International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 6 Issue 5, July-August 2022 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Comparative Study of Different Video Shot Boundary Detection Techniques

Swati Hadke[1], Ravi Mishra[2]

[1]Assistant Professor, [2]Associate Professor,
[1,2]Department of Electronics & Telecommunication Engineering,
G. H. Raisoni Institute of Engineering & Technology, Nagpur, Maharashtra, India

ABSTRACT
Video shot boundary detection is the crucial step in the field of research of video processing. This makes the task of video retrieval based on contents, indexing and browsing. This paper contains the review of different techniques and methods which implemented to achieve the task of SBD along with key performance measurement parameters. This explains different preprocessing techniques, feature extraction methodologies, similarity computation techniques, etc. The outcomes of different approaches are framed with reference to accuracy, speed of computation along with comparison of precision, recall and F1 score.

KEYWORDS: *Shot boundary detection, Speeded-Up-Robust Features, High Level Fuzzy Petri Net, Scale Invariant Fourier Transform, Dual Tree Complex Wavelet Transform, Walsh-Hadamard Transform*

How to cite this paper: Swati Hadke | Ravi Mishra "Comparative Study of Different Video Shot Boundary Detection Techniques" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-6 | Issue-5, August 2022, pp.1300-1308, URL: www.ijtsrd.com/papers/ijtsrd50630.pdf

IJTSRD50630

INTRODUCTION

Videos are consisting of different shots and shots made up of frames that connected together in consecutive manner. Thus for a user it is always difficult to find or retrieve the desired content of video from the large database. This problem then motivated to research on how the shot boundary be detected, thus solution for this can be given by video shot boundary detection techniques. In this, the detection of discontinuity in a stream of video is to be taken place. The discontinuity refers between the shots which called the primitive component of video. This is advantageous for many purposes of video browsing, indexing and retrieval. The shot may take different forms as Abrupt and Gradual. The gradual transitions further divided to Fade in/out, dissolve and wipe. Generally, the SBD framework designed with steps mentioned in fig 2. For movies, media, news, entertainment, etc. the detection of shot boundary is very crucial. As there are many approaches which does a lot much of research work to achieve the said task above of detection of shot boundary in a video stream. This undergoes several steps to achieve the accurate detection. Fig.1 shows the video stream structure.

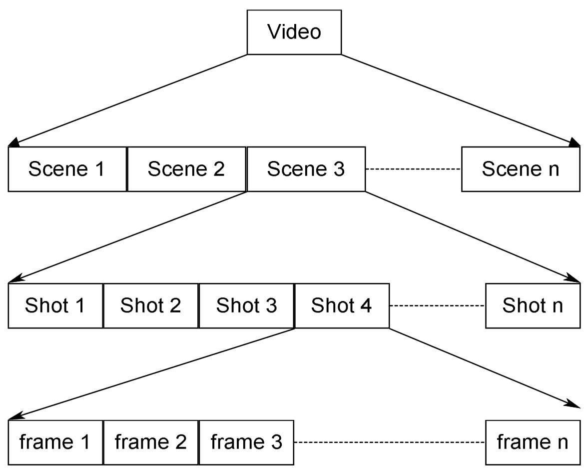

Fig.1 Video Stream Structure

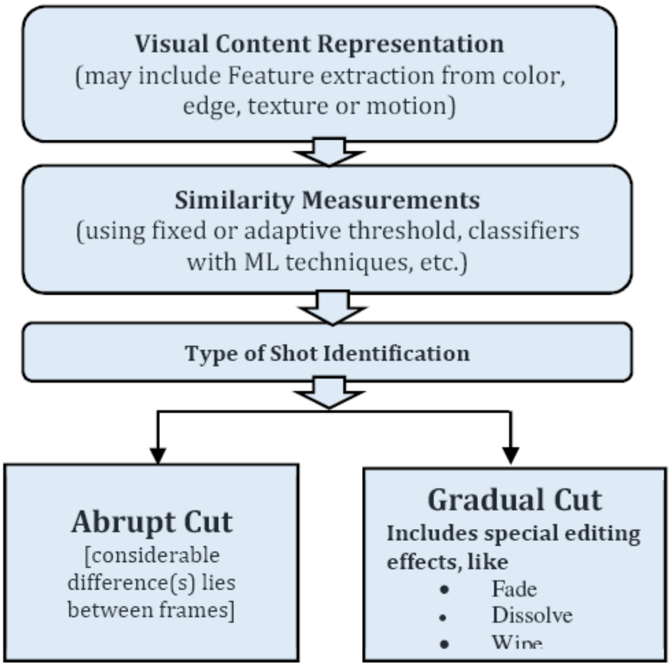

Fig. 2. Steps in SBD

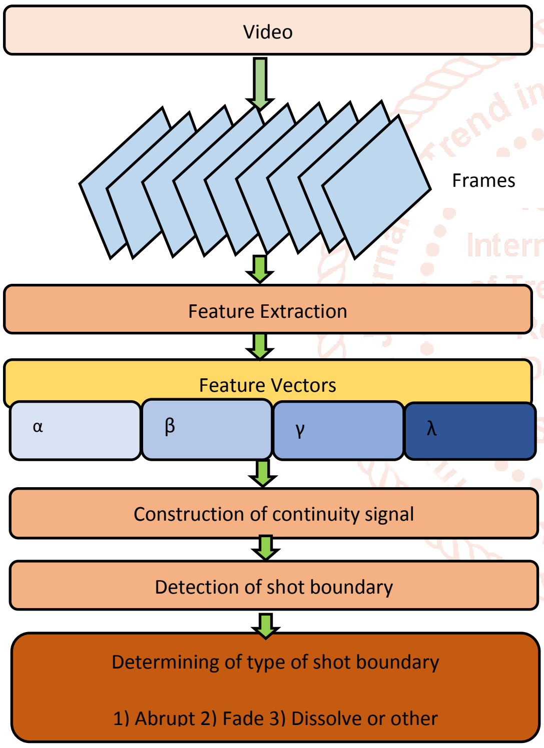

Fig. 3. Algorithm of WHT method

The identification of Abrupt cut or Hard cut is comparatively easy with respect to Gradual cut. Since gradual transition (GT) includes various special editing effects. In addition, the flashlight, object/camera movements also lead to false results in detection of shot(s).

In GT, effects can be Fade, Dissolve and Wipe. From fig.2,it is clear that, the detection process consists of the steps such as, segmentation of video into several frames, preprocessing, similarity matching or formation of continuity function and then the identification of shot boundary. Further it can be classified as Gradual or Abrupt.

The video segmentation provides the frames to perform the preprocessing. The preprocessing involves feature extraction, there are several feature descriptors are available which can take various forms in terms of RGB, HSV and LAB.

The above descriptors have their own advantages and disadvantages. Taking the specific histogram into consideration, one or more can be utilized for the preprocessing purposes.

With reference to literatures regarding the research in video processing, especially in shot boundary detection, Walsh Hadamard Transform, Dual Tree Complex Wavelet Transform, Discrete Cosine Transform, Fuzzy Logics, Convolutional Neural Network, High Level Fuzzy PetriNet, etc. contributed to achieve the remarkable outcomes.

All above mentioned, having their own benefits to achieve good score in Precision and Recall, hence also the combined score (F1 Score) which helps in judging the methodology. As seen in detail there are the compromises made with accuracy and speed of execution, user oriented approach and complex approach, etc.

Review of Literature

The Video shot boundary can be referred to as the primitive component for research in the field video processing for wide applications. With reference to the types of transitions that mentioned here, various methodologies were implemented and existed. LaxmiPriya [3] presented an approach of WHT for the SBD where, the features of edge, color, texture and motion were extracted by WHT kernel matrix. Thus, matching can be evaluated for various extracted video frames then followed by continuity function to design which decides the detection of boundary.

$\emptyset=\omega 1\alpha+\omega 2\beta+\omega 3\gamma+\omega 4\lambda$ …. 1)

Where, ω1, ω2, ω3, ω4 are the weight coefficients which calculated using feature weighing methods.

α,β,γ and λ are the extracted feature vectors of color, edge, texture and motion respectively.

The type of transition can be identified with reference to the peak that formed for a continuity value, shown as per fig.3.The initial process begins with feature extraction which can be global and/or local features [6] [7] and use of machine learning approach can classify the transitions as shot or non-shot categories [7]. The SVM used most preferably. Considering the types of transitions as gradual and abrupt, the separate detection algorithms have also been implemented in [15] making use of convolutional neural networks. Abrupt transitions detected with the help of fusing color histogram and deep features, gradual transition detection taken place with 3D CNN, this helps in classification of clips into specific shot change type. Victor R.L. Shen [2] given an approach of HLFPN to reduce the computational time for shot detection but the accuracy not promising. While Rong-Kuan Shen [23] presented a hybrid approach of HLFPN and keypoint matching, in which detection of gradual transitions and reduced false matches were the benefits mentioned for this method in addition different video types also can undergo the SBD task. Frame rate up conversion [4] is another method adopted to detect gradual shot change, where the combine features of motion and luminance were used. To identify the location of shot, the threshold changes adaptively and both methods with independently performed. The recall ratewas achieved well but initial setting of threshold is to be made manually. Another approach of GPU accelerated hard cut detection [9] claimed faster execution by making the use of SURF from fig. 4, both local and global features were extracted along with this the machine learning approach was implemented. This benefited less computational time but with lower recall rate.

Another approach of Fully Convolutional Neural Networks [13] for SBD enables to run as faster as 120x real time, but missing of long dissolves and also not accurately performed for partial scene changes and faster scenes with blur motion. In this, the flashes were artificially added and which made network invariant to flashes that may otherwise lead to false matches. N. J. Janwe and K. K. Bhoyar[1]; suggested a technique of SBD based on JND color histogram, in which adaptive threshold was used based on sliding window. Dissolve and fade of gradual transition along with hard cuts were appropriately detected but the performance highly depends upon the window size. Wu Z., Xu P. [18] mentioned the use of SURF for said purpose, where the pretreatment

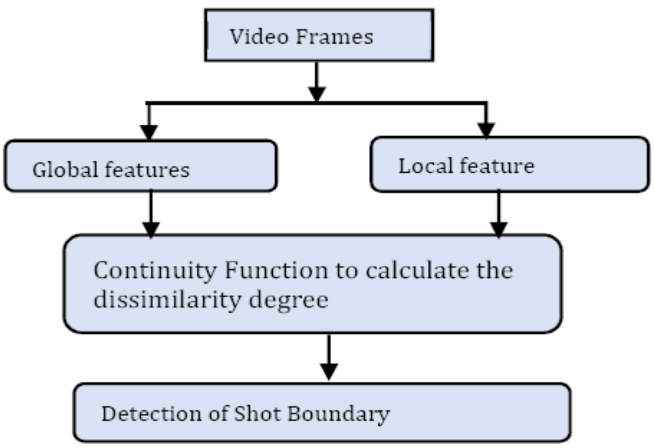

Fig. 4. SBD approach using SURF

consists of frame difference measurement and use of adaptive threshold to detect gradual transitions. Alan Hanjalic [24] suggested the solution for SBD using statistical methods, such as shot-length distribution, visual discontinuity patterns at shot boundaries, and characteristic temporal changes of visual features around a boundary. Victor R. L. Shen [21] proposed MLPN approach which has given flexibility in learning and multiple heterogeneous outputs can be drawn. Also offers an advantage of faster learning. WeimingHu, [22], et al; explained the overall concepts in video retrieval on content based in which the shot boundary detection was the primitive step. Here the video structure analysis has been discussed with SBD, with the same reference the threshold based approach and statistical learning based approaches were discussed. In addition, Supervised and unsupervised algorithms, key frames extraction techniques, scene segmentation, feature extraction from frames, video data mining and classification, etc. were discussed. Ravi Mishra [5] explained the SBD using Dual Tree Complex Wavelet Transform for real time and non-real time videos, here, adaptive threshold was used to compare the things with reference to locate the shot boundary. The system claimed to be efficient for different AVI videos to detect the transitions. Hannane, R., Elboushaki, A., Afdel, K. [8] presented SIFT-point distribution histogram approach and claimed to be efficient in presence of illumination to detect both the transition types but on other side compromise to be made with accuracy due to false matches met due to large noticeable motion. T. Kar, P. Kanungo [10], et al; suggested a work to detect the shot in presence of motion and illumination, in this the generated feature frames converted to gradient oriented feature and then adaptive threshold acted upon to locate the shot boundary. There are different feature extraction techniques are available [11] along with their similarity measures for distance calculations. TejaswiniKar, Priyadarshi Kanungo [12] proposed cut detection based on weber features, which claimed good implementation in presence of motion and illuminations, but limited to light environment, the same has given the false matches for dark environments. Claudia C. Oprea [14], et al; proposed an approach of SBD for low complexity HEVC coders, this has given the faster response but limited to Chroma components and not able to work with greyscale sequence. Shunmugam Karpagavalli [16] proposed technique that combines the Hessian matrix of point of interest and the minimum Eigen values of the region of interest, thus made the hybrid key point detection with several benchmark algorithms. Ahmed Khazaal Sulaiman [17], et al; suggested to have SBD for static and dynamic objects with static and dynamic camera motion.

The preprocessing can be done as, Feature extraction, for this various techniques introduced as HSV [6], RGB, block HSV histogram [7], YCbCr [3]. The color feature assumed to be primitive one for the purpose to form then continuity function to compare the adjacent frames. As the pixel values for them signifies the similarity or differences. The distribution of pixel intensities also the parameters useful to take part in similarity calculations. Fast detection approach in SBD proposed in [25] where the frames considered were half in number and then undergone the process of feature extraction using SURF shown in fig. 5, hence benefited to have less computation time since all frames need not be processed.

Shot 1				**Shot 2**			
F1	F2	F3	F4	F5	F6	F7	F8
F1		F3		F5		F7	
feat 1		feat 2		feat 3		feat 4	

Fig. 5. Illustration how frames were selected

There are many approaches of color descriptors, preferably HSV has been chosen as far as the literatures under review. The another component can be the edge descriptor. The edges are less susceptible to lighting and camera movements. the magnitude of gradient vector defines the strength of edge. Another is spatial arrangement of pixels called texture is contributing to the feature component to take care of similarity measure. The motion effects can be reduced by computing motion strength [3]. X. Qian [26], et al; explained the method for fade and flashlight detection using an approach of accumulating histogram difference, in this the classification of AHD characteristics of fades was done by forming the mathematical models and then comparison was made for the grey values of the corresponding monochrome color with maximum and minimum grey values of the solid color frames during fades transitions, thus detection of fades and flashlights was made.

Methodologies implemented and their outcomes for Shot boundary detection

As earlier mentioned in literature review, lot many efficient techniques were developed for the said task and those will be discussed here with their results.

There different coding languages, different software tools and different approaches are available, with which whatever the results obtained are frames in the reviewed literatures, for that the comparative analysis of the performances in terms of accuracy, computation speed, simple and complex techniques, etc. are mentioned in the article. As in Fig 6, the general procedure starts with video frame extraction and segmentation. Then preprocessing includes feature extraction. Literature reviewed here presented different ways, like the use of JND color histogram [1] offered several advantages over RGB/HSV to compute the color index, but on other side drawback of wastage of space for the colors which are not even available. The HSV histogram [9] [18] as considered to be more effective for the perception of color as compared to RGB [6], in addition the combined local and global features of block HSV histograms can be used instead of RGB color histograms [7] which overcomes the issue of sensitivity to camera/object movements that can arise with RGB. Also the SIFT-PDH used to achieve said purpose along with adaptive filter [8], SURF and SIFT have their own advantages and disadvantages but application orientation can make their better use to fit for the desired outcome.

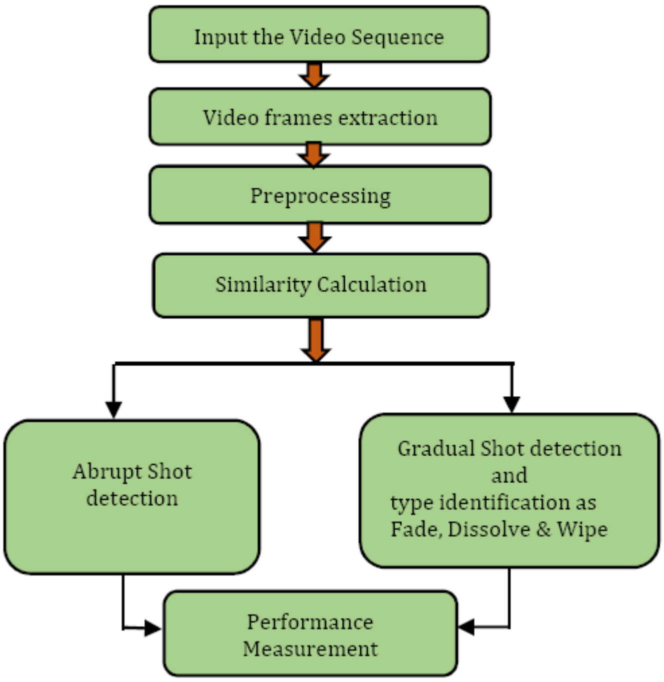

Fig. 6. General procedure for SBD

The use of HOG descriptor explained in [14] for performance improvement of an algorithm, here for primary shot detection of video HSV color histogram was used, and for secondary shot detection HOG features were used.

Statistical functions [24] that helps to measure the visual content discontinuity by computing mean absolute change of intensity I (x, y) for frames *k* and *k+1* for all frame pixels. Here the values of x and y from I (x, y) varies from $1 \le x \le X$ and $1 \le y \le Y$, where X and Y are the dimensions of frames. Further compared with threshold *Th,*

$$Z(k,k+1) = \frac{1}{XY}\sum_{x=1}^{X}\sum_{y=1}^{Y} D\,k,k+1(x,y)$$

with

$$Dk,k+1\,(x,y) = \begin{cases} 1, if\ |Ik(x,y) - Ik+1(x,y)| > Th \\ 0, Else \end{cases}$$

……… 2)

Thus adopting one or more approaches, the feature(s) can be extracted. Afterwards the feature matching is to be conducted where two feature matrices may be *f (i)* and *f (i+1)* can be compared and distance calculation among them gives the degree of similarity. Fixed and adaptive thresholds then used to decide the occurrence of boundary. In addition, the machine learning approaches also helps to classify the types of shot boundary situations. Precision, recall and F1 score values will decide the effectiveness and reliability of an algorithm.

Recall = T / (T+M)
Precision = T / (T+F)
F1 measure = 2*R*P / (R+P)
Where, T = Correct transitions detected
M = Missed transitions
R = False transitions detected

Precision is parameter shows the relevancy from total detected frames, Recall refers to correctly detected relevant shots, and F1 score is the weighted average of both P & R.

There are many methods have been compared and shown in the table 1. The comparison shown in terms of precision, recall and F1 score, which are assumed to be the performance determining parameters. Table 2 shows the comparative analysis of transitions where different methods were compared with HKD i.e. hybrid keypoint detection method [16] in terms of number of correctly detected transitions for both abrupt and gradual cut. Fig. 7 shows the statistical analysis of performance of different approaches. Fig. 8 gives the comparison of F1 score for the approaches or methodologies under review. For this, different machine learning approaches, neural networks, classifiers, etc. were used for determining the semantic factors and then identifying the discontinuity. Followed by this, the type of the transitions also determined. There are the tradeoffs for the accuracy and speed. The comparison shown below presented the numbers for judgment of effectiveness of the implemented method(s). Also the numbers may vary as per the datasets used, since along with shot detections and its type identification, the illuminations and lighting effects also to be considered for avoiding the false matches. There are special attentions also given for this problems [26]. Fig. 9 shows the statistics for comparative analysis on transitions of different algorithms in reference with HKD method [16].

Table I Performance Comparison for different approaches

Label	Method	Precision	Recall	F 1
M1	Unsupervised Clustering	98.04%	98.04%	98.04%
M2	WHT	89.30%	86.30%	87.70%
M3	Global & Local Feature descriptors	88.50%	65.90%	74.60%
M4	Supervised clustering	94%	69%	82%
M5	Fast algorithm for SBD with SURF	100%	99.84%	99.91%
M6	HLFPN &keypoint matching	86.56%	72.39%	76.90%
M7	Time warping & Mean Shift	95.40%	100%	97.50%
M8	Blocked HSV	95.70%	86.10%	96.60%
M9	C3D	81.30%	79.60%	80.40%
M10	HSV +DPHA	90.70%	90.10%	90.40%
M11	TSSBD	93.20%	93.80%	93.50%
M12	GPU + SURF	90.35%	63.59%	72.60%
M13	Graph Model	96.47%	94.41%	95.37%
M14	Combined Features + SVM	92.84%	75.87%	82.10%

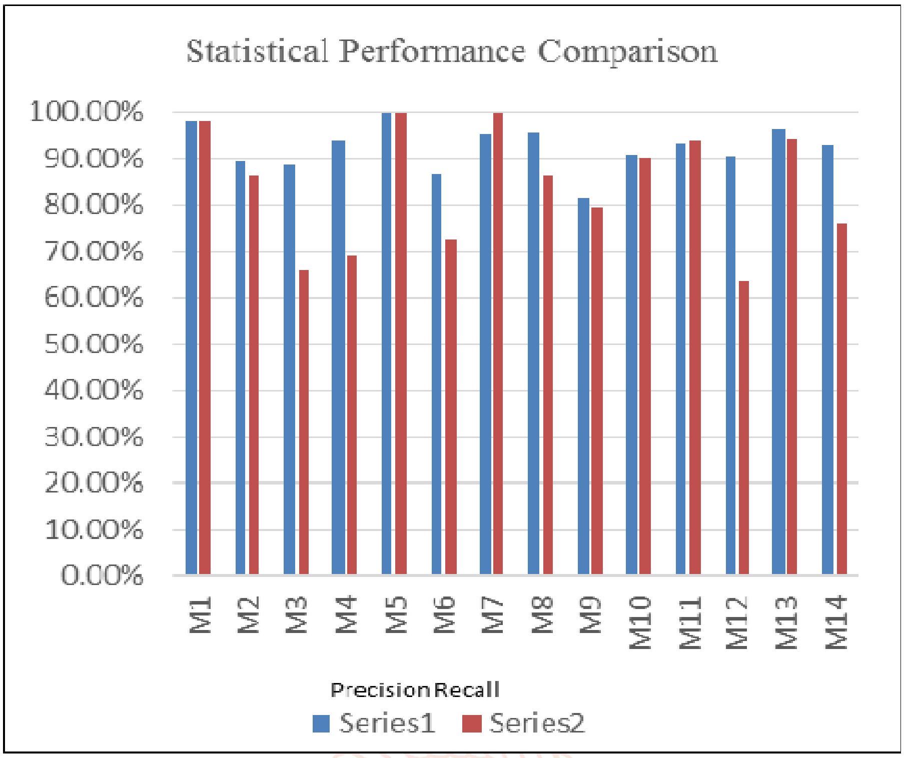

Fig. 7. Statistical Performance Comparison

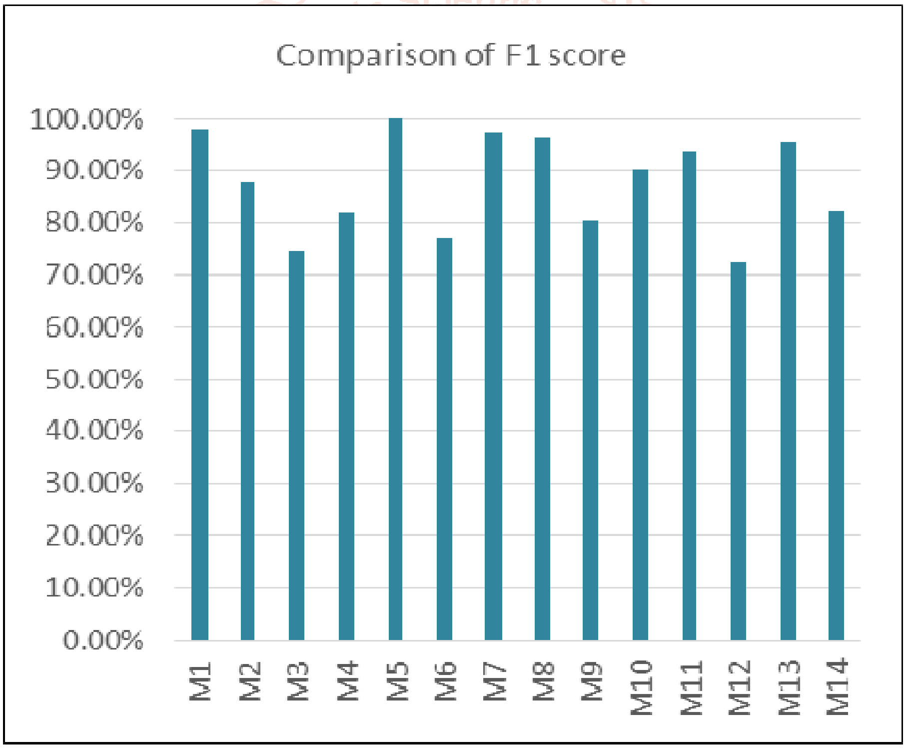

Fig. 8. Comparison of F1 score

Table II Comparative analysis on transitions [16]

Methods	Transition type	
	Abrupt	Gradual
MEV	117	221
CD	49	107
FAST	14	21
SURF	60	108
MSER	31	55
BRISK	3	6
HKD	126	238

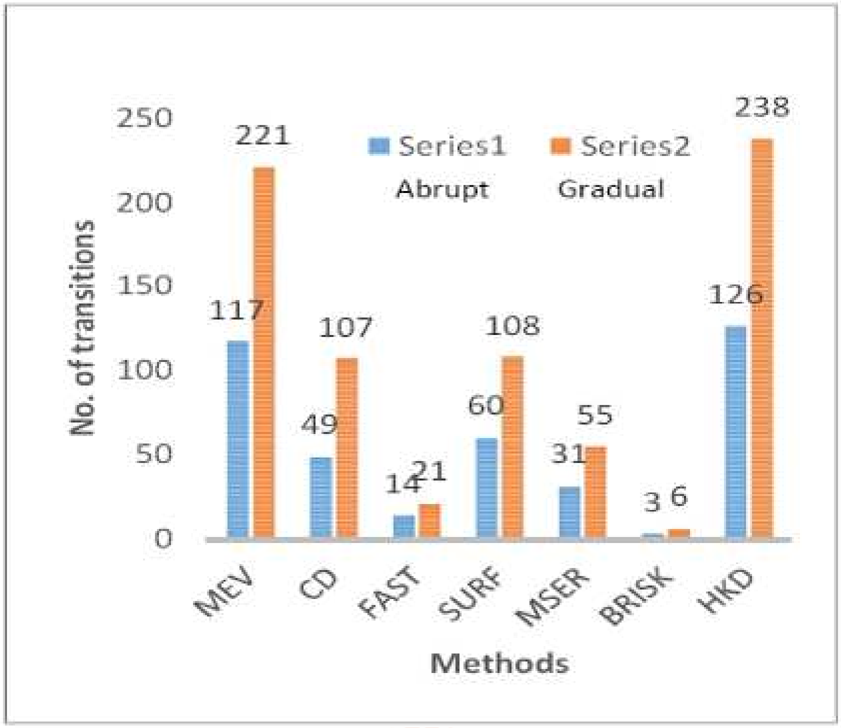

Fig. 9. Comparative analysis on transitions

Performance comparison for validation

Most methods have implemented and validated their results comparing with top performers of TRECVID datasets. The WHT + SVM methods mentioned in review was implemented to achieved task of SBD and have been compared with top performers of TRECVID 2007. Also in addition SVM classifiers were used for evaluation of extracted features, for this the TRECVID 2005 and 2006 datasets were considered. TRECVID 2007 dataset was tested in SVM model, another TSSBD method used the TRECVID 2005 as testing set, also used the untrimmed videos of TRECVID 2005.

Conclusion

This paper contains the review of various methods adopted for SBD and the comparison for the same is made by keeping attention towards precision, recall and F1 score. In this review, several outcomes of different approaches are discussed where they achieved the detection of hard cut and gradual transitions. Different methods using fixed and/or adaptive thresholds are discussed, which may have tradeoff between accuracy, speed and complexity. Also the advantages offered by fully CNN, supervised and unsupervised methods, classifiers are framed here. Several approaches to improve the computation speed also discussed. The contribution made towards detection of shot(s) in presence of flashlight, fades, camera/object motion also discussed here.

Acknowledgment

We are thankful to all the authors mentioned here for motivating us to make research in the field of video processing.

References

[1] N. J. Janwe and K. K. Bhoyar, "Video shot boundary detection based on JND color histogram," 2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013), Shimla, 2013, pp. 476-480, doi: 10.1109/ICIIP.2013.6707637.

[2] V. R. L. Shen, H. Tseng and C. Hsu, "Automatic video shot boundary detection of news stream using a high-level fuzzy Petri net," 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC), San Diego, CA, 2014, pp. 1342-1347, doi: 10.1109/SMC.2014.6974101.

[3] L. P. G. G. and D. S., "Walsh–Hadamard Transform Kernel-Based Feature Vector for Shot Boundary Detection," in IEEE Transactions on Image Processing, vol. 23, no. 12, pp. 5187-5197, Dec. 2014, doi: 10.1109/TIP.2014.2362652.

[4] S. Kim, H. Hong and J. Nang, "A Gradual Shot Change Detection Using Combination of Luminance and Motion Features for Frame Rate Up Conversion," 2015 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Bangkok, 2015, pp. 295-299, doi: 10.1109/SITIS.2015.33.

[5] R. Mishra, C. V. Raman, S. K. Singhai and M. Sharma, "Real time and non real time video shot boundary detection using dual tree complex wavelet transform," 2015 International

Conference on Industrial Instrumentation and Control (ICIC), Pune, 2015, pp. 1495-1500, doi: 10.1109/IIC.2015.7150986.

[6] Hong Shao, Yang Qu, Wencheng Cui, "Shot Boundary Detection Algorithm Based on HSV Histogram and HOG Feature," 2015 International Conference on Advanced Engineering Materials and Technology, Atlantis Press.

[7] Youxian Zheng and Yuan Zhang, "Abrupt shot boundary detection with combined features and SVM," 2016 2nd IEEE International Conference on Computer and Communications (ICCC), Chengdu, 2016, pp. 409-413, doi: 10.1109/CompComm.2016.7924733.

[8] Hannane, R., Elboushaki, A., Afdel, K. et al. "An efficient method for video shot boundary detection and keyframe extraction using SIFT-point distribution histogram",. Int J Multimed Info Retr 5, 89–104 (2016). https://doi.org/10.1007/s13735-016-0095-6

[9] Y. Zheng and Y. Zhang, "GPU-accelerated abrupt shot boundary detection," 2016 16th International Symposium on Communications and Information Technologies (ISCIT), Qingdao, 2016, pp. 141-145, doi: 10.1109/ISCIT.2016.7751609.

[10] Kar, T., Kanungo, P. A motion and illumination resilient framework for automatic shot boundary detection. SIViP 11, 1237–1244 (2017). https://doi.org/10.1007/s11760-017-1080-0

[11] Tyagi V. (2017) Content-Based Image Retrieval Techniques: A Review. In: Content-Based Image Retrieval. Springer, Singapore. https://doi.org/10.1007/978-981-10-6759-4_2.

[12] T. Kar and P. Kanungo, "Motion and illumination defiant cut detection based on Weber features," in IET Image Processing, vol. 12, no. 10, pp. 1903-1912, 10 2018, doi: 10.1049/iet-ipr.2017.1237.

[13] M. Gygli, "Ridiculously Fast Shot Boundary Detection with Fully Convolutional Neural Networks," 2018 International Conference on Content-Based Multimedia Indexing (CBMI), La Rochelle, 2018, pp. 1-4, doi: 10.1109/CBMI.2018.8516556.

[14] C. C. Oprea, R. O. Preda, I. Pirnog and R. A. Dobre, "Video Shot Boundary Detection for Low Complexity HEVC Encoders," 2018 10th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Iasi, Romania, 2018, pp. 1-4, doi: 10.1109/ECAI.2018.8678954.

[15] . L. Wu, S. Zhang, M. Jian, Z. Lu and D. Wang, "Two Stage Shot Boundary Detection via Feature Fusion and Spatial-Temporal Convolutional Neural Networks," in IEEE Access, vol. 7, pp. 77268-77276, 2019, doi: 10.1109/ACCESS.2019.2922038.

[16] S. Karpagavalli, V. Balamurugan and S. R. Kumar, "A Novel Hybrid Keypoint Detection Algorithm for Gradual Shot Boundary Detection," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India, 2020, pp. 1-5, doi: 10.1109/ic-ETITE47903.2020.343.

[17] A. K. Sulaiman and S. A. Mahmood, "Shot Boundaries Detection Based Video Summary Using Dynamic Time Warping and Mean Shift," 2020 International Conference on Computer Science and Software Engineering (CSASE), Duhok, Iraq, 2020, pp. 278-283, doi: 10.1109/CSASE48920.2020.9142116.

[18] Wu Z., Xu P. (2014) A Fast Gradual Shot Boundary Detection Method Based on SURF. In: Wen Z., Li T. (eds) Practical Applications of Intelligent Systems. Advances in Intelligent Systems and Computing, vol 279. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-54927-4_66

[19] J. Yuan et al., "A Formal Study of Shot Boundary Detection," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 2, pp. 168-186, Feb. 2007, doi: 10.1109/TCSVT.2006.888023

[20] Y. Du, C. Jiang and M. Zhou, "A Petri-Net-Based Correctness Analysis of Internet Stock Trading Systems," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 38, no. 1, pp. 93-99, Jan. 2008, doi: 10.1109/TSMCC.2007.896995.

[21] V. R. L. Shen, Y. Chang and T. T. Juang, "Supervised and Unsupervised Learning by Using Petri Nets," in IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, vol. 40, no. 2, pp. 363-375, March 2010, doi: 10.1109/TSMCA.2009.2038068.

[22] Hu, Weiming & Xie, Nianhua & Li, Li & Zeng, Xianglin & Maybank, Stephen. (2011). "A

Survey on Visual Content-Based Video Indexing and Retrieval," IEEE Transactions on Systems, Man, and Cybernetics, Part C. 41. 797-819. 10.1109/TSMCC.2011.2109710.

[23] R. Shen, Y. Lin, T. T. Juang, V. R. L. Shen and S. Y. Lim, "Automatic Detection of Video Shot Boundary in Social Media Using a Hybrid Approach of HLFPN and Keypoint Matching," in IEEE Transactions on Computational Social Systems, vol. 5, no. 1, pp. 210-219, March 2018, doi: 10.1109/TCSS.2017.2780882.

[24] A. Hanjalic, "Shot-boundary detection: unraveled and resolved?," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, no. 2, pp. 90-105, Feb. 2002, doi: 10.1109/76.988656.

[25] E. Hato and M. E. Abdulmunem, "Fast Algorithm for Video Shot Boundary Detection Using SURF features," 2019 2nd Scientific Conference of Computer Sciences (SCCS), Baghdad, Iraq, 2019, pp. 81-86, doi: 10.1109/SCCS.2019.8852603.

[26] X. Qian, G. Liu and R. Su, "Effective Fades and Flashlight Detection Based on Accumulating Histogram Difference," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, no. 10, pp. 1245-1258, Oct. 2006, doi: 10.1109/TCSVT.2006.881858.