentation method based on histogram frame difference,'' Comput. Eng. Appl., 2010, vol. 46, no. 9, pp. 186-189.

[9] C. Cai, K. M. Lam and Z. Tan, "TRECVID2005 experiments in the Hong Kong polytechnic University: Shot boundary detection based on a multistep comparison scheme,'' TREC Video Retr. Eval. Notebook Papers, Austin, TX, USA, Tech. Rep., January 2005.

[10] Z. Cernekova, I. Pitas, and C. Nikou, "Information theory-based shot cut/fade detection and video summarization,'' IEEE Trans. Circuits Syst. Video Technol., January 2005, vol. 16, no. 1, pp. 82-91.

[11] J. Wang and W. Luo, "A self-adapting dual-threshold method for video shot transition detection,'' in Proc. IEEE Int. Conf. Netw., Sens. Control, April 2008, pp. 704-707.

[12] H. J. Zhang, J. Wu, D. Zhang, and S. W. Smoliar, "An integrated system for content-based video retrieval and browsing,'' Pattern Recognition, 1997, vol. 30, no. 4, pp. 643-658.

[13] S. Lian, ``Automatic video temporal segmentation based on multiple features,'' Soft Computing, 2011, vol. 15, no. 3, pp. 469-482.

[14] Z.-M. Lu and Y. Shi, ``Fast video shot boundary detection based on SVD and pattern matching,'' IEEE Trans. Image Process., December 2013, vol. 22, no. 12, pp. 5136_5145.

[15] J. Yuan, H. Wang, L. Xiao, W. Zheng, J. Li, F. Lin and B. Zhang, ``A formal study of shot boundary detection,'' IEEE Trans. Circuits Syst. Video Technol., February 2007, vol. 17, no. 2, pp. 168-186.

[16] M. M¨uhling, R. Ewerth, T. Stadelmann, C. Z¨ofel, B. Shi and B. Freisleben, "University of Marburg at TRECVID 2007: Shot Boundary Detection and High Level Feature Extraction," in Proceedings of the TRECVID 2007 workshop participants notebook papers, Gaithersburg, MD, USA, November 2007.

[17] J. Yuan, Z. Guo, L. Lv et al., "THU and ICRC at TRECVID 2007," in Proceedings of the TRECVID 2007 workshop participants notebook papers, Gaithersburg, MD, USA, November 2007.

[18] Z. Liu, D. Gibbon, E. Zavesky, B. Shahraray and P. Haffner, "At&t research at trecvid 2007," in Proceedings of the TRECVID 2007 workshop participants notebook papers, November 2007, pp. 19–26, Gaithersburg, MD, USA.

[19] S. H. Abdulhussain, A. R. Ramli, M. I. Saripan, B. M. Mahmmod, S. A. R. Al-Haddad and W. A. Jassim,

"Methods and challenges in shot boundary detection: a review," Entropy, 2018, vol. 20, no. 4, pp. 214.

[20] U. Gargi, R. Kasturi and S. H. Strayer, "Performance characterization of video-shot-change detection methods," Circuits Syst. Video Technol, 2000, vol. 10, no. 1, pp. 1–13.

[21] S. Teng and W. Tan, "Video temporal segmentation using support vector machine," in Asia Information Retrieval Symposium, H. Li, T. Liu, W. Y. Ma, T. Sakai, K. F. Wong, and G. Zhou, Eds., 2008, vol. 4993, pp. 442–447, Springer, Berlin, Heidelberg.

[22] Z. Li, X. Liu and S. Zhang, "Shot boundary detection based on multilevel di9erence of colour histograms," in Proceedings of the First International Conference on Multimedia and Image Processing (ICMIP), June 2016, pp. 15–22, IEEE Press, Bandar Seri Begawan, Brunei.

[23] S. Tippaya, S. Sitjongsataporn, T. Tan, M. M. Khan and K. Chamnongthai, "Multi-modal visual features-based video shot boundary detection," IEEE Access, 2017, vol. 5, pp. 12563–12575.

[24] C.-H. Lin, M.-D. Hsiao and Li-J. Fu, "Analysis of shot boundary based on color and texture features of frame IOP Conference Series: materials Science and Engineering," IOP Conference Series: Materials Science and Engineering, 2013, vol. 53, no. 1, Article ID 012004.

[25] Zhang HJ, Kankanhalli A., Smoliar SW and Tan SY, "Automatic partitioning of full motion video," ACM Multimedia Systems, 1993, vol. 1, issue 1, pp. 10–28.

[26] Meng J., Juan Y., Chang SF.,"Scene change detection in a MPEG compressed video sequence'" In: Proc. SPIE/IS&T Symp. Electronic Imaging Science and Technology: Digital Video Compression: Algorithms and Technologies, 1994, vol. 2419, pp. 14–25.

[27] Yeo BL and Liu B., "Rapid scene analysis on compressed video," IEEE Trans Circuits Syst. Video Technol. 1995, vol. 5, issue 6, pp. 533–544.

[28] Song S., Kwon T., and Kim W., "Detection of gradual scene changes for parsing of video data," In: Proc IS&T/SPIE, 1998, vol. 3312, pp. 404–413.

[29] Truong BT, Dorai C. and Venkatesh S., "New enhancements to cut, fade, and dissolve detection processes in video segmentation," ACM International Conference on Multimedia, 2000, pp. 219–227.

[30] Sasithradevi A., Roomi S. M. M., "A new pyramidal opponent color-shape model

based video shot boundary detection, " Journal of Vis. Communication & Image Represent. 2020, vol. 67, 102754.

[31] Mondal J., Kundu M. K., Das S. and Chowdhury M., "Video shot boundary detection using multiscale geometric analysis of nsct and least squares support vector machine," Multimedia Tools Appl. 2017, vol. 77, pp. 8139–8161.

[32] Yazdi, M.; Fani, M. Shot boundary detection with effective prediction of transitions' positions and spans by use of classifiers and adaptive thresholds. In Proceedings of the 2016 24th Iranian Conference on Electrical Engineering (ICEE), Shiraz, Iran, 10–12 May 2016, pp. 167–172.

[33] Xu J., Song L. and Xie R., "Shot boundary detection using convolutional neural networks," In Proceedings of the 2016 Visual Communications and Image Processing (VCIP), Chengdu, China, 27–30 November 2016, pp. 1–4.