JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

CONTENT-BASED VIDEO SHOT BOUNDARY DETECTION USING CANNY EDGE FEATURES

[1]D. Asha, [2]Y. Madhavee Latha

[1]Research Scholar (H.T. No.: 1503PH0832), [2]Professor
[1]Department of Electronics and Communication Engineering,
[1] Jawaharlal Nehru Technological University Hyderabad, Hyderabad, Telangana, India
[2]Department of Electronics and Communication Engineering,
[2] Malla Reddy Engineering College for Women affiliated to JNT University, Hyderabad, Telangana, India

Abstract : Video Shot Boundary Detection (VSBD) is a vital preliminary step in content-based video analysis, enabling efficient indexing, retrieval, summarization, and editing workflows. This paper introduces a robust and computationally efficient method for detecting shot boundaries especially abrupt cuts using the Canny edge features. Our approach employs the following pipeline: frames are converted to grayscale and Canny edge features are extracted. Normalized ECR is computed on consecutive frames, then threshold is set and shot transitions are flagged. We benchmark our method on standard datasets such as TRECVID, evaluating performance in terms of precision, recall, F1-score. Our proposed method average F1-score 93.17.

***IndexTerms* - Video Shot Boundary detection (VSBD), Cut Transition (CT), Gradual Transitions (GT), Edge Change Ratio (ECR)**

I. INTRODUCTION

In today's digital era, the volume of video content encompassing everything from streaming media to social platforms has surged dramatically. Effective organization and retrieval of such massive data demand robust content-based analysis techniques. Video Shot Boundary Detection (VSBD) plays a pivotal role in this context by automatically segmenting videos into distinct, meaningful shots [1]. A shot represents a continuous sequence of frames captured by a single camera, forming the fundamental unit for higher-level analysis.

VSBD is a fundamental process in video analysis that involves segmenting a video into distinct shots. A shot is a sequence of consecutive frames captured without interruption by a single camera, representing a continuous action in time and space. SBD is crucial for various applications, including video editing, summarization, indexing, and content-based retrieval. By identifying the transitions between shots, VSBD enables more efficient and effective video processing and analysis. There are primarily two types of shot transitions. Cut Transitions (CT) are instantaneous changes from one shot to another. Gradual Transitions (GT) include dissolves, fades, and wipes, where the transition occurs over multiple frames [1], shot transitions are shown in Fig.1. The rest of this paper is organized as follows. Section II reviews existing VSBD techniques.
Section III explains the canny edge feature extraction in detail. Section IV escribes the proposed VSBD technique. Section 5 presents and analyzes the results. Finally, Section VI concludes the paper and suggests possible future work.

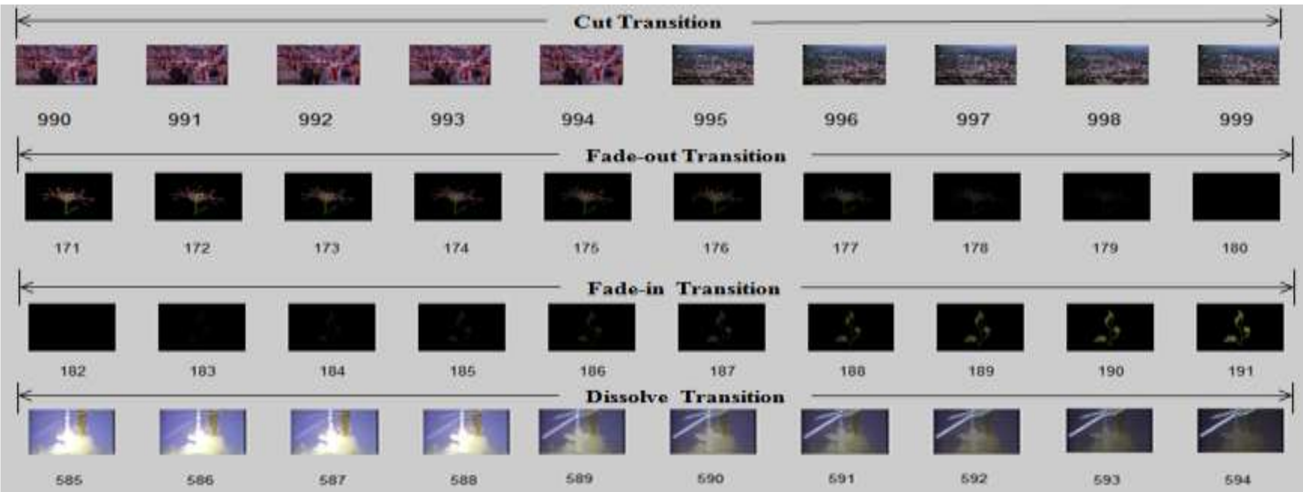

Fig.1 Different Shot Transitions

II. LITERATURE SURVEY

Early and widely used approaches such as Classical (hand-crafted) methods compute frame-to-frame differences using global features such as color histograms, edge histograms, or block-based differences. Simple thresholding on histogram distance or pixel-difference peaks detects many cuts effectively; additional heuristics like temporal smoothing, adaptive thresholds were developed to reduce false positives from camera/object motion and lighting changes. In table. 1 literature survey on VSBD methods is presented.

Table. 1 Literature survey on VSBD methods

Author /Year of Publications	Features	Key Contribution	Dataset Used
Sharma S et al (2024) [2]	dynamic frame-to-frame histogram comparison	adaptive thresholds; emphasizes computational efficiency	Various video datasets (unspecified)
Rashmi BS et al (2021) [3]	Mean Cumulative Sum Histogram (MCSH)	Block-based cumulative sum on fuzzified gradients	General datasets
Asha et al (2019) [4]	Multiple Haar Transform features	Procedure based SBD	TRECVID 2001
Kar T et al (2017) [5]	gradient-orientation feature	adaptive thresholding robust to motion and lighting changes	TRECVID 2001
G Lakshmi Priya et al (2014) [6]	Walsh-Hadamard transform	Procedure Based shot transition Identification process (PBI)	TRECVID 2007
Z M Lu et al (2013) [7]	SVD and pattern matching	Fast Shot boundary detection	TRECVID 2001
G L Priya et al (2013) [8]	Split-and-merge scheme guided by Fisher Linear Discriminant (2D² FLD)	texture region features and spectral clustering for dimensionality reduction	Cricket video (case study)

III. CANNY EDGE FEATURE EXTRACTION

The Canny edge detector is an optimal multi-stage algorithm designed to detect edges with low error rate, accurate localization and single response per edge. Its popularity stems from both its theoretical robustness and practical effectiveness in various computer vision systems. Canny is a foundational edge detection algorithm in image processing and computer vision [9]. Its algorithmic clarity and balance between precision and robustness make it ideal for a variety of applications ranging from object recognition to scene understanding. The Block diagram of canny edge detection is shown in fig. 2.

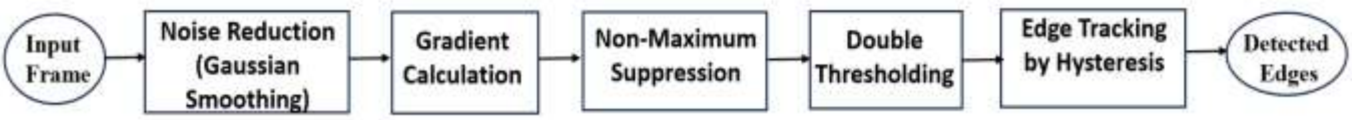

Fig.2 Block Diagram of Canny Edge Detection

- **Noise Reduction (Gaussian Smoothing):** The input image is first smoothed using a Gaussian filter to minimize the impact of noise.
- **Gradient Calculation:** The smoothed image is processed with Sobel operators to compute horizontal (G_x) and vertical (G_y) gradients. These are used to calculate both gradient magnitude (G) and direction (θ) as in (1) &(2).

$$G = \sqrt{|G_x|^2 + |G_y|^2} \tag{1}$$

$$\theta = tan^{-1}\left(\frac{G_y}{G_x}\right) \tag{2}$$

$$where\ G_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}, G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

- **Non-Maximum Suppression:** To refine edge detections, this step suppresses non-maximal pixel values, ensuring only the local maxima along the gradient direction remain producing thin, precise edges.
- **Double Thresholding:** Using two thresholds (high and low), pixels are categorized as strong edges, weak edges, or non-edges. Strong edges are retained, weak edges are conditionally retained, and the rest are suppressed.
- **Edge Tracking by Hysteresis:** Weak edges are preserved only if they connect to strong edges. This ensures continuity of true edges while eliminating isolated noise.

IV. PROPOSED CONTENT-BASED VIDEO SHOT BOUNDARY DETECTION USING CANNY EDGE FEATURES

Edge Change Ratio (ECR) is a widely recognized feature used to detect shot boundaries particularly effective for hard cuts and some types of gradual transitions. ECR provides a structurally informed signal, it captures geometric content changes, which often align closely with actual scene segmentation [10-12]. The Flowchart of proposed method is shown in Fig.3 and algorithm is given below:

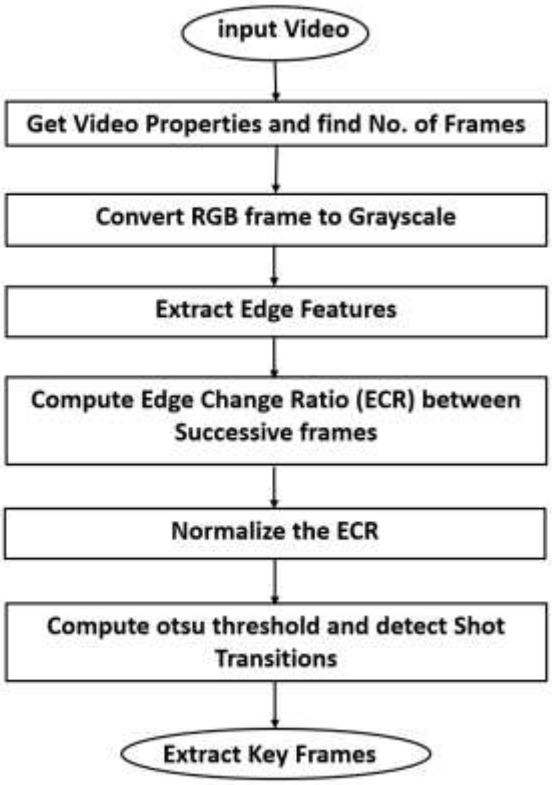

Fig.3 Flowchart of Proposed Content-Based Video Shot Boundary Detection

Algorithm of Proposed Content-Based VSBD method

Step 1: Read a video and calculate the No. of Frames
Step 2: convert RGB frame to grayscale
Step 3: Apply an edge detector (e.g., Canny) to consecutive frames.
Step 4: Calculate Edge Changes Entering Edges (ECR_{in}) and Exiting Edges (ECR_{out}) as in (3).
Where Edges (ECR_{in})=Edges present in the current frame but absent in the previous frame.
Edges that exist in the previous frame but vanish in the current frame are called Exiting Edges (ECR_{out}).
Step 5: Compute Edge Change Ratio (ECR)

$$ECR_n = \max\left(\frac{ECR_{in} in\ F_n}{Total\ Edges\ in\ F_n}, \frac{ECR_{out} from\ F_{n-1}}{Total\ Edges\ in\ F_{n-1}}\right) \tag{3}$$

$where\ ECR_{in} = Edges\ present\ in\ the\ current\ frame\ F_n\ but\ absent\ in\ previous\ frame\ F_{n-1}$
$ECR_{out} = Edges\ that\ exist\ in\ the\ previous\ frame\ F_{n-1}\ but\ vanish\ in\ the\ current\ frame\ F_n$
$n = 1,2, \dots \dots N\ and\ N\ is\ Total\ No.of\ frames\ in\ video$
Step 6: Normalize ECR
Step 7: Compute otsu threshold from ECR values. Otsu's method is a data-driven technique that selects an optimal threshold by minimizing intra-class variance (or equivalently maximizing inter-class variance).
Step 8: Detect Shot transitions and Extract Key Frames.

V. Experimental Results and Discussions

The performance of our proposed Content-Based VSBD is tested on TRECVID dataset [13-15]. We implemented our proposed method in MATLAB 7.6 on the Windows Xp Professional with CPU of 2.10 GHz. Our method is evaluated by calculating precision (P), Recall (R) and F1-Score (F_1) as in (4)-(6). Canny edge features are shown in fig.4. VSBD Plot and extracted key frames for documentary video are shown in Fig. 5 & Fig. 6 and for News video are shown in Fig.7 & Fig.8. in table. 2 results of proposed VSBD method for 8 videos are shown. For documentary video the threshold value is 0.4933 and it detected 3 shot transitions and extracted 4 key frames. For news video the threshold value is 0.5049 and it detected 11 shot transitions and extracted 12 key frames. Our proposed method is compared with VSBD using SVD [7] and shown in table.3 and bar graph is shown in Fig.9.

$$P\ (\%) = \frac{N_{Correct}}{N_{Correct}+N_{False}} * 100 \tag{4}$$

$$R(\%) = \frac{N_{Correct}}{N_{Correct}+N_{Missed}} * 100 \tag{5}$$

$$F_1 - Score(\%) = 2 * \frac{P*R}{(P+R)} * 100 \tag{6}$$

Fig.4 Canny Edge Features

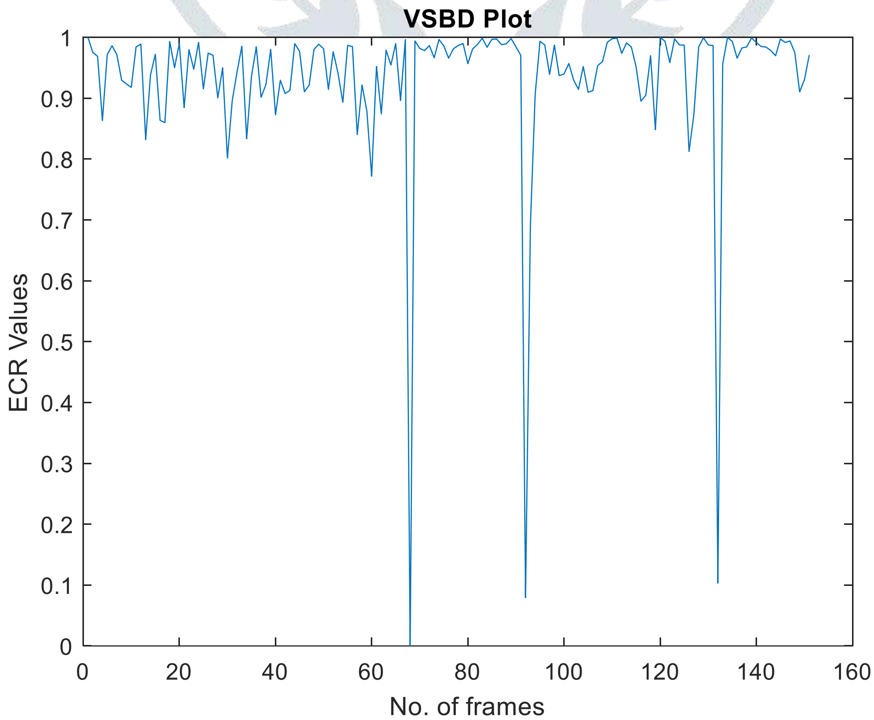

Fig.5 VSBD Plot for Documentary Video.

Fig. 6 Key Frames extracted for Documentary Video.

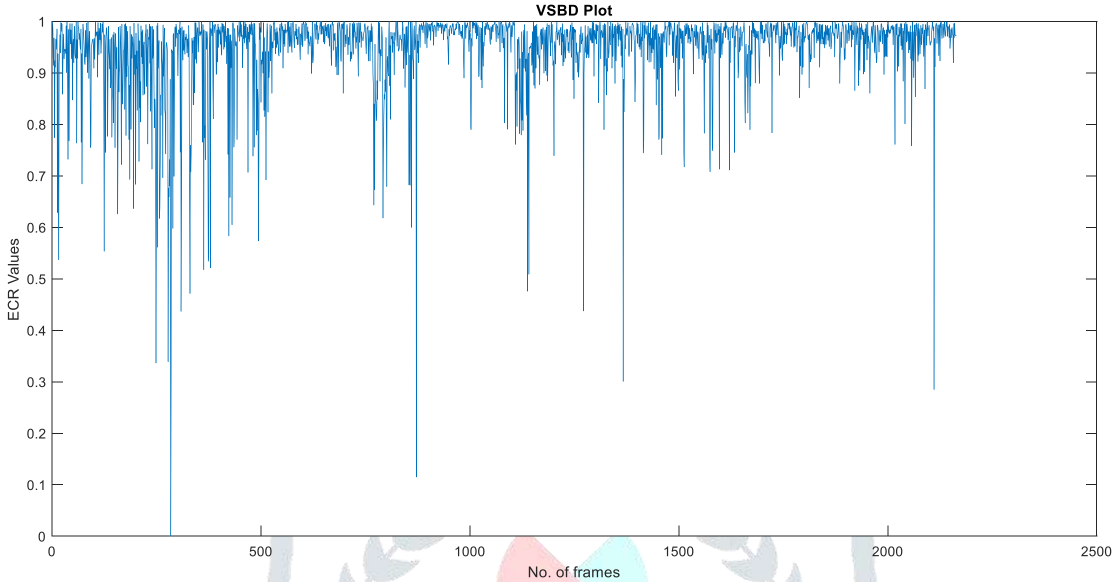

Fig. 7 VSBD Plot for News Video.

Fig. 8 Key Frames extracted for News Video.

Table. 2 Results of VSBD using ECR on 12 Dataset.

Dataset	Videos	No. of Frames	No. of Shot Transitions	P (%)	R (%)	F_1-Score (%)
TRECVID 2001 test Data [13], IAAC2 2013[14], Videos downloaded from Google [15]		3525	4	100.00	100.00	100.00
		2164	11	90.91	90.91	90.91

		1101	12	90.91	83.33	86.96
		2762	7	85.71	75.00	80.00
		126	4	100.00	100.00	100.00
		2970	5	100.00	100.00	100.00
		151	4	100.00	100.00	100.00
		1726	8	87.50	87.50	87.50
Averages				94.38	92.09	93.17

Table.3 Comparison chart of proposed VSBD using ECR with VSBD using SVD

Method	Average Precision (%)	Average Recall (%)	Average F1-Score (%)
VSBD using SVD [7]	92.34	84.57	86.29
Proposed VSBD Method	94.38	92.09	93.17

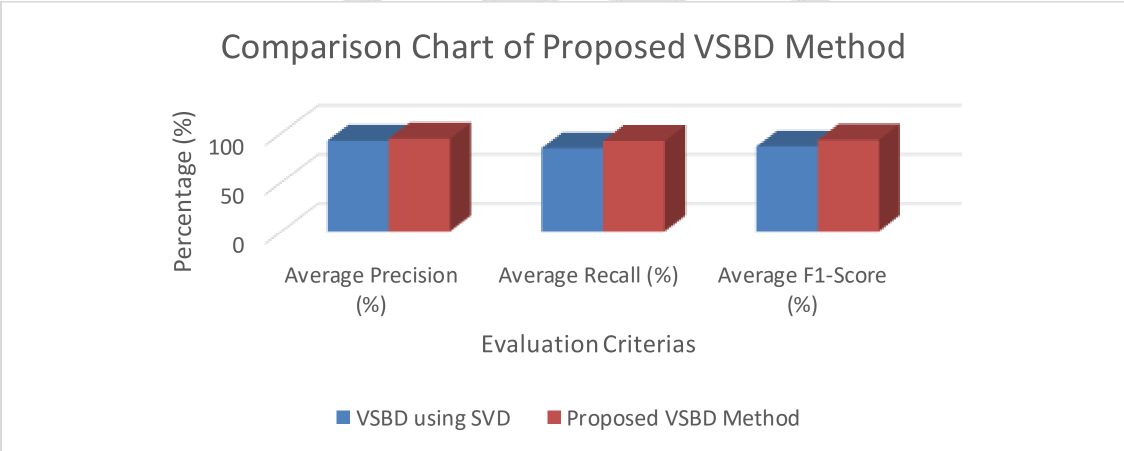

Fig. 9 Bar Graph of Comparison chart of proposed VSBD using ECR with VSBD using SVD

VI. Conclusion

In this work, we have presented a content-based VSBD method using Canny Edge Features focusing particularly on abrupt cuts. Our pipeline, grounded in efficient Canny edge extraction and ECR computation, demonstrates that even a single, well-designed structural cue can achieve strong performance, with an average F1-score of 93.17% on the standard TRECVID dataset. In future we can still improve by feature Fusion with Learning-based and Multi-modal Models.

References

[1] Kar, T., Kanungo, P., Mohanty, S.N. et al. Video shot-boundary detection: issues, challenges and solutions. Artif Intell Rev 57, 104 (2024). https://doi.org/10.1007/s10462-024-10742-1

[2] Sharma, S., & Malakar, P. K. (2024). Threshold-Based Video Shot Boundary Detection: A Comprehensive Approach for Transition Identification. ShodhKosh: Journal of Visual and Performing Arts, 5(1), 3047–3053. https://doi.org/10.29121/shodhkosh.v5.i1.2024.3248

[3] Rashmi BS, Nagendra swamy HS (2021) Video shot boundary detection using block based cumulative approach. Multimed Tools Appl 80:641–664

[4] Asha D., Madhavee Latha Y. (2019) Content-Based Video Shot Boundary Detection Using Multiple Haar Transform Features, Soft Computing and Signal Processing. Advances in Intelligent Systems and Computing, vol 900. Springer, Singapore. https://doi.org/10.1007/978-981-13-3600-3_67.

[5] 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

[6] G. Lakshmi Priya, S. Domnic, Walsh-Hadamard transform Kernel-based feature vector for shot boundary detection. IEEE Trans. Image Process. 23(12), 5187–5197 (2014)

[7] Z.-M. Lu, S. Yong, Fast video shot boundary detection based on SVD and pattern matching. IEEE Trans. Image Process. 22(12), 5136–5145 (2013)

[8] G.L. Priya, S. Domnic, Video cut detection using block-based histogram differences in RGB color space, in International Conference on Signal and Image Processing, pp. 29–33 (2010)

[9] Refael C. Gonzalez, Richard E. Woods, Digital Image processing, Second edition, Pearson.

[10] X. Liu, J. Sun, J. Liu, J. Sun and J. Liu, "Shot-based temporally respective frame generation algorithm for video hashing," 2013 IEEE International Workshop on Information Forensics and Security (WIFS), Guangzhou, China, 2013, pp. 109-114, doi: 10.1109/WIFS.2013.6707803.

[11] Guru, D.S., Suhil, M., Lolika, P. (2013). A Novel Approach for Shot Boundary Detection in Videos. In: Swamy, P., Guru, D. (eds) Multimedia Processing, Communication and Computing Applications. Lecture Notes in Electrical Engineering, vol 213. Springer, New Delhi. https://doi.org/10.1007/978-81-322-1143-3_17

[12] D. Asha, Y. Madhavee Latha, V. S. K. Reddy, “Content Based Video Retrieval system using Multiple Features”, International Journal of Pure and Applied Mathematics (IJPAM), Vol. 118, No.14, February 2018, pp 287-294, https://acadpubl.eu/jsi/2018-118-14-15/articles/14/42.pdf.

[13] TREC Video Retrieval Test Collection [online] (2001), Available on website: http://trevid.nist.gov/ and www.open-video.org

[14] TRECVID. April 12, 2019 [Online]. Available: https://ir.nist.gov/tv2013/iacc.2/iacc.2.a/.

[15] Google video.URL http://video.google.com/