RESEARCH ARTICLE

INTERNATIONAL RESEARCH JOURNAL OF MULTIDISCIPLINARY TECHNOVATION

Asian Research Association

DOI: 10.54392/irjmt26320

A Robust Shot Detection Method for Regional Language Videos using Edge and Histogram Statistics

Avani Bhuva [a], Dhirendra Mishra [a *]

[a] Department of Computer Engineering, MPSTME-NMIMS, Mumbai, 400056, India

* Corresponding Author Email: dhirendra.mishra@nmims.edu

DOI: https://doi.org/10.54392/irjmt26320

Received: 04-09-2025; Revised: 29-04-2026; Accepted: 11-05-2026; Published: 22-05-2026

Abstract: The huge amount of digital content is being uploaded by users every day, and processing these videos specifically for indexing and retrieval purposes is extremely important. This research is contributing to society by providing a baseline approach for Gujarati news channel video retrieval for regional users. The Gujarati language is spoken by more than 22 million people worldwide. To retrieve these videos correctly, we need to process them frame by frame; hence, shot segmentation is an essential component, followed by keyframe extraction, and, lastly, extracting text features for the retrieval of news videos. A three approaches were developed to detect shot boundaries and identify frames containing meaningful Gujarati text in broadcast news video. The three methods were: (1) An Shot Boundary Detection (SBD) methods based on global threshold (SBD-GT) with 5 distinct frame level features (2) SBD methods employing statistical adaptive thresholding (SBD-AT) with 5 distinct frame level features (3) Development of the proposed Edge Histogram Adaptive Shot Detection (EHASD) system with the fusion of edge and histogram based features. EHASD applied Statistical Edge-Histogram Thresholding (SEHT) together to evaluate edge deviation and histogram difference to improve the detection rate of both abrupt and gradual transitions. The proposed research was evaluated using a TV9 Gujarati news video dataset with 45920 total frames from different categories like weather, Cricket, and politics. The proposed approach, EHASD, achieved a precision of 93.25% and a recall of 92.01%, surpassing the values obtained by the global and adaptive thresholds in SBD approaches. Additionally, the z test & paired t test was also performed on the mean and standard deviation to validate the results of the proposed approach. In addition, the research confirmed that the proposed method can be used further to perform key frame selection followed with text extraction for the video retrieval purpose.

Keywords: Shot Boundary Detection (SBD), Edge Histogram Adaptive Shot Detection (EHASD), Statistical Adaptive Thresholding, Edge Change Ratio and Histogram Difference, Frame-level features.

1. Introduction

The rapid growth of digital video content across news channels and online platforms poses significant challenges for efficient video indexing and retrieval. News channel videos generate a large volume daily, making manual annotation and frame-level processing difficult and computationally expensive. Therefore, automated content-based video retrieval (CBVR) systems have become crucial for handling large video archives [1]. Significant research has been conducted on global languages, such as English and Chinese. Hence, video retrieval systems remain comparatively underdeveloped for the regional languages. In India, there are a total of 22 regional languages and 912 private satellite [2, 3] for regional news video with large volumes. Despite the lack of a framework for automated video retrieval approaches based on user queries. A single broadcast video [4] may consist of thousands of frames, which contain a lot of information. Processing such information is very important to retrieve the relevant video [1, 3]. To address this issue, video segmentation techniques are developed that analyze video frames by dividing them into small units, called shots, and extracting frame features. Shot Boundary Detection (SBD) approaches are used to identify these boundaries, from which we can select the key frame; from the key frame, we can extract text features; and, lastly, we can generate a retrieval index. Effective SBD reduces redundancy, improves efficiency and increases the visibility on relevant frames [5-8].

In broadcast news videos, on-screen textual information is a very important for understanding the content. Like captions, headlines, tickers, and location deliver crucial information that cannot be inferred directly from visual content. This is particularly important for regional languages such as Gujarati, which is spoken by more than 22 million people worldwide. Gujarati news

videos gives textual overlays by providing primary contextual information on events like weather updates, Cricket matches and political developments. However, extracting meaningful text from such videos is challenging due to the complexity of Gujarati text, which uses different font styles and has more vowels (14) and consonants (34) than the English language [3]. The Gujarati news channel videos has background clutter, frequent scene transitions and camera movements. These challenges underscore the importance of accurate shot segmentation to ensure that text extraction is applied only to visually meaningful, text-rich frames [4-9]. Many conventional methods employ Global thresholds to identify shot transitions and single-frame-level features [10-18]. Although computationally efficient, Global threshold-based approaches [19, 20] are highly sensitive to variations in video content and often fail to generalize across different video categories. Broadcast news videos exhibit diverse visual characteristics, including studio shots, outdoor scenes, rapid camera movements, and scrolling text overlays. As a result, a single global threshold often leads to missed detections or excessive false positives.

To overcome these limitations, initially, we have implemented basic SBD approaches with a global threshold and extracted 6 distinct frame features. The SBD approaches again experimented with adaptive thresholding. Frame statistical adaptive thresholding techniques have been implemented, in which thresholds are dynamically computed using statistical measures such as the mean and standard deviation of feature differences. Adaptive methods improve robustness for different categories. However, many adaptive approaches that rely on a single frame feature [18, 21-26] handle both abrupt and gradual transitions in complex news videos with changes in illumination, background, or text overlays, but may not align with the actual shot boundaries. These shots are highly susceptible to issues such as camera motion, object movement, or lighting variations, leading to severe segmentation [27]. With respect to these limitations, this research presents a three-stage framework for shot boundary detection of Gujarati news videos. In the first stage, SBD methods with global thresholds are used to create baseline performance. The second stage assesses SBD techniques using statistical adaptive thresholding to analyze performance relative to the first-stage SBD-GT. In the third stage, we introduced a novel Edge Histogram Adaptive Shot Detection (EHASD) framework for regional-language news content. The proposed EHASD method extracts edge and histogram features with the Statistical Edge-Histogram Thresholding (SEHT) framework. Edge features effectively capture structural changes between consecutive frames, while histogram differences capture changes in intensity and color distribution across video frames. By jointly analyzing these features with an adaptive statistical threshold, EHASD can detect both abrupt and gradual transitions while reducing false detections. The proposed framework will improve retrieval accuracy while reducing computational overhead, thereby addressing a critical gap in regional-language CBVR research is extremely important.

The rest of this paper is organized as follows. Section 2 is about related work on shot boundary detection approaches. Section 3 details the implementation of existing shot boundary detection methods, including Global threshold approaches, Statistical Edge-Histogram Thresholding and the proposed EHASD method. Section 4 presents the experimental results, including datasets, evaluation criteria for existing SBD with Global/global thresholds, existing SBD with statistical thresholds, the proposed EHASD, and, lastly, a z-test statistical and validation of the proposed research. Section 5 conclusion and future scope.

2. Literature Review

SBD has gained considerable attention in video content analysis and has become an essential step for applications like video summarization, retrieval, classification, and text extraction. The traditional approaches use basic features like edge, HOG, Color [11, 13, 14]. Currently the different Deep Learning approaches to perform the video segmentation to detect the Deep features [21-23] and to detect the shots as mentioned by the researchers [15, 16, 24-26].

2.1 Feature Based Approaches

Early studies on shot boundary detection (SBD) approaches depend on changes in pixel intensity, color distributions, or other features between consecutive frames. These approaches generate relevant shots by comparing frames using methods such as frame differencing, histogram intersection, and chi-square distance. [4-5]. Histogram-based techniques examine differences in color distributions across neighboring frames as correctly highlighted by T. Kar *et al.* [8], Gargi *et al*. [9] and Y. Bendraou *et al.* [10]. However, these approaches mainly used Global thresholds and Otsu thresholding [20, 21] to detect shots, as shown in Table 1. To overcome these limitations, the researchers moved to feature-based approaches that exploited characteristics like edges, textures, etc. The different features like edge and histogram [11, 13, 14] and deep features [21-26, 28, 29] are extracted using feature extraction approaches. The Edge Change Ratio (ECR) method efficiently detects significant structural changes between frames by analyzing edge features. While ECR provided greater flexibility for minor lighting changes, it required proper threshold settings to avoid false detections. A combined approach using color histograms and Histogram of Oriented Gradients (HOG)

has shown potential results for video summarization and retrieval by reducing redundant frames.

R. M. Bommisetty *et al*. [30] and E. M. Saoudi *et al.* [31] have inspected the basic concepts of Shot Boundary Detection (SBD), focusing on techniques to enhance text extraction accuracy in video content analysis. Edge-based methods can be affected by noise or camera movement, leading to false detection. To address various video content, researchers have explored multiple shot detection strategies, including color histogram analysis, block motion matching and the MPEG approach to compressed data, each of them will be useful for specific transition types and content characteristics [32]. Li *et al*. [33] Proposed a multilevel color histogram difference approach for shot boundary detection, which improves detection accuracy by capturing gradual and abrupt transitions; however, performance is sensitive to threshold selection and lighting variations. V. L. Narla *et al.* [34] have introduced an MFCC and DTW-based method for signal analysis, demonstrating effective feature matching capabilities; however, its applicability is limited to audio signals and does not directly address visual shot boundary detection.

2.2 Thresholding Techniques

Threshold plays an important role in extracting shots from a video [5-8]. P. K. Sahoo et al. [20, 27, 32] discussed the role and importance of threshold in image segmentation. Image segmentation is the primary approach that is being used for text extraction, region extraction and in many more applications. There are different types of thresholding techniques, like bimodal, Otsu, local, and global thresholding [8, 20, 27]. Few drawbacks of Global-thresholding methods as mentioned in [8], including the need for adaptive or hybrid techniques that can respond to the varying characteristics of video content, and highlighted the importance of integrating fusion features with statistical thresholds. Research papers [7, 18-20] have applied the SBD concept to TRECVID datasets, using both Global and adaptive thresholds. Adaptive threshold methods use local statistical metrics, such as the mean and standard deviation of frame differences. Although adaptive techniques provide some flexibility, they remain vulnerable to false detections in highly textured or graphically elaborate content, such as news broadcasts with overlays and banners. Nevertheless, many of these approaches are either content-agnostic or overly specialized for specific datasets, limiting their applicability across diverse video genres, particularly in low-resource languages such as Gujarati.

2.3 Learning-Based Approaches

SBD approaches have now been adopted by Machine learning, especially with various deep learning methods. Researchers have turned to Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to extract spatial and temporal features from annotated datasets [28]. These techniques set a high bar for accuracy but require significant computational resources and substantial labeled data, which are often unavailable for regional languages like Gujarati or for older broadcast archives. Esteve Brotons *et al*. [22] used a CNN model to process multiple frames at one go to capture spatiotemporal patterns. Between 2022 and 2024, numerous new and advanced deep learning models have significantly enhanced the accuracy and speed of video retrieval systems [28, 30, 31]. A recent study combined a 3D CNN with depth-wise convolutions, achieving an F1 score of 93% on the Clip Shots dataset. Soucek *et al.* [28] have applied TransNet V2, a transformer-based model with attention mechanisms, to the same dataset. With an F1 score of 77.9%, representative robust performance in capturing both abrupt and gradual transitions.

The team of Abdul Hussain *et al*. [32] conducted a literature review and listed the key challenges with respect to shot boundary detection approaches. Traditional performance metrics like Euclidean, Bhattacharyya and Manhattan distances are used to detect the shot. As T. Kar [8] *et al*. and Gargi *et al.* [9] mentioned the system performance depends on metrics such as Recall, Precision and F1 scores and have evaluated the performance of the SBD algorithm's performance. Also categorizing SBD methods into pixel-based, feature-based, transform domain and machine learning-driven approaches. The detailed literature is mentioned in the table 1, which focuses on extraction frame features and evaluates the performance of correctly identified shots.

The findings are: sensitivity to camera or object movement, sudden changes in lighting and the difficulty for gradual transitions like fades or dissolves. They have also mentioned important directions for future research, but mostly focused on general video datasets, overlooking region-specific or language-diverse content such as the regional language Gujarati. While they mentioned the potential for merging handcrafted and deep features, they didn ' t provide any specific strategies for handling the complex structure, including broadcast footage or any video category in low-resource languages. Existing research indicates that feature-based approaches are effective at detecting scene changes but are heavily dependent on Global thresholds, which can cause them to fail when lighting or motion changes occur in the scene. Compared with the adaptive and hybrid thresholds, the adaptive and hybrid thresholds may better adjust to local differences, though their performance still needs to be measured on regional videos.

Table 1. Literature Review on feature-based approaches [9, 10, 13, 14, 16, 20, 29, 32]

Dataset	Feature Approach	Threshold		Transition types		Accuracy
		Other	AT	HT	GT	
IACC3 datasets	CNN,Self-Attention	-	√	-	-	Transnet-94.4%, TransnetV2-9 2.7%
BBC RAI	CNN, Self-Attention	-	√	-	-	Transnet- 91.1%, TransnetV2- 91.50%
Sports news video	MPEG-based shot-detection algorithms	Twin threshold	-	-	√	90-70%
Different category of videos	Histogram Based	√	-	-	-	100% recall
TRECVID IACC.3	TransNet V2	√	-	√	√	-
Different categories of video	Histogram Based	√	-	√	-	-
TRECVID	Histogram Based	√	-	√	-	Precision 84.7% Recall-80.6%
Different categories of video	Histogram Based	√	-	√	-	Precision-82.40% Recall-98.10%
TRECVID	Histogram Based	-	√	√	-	-
Different categories of video	Histogram Based	√	-	√	-	-

Recent deep learning models, specifically CNNs and transformers, have significantly improved performance accuracy; however, they require large, labeled datasets, which are not readily available for languages like Gujarati. So, the main problem remains the same: delivering accurate, efficient shot boundary detection across all kinds of video with large amounts of data. These research gaps suggested the need for efficient shot detection and multi-feature frameworks. That's where the proposed Edge Histogram Adaptive Shot Detection (EHASD) method comes in to address the challenges mentioned by Kar *et al*. [8] and aims to improve SBD performance for regional language video analysis.

2.4 Unique Contribution to Underlying Issues and Challenges

This research introduces the Edge Histogram Adaptive Shot Detection (EHASD) method with targeting the exact issues found by the researchers as mentioned correctly in their research work [5-8,20,27,32].

2.4.1 Edge and Histogram Fusion Feature Approach

EHASD combines the Edge Change Ratio with a histogram-based approach to detect structural transitions in video and the distribution of pixel intensity, thereby reducing false positives. This fusion works particularly well with complex joins, graphics and the gradual transitions you frequently see in broadcast news videos. By developing localized edge histograms based on Sobel edge gradients, this method can track structural changes in video frames rather than pixel or color changes.

2.4.2 Domain-Specific Evaluation for Regional Content

Majorly, shot boundary detection approaches have been validated on large global datasets, such as TRECVID. EHASD, on the other hand, the proposed EHASD worked on the TV9 Gujarati news footage targeting three domains: weather, cricket, and politics. The aim is to achieve high precision across both abrupt and gradual shot changes in region-based video.

2.4.3 Statistical Edge Histogram Thresholding

Global thresholds can be inflexible and unpredictable, as we have to select different threshold values for the various SBD approaches, which may or may not work effectively for the dataset. The proposed EHASD introduces Statistical Edge-Histogram Thresholding, which calculates the mean and standard deviation within the frame based on edge and histogram changes.

2.4.4 Enhanced Segmentation through Multi-Feature Fusion

Traditional approaches used ECR and Bhattacharyya-based histograms had success but they struggled when scenes got dynamic. EHASD's fusion of local contrast and structure provides a stronger, more reliable segmentation solution, which will support a better key frame selection approach for tasks like text extraction and video retrieval.

3. Methodology

A framework was proposed to detect the no of distinct shots which will further help to select the best key frame with Gujarati text in order to achieve the best key frame selection. As shown in Figure 1, Gujarati news channel videos are fed into the pipeline throughout the methodology, producing final outcomes. The design of this multi-stage framework is to extract k. no of most significant shots. A comprehensive explanation of the proposed diagram is provided in the following section, organized stage by stage to reflect each step of the process.

3.1 Dataset and Experimental Setup

The experiments were conducted on the TV9 Gujarati News Video Dataset, which includes videos from three major categories: weather, Cricket and political news. Frames from each video are extracted, and each frame was converted with resolutions of 640 × 360 to 1280 × 720 pixels.

The dataset consists of both short and long duration clips, ranging from 120 to 511 seconds. A total of 45920 video frames were processed and identified with Gujarati text such as "વરસાદ" (rain), "ક્રિકેટ" (cricket), and "મોદી" (Modi), which were manually verified to support the text detection and retrieval stages. This dataset provides a variety of combinations of visual and linguistic content, making it well-suited for evaluating the effectiveness of the proposed EHASD framework for regional-language video segmentation. Table 2 describe of the TV9 Gujarati News Video Dataset used for evaluating the proposed EHASD framework.

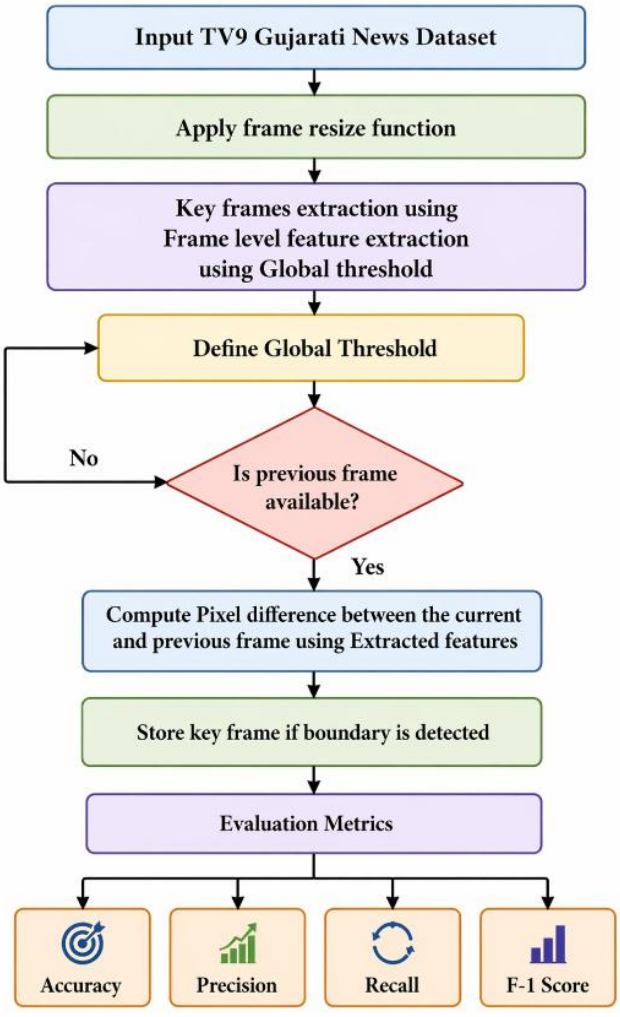


Figure 1. Methodology framework for Shot Detection and Key Frame Selection.

Table 2. Description of the TV9 Gujarati News Video Dataset used for evaluating the proposed EHASD framework [35]

Video Label	Number of frames	Frame Size (Width × Height)	Video Duration (Sec.)
weather-1	2996	1280x720	120
weather-2	4618	1280x720	183
Cricket-1	7107	1280x720	266.4
Cricket-2	7883	640x360	309
politics-1	13341	1280x720	511.8
politics-2	9975	640x360	383.4
Total	45920	--	1773.6

Ground-Truth Shot Boundary Annotation Methodology:

1. Annotation Source and Type: Ground truth annotations pertaining to Gujarati words are used for the purpose of validating the shot detection framework. Each category of video is associated with the frequency of occurring Gujarati text, which has been manually identified from the frames. The manual annotation is necessary because there are no preexisting shot labels for the video dataset.
2. Annotation Procedure: The videos were examined during a shot boundary detection. Annotated transitions include abrupt cuts between studio and field scenes and switches between anchor, reports, debates and during full-screen graphics
3. Use of Ground Truth in Evaluation: The finalized annotations were used as the reference ground truth for evaluating detected shots across all Gujarati text, which will be further useful for video retrieval purposes and to calculate precision, recall, and F1-score.

These language annotations help as semantic ground truth references for evaluating the accuracy of detected shots in relation to relevant textual content within the TV9 Gujarati news dataset.

3.2 Methodology and Implementation of Shot Boundary Detection Approaches

This section describes the complete methodology for detecting shot boundaries and mapping Gujarati text from the ground-truth table 3 of Gujarati news videos. The implementation has been organized into three stages: the first stage focuses on understanding the basic structure of video data and applying frame resizing for uniform processing, the second stage explains the use of existing Shot Boundary Detection techniques with the use of Global and adaptive thresholds and the final stage introduces the proposed Edge Histogram Adaptive Shot Detection framework, which integrates SEHT and feature fusion to improve detection accuracy to reduce redundant shots. Each step is explained in the below section.

Fundamentals of video representation:

Let the video be represented as a continuous function

$$V(t) = F(x, y, t) \quad (1)$$

where V(t) is the video function depending on time t, F(x,y,t) is the intensity at pixel coordinates (x,y) at time t, x and y are the spatial coordinates of the image. Extracting frames at regular intervals with a frame rate of f frames per second and the time between two consecutive frames is calculated by:

$$\triangle \mathrm{t} = 1/\mathrm{f} \quad (2)$$

The set of extracted frames $F_k(x,y)$ at times can be expressed as $F_k(x,y)=F(x,y,t_k)$.

$$\mathrm{t_k} = \mathrm{k} \cdot \Delta \mathrm{t} \quad (3)$$

Δt is the time interval between consecutive frames k is the frame index and N is the total number of extracted frames.

3.2.1 Frame resize function

The frame resize approach helps standardize input dimensions, especially during model training. Which is essential for the video processing, OCR and text extraction process. To improve processing speed, reduce memory usage, enhance algorithm performance, and reduce noise & irrelevant details, we have used the frame resize function. As we can see in the table 2, the frame size is different in each category of the video, so the first stage is to make all the frames the same size so the processing of the frames will be easier and faster. In this stage, all video frames have been resized to 640x480.

3.2.2 Global Threshold-Based Evaluation of Existing Shot Boundary Detection Approaches

`In this research, we conducted experiments on existing SBD approaches to validate and assess the performance of our Video dataset.

Table 3. ground truth gujarati word annotations for shot detection evaluation

Video Category	Gujarati Words Identified	English Translation / Meaning	Contextual Category
Weather Videos	વરસાદ, ગરમી, બરફ, વરસાદી	Rain, Heat, Ice, Rainy	Weather Reports and Forecasts
Cricket Videos	રમત, ક્રિકેટ, ઇન્ડિયા, ઇન્ડિયન, વર્લ્ડ કપ, ક્રિકેટર, ટીમ, ભારત	Play, Cricket, India, Indian, World Cup, Cricketer, Team, Bharat	Match Highlights
Politics Videos	રાજકારણ, ભાજપ, વડાપ્રધાન, મોદી, નરેન્દ્ર મોદી, કોંગ્રેસ	Politics, BJP, Prime Minister, Modi, Narendra Modi, Congress	Political News and Debates

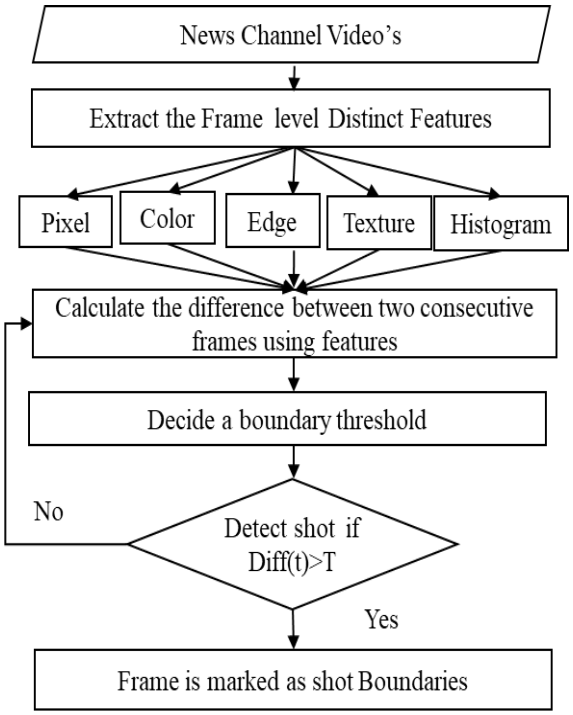


Figure 2. Flowchart for detecting Shot Boundaries by extracting frame level features using global threshold

The Pixel Difference Method, Edge Change Ratio, Color Difference, Gabor SBD and HBA Bhattacharya Distance (BD) approaches have been considered for experimental purposes. The conventional Shot Boundary Detection (SBD) process forms the baseline for evaluating the proposed approach. As illustrated in Figure 2, the process begins by collecting Gujarati news videos from the TV9 channel. Each video is decomposed into individual frames to enable frame-level feature analysis. During the first phase, features will be extracted from each frame using image processing techniques. The selected features, like pixel, color, edge, texture, and histogram, together capture the visual and structural content of each frame. The frame features are calculated using following:

PDM: $D_t = \sum_{x=1}^{W} \sum_{y=1}^{H} \sum_{C=1}^{C} |F_t\ (x,y,c) - F_{t-1}(x,y,c)|$ (1)

Where F_t = current frame, F_{t-1}=next frame

ECR: $ECR(t) = \frac{N_{enter}(t)+N_{exit}(t)}{N_{edge}(t)}$ (2)

$N\_enter\ (t)$= Number of new edge pixels appearing in frame Ft+1 not in Ft

$N\_exit\ (t)$=Number of old edge pixels disappearing from frame in Ft and not in Ft+1

$N\_edge\ (t)$=otal number of edge pixels in frame Ft

CDM: $D_{total} = \sum_{x,y}(D_H(x,y) + D_s(x,y) + D_v(x,y)$ (6)

D_H =Measures the difference in Hue, D_s =Represents the difference in Saturation and D_v = Computes the difference in Value

HBA: $Bhattacharyya(H_1, H_2) = \sqrt{1 - \frac{1}{\sqrt{H_1 \cdot H_2 \cdot N}}}$ (7)

H_1 = Normalized histogram of the previous frame, H_2 =Normalized histogram of the current frame and N=256.

Gabor: $g(x,y) = \exp(-\frac{x'^2+y^2+y'^2}{2\sigma^2}) . \exp(j2\pi\frac{x'}{\lambda})$ (8)

ϭ= standard deviation of the Gaussian envelope, Θ= Orientation of Gabor Filter, λ= is the wavelength of the sinusoidal factor, γ= Spatial aspect ratio.

Once the features are extracted, the difference between consecutive frames is calculated using the global threshold values. If the computed frame difference exceeds the threshold, then it is recorded as a scene change and stored as a shot boundary. This method provides a simple but effective mechanism for detecting abrupt transitions such as cuts. Due to the global threshold, this approach struggles with gradual transitions like fades, dissolves, and is also sensitive to illumination changes or camera motion. Existing Shot Boundary Detection techniques yield different results depending on which frame features are considered. Pixel and color-based approaches are computationally simple but highly sensitive to lighting and motion changes. Whereas edge-based and Gabor filters perform better, as these approaches focus on structural and texture variations along with threshold values. Histogram-based methods provide moderate stability against noise but again depend on a Global threshold. Overall, these approaches highlight the difference between accuracy and adaptability, motivating the need for an adaptive, multi-feature framework such as the proposed EHASD.

3.3 Adaptive Threshold-Based Evaluation of Existing Shot Boundary Detection Approaches

In the PDM approach, the global threshold is determined based on the total no of pixels available in the frame. The threshold we set for PDM is 50000, which might not work well across different video categories. If we keep the threshold below 50000, it may detect false shot boundaries; if we keep it high, it may miss the actual shot boundaries. Likewise, we have applied different thresholds for the SBD approaches we discuss in the results section. This method extracts frame features such as pixel, color, edge, texture, and histogram in a similar way as explained earlier and computes inter-frame differences. But this approach dynamically determines a boundary threshold using statistical parameters, mean and standard deviation. A frame is marked as a shot boundary when the consecutive frame difference exceeds the adaptive threshold. Every time we need to set the manual threshold for detecting the shots to overcome this problem, we have used the adaptive threshold, which is based on statistical measures like mean and standard deviation, as shown below in figure 3.

The mean term refers to the average frame value derived from a TV9 video dataset. The Standard Deviation statistic measures the extent of variation of values around the mean. A high standard deviation implies considerable unevenness in the method's performance across different shots, while a low standard deviation indicates more consistent detection results. Adaptive Thresholding (AT) can be calculated as:

$$AT = \mu + \kappa\sigma \quad (9)$$

μ=Mean pixel difference across frames, σ=Standard deviation of the differences, k = A tuning factor.

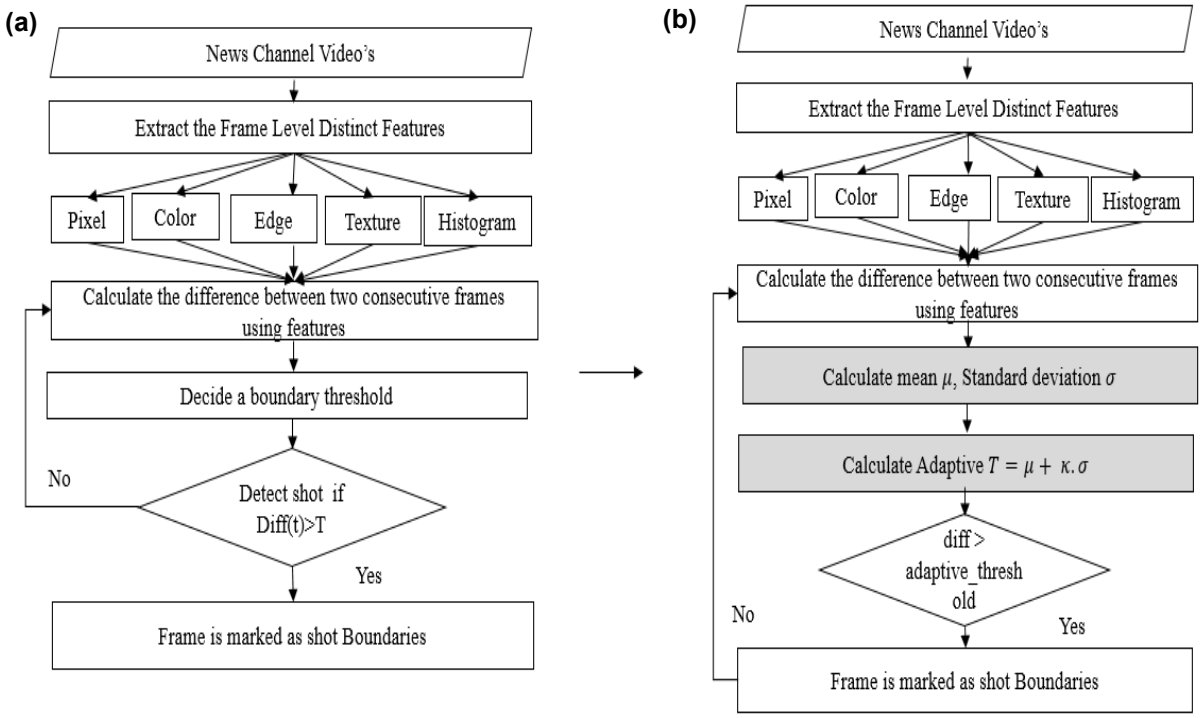


Figure 3. Flowchart **(a)** Shot Boundaries by extracting frame level features using global threshold (**b**) Shot Boundaries by extracting frame level features using adaptive threshold

Here, the k value is selected as 2.0. To ensure statistical permanency and effectiveness across multiple Gujarati video datasets. The experiment was also conducted with k=1, but we did not choose it because more false positives were recorded, whereas k=3 detected more false negatives due to a high k tuning factor. So, for future purposes, k=1 may be more suitable for slow-motion movie scenes, etc., and k=3 for fast-moving videos like live cricket matches, where the ball is hitting a bat and the motion is very high. Hence, for this experiment, k=2 was kept to get the most number of consistent shots.

3.4 Proposed EHASD using SEHT

The existing SBD approaches still suffer from a larger number of shots, leading to inaccurate keyframes. Also, the ECR Canny edge detector detects more edge features, which increases the time complexity, and the statistical threshold still fails to produce accurate results. To overcome these issues, we have proposed the EHASD approach, which resolves the problem of a Global threshold by adopting the AT, uses the Sobel gradients to extract the meaningful edge features, and also uses the concept of the Histogram-based Bhattacharya distance approach to get the accurate and fewer shots, as we can see in the result section. This fusion is called Edge Histogram Adaptive Shot Detection with Statistical Thresholding. As a result, EHASD can: remove the need for algorithmic adjustment by automatically setting thresholds based on edge variations, accurately detect hard cuts and soft transitions (such as dissolves and fades) without eliminating or splitting too much, and reduce the number of false positives and false negatives. The figure 4 below shows the left side with the Global threshold and existing SBD approaches, and the right side with the proposed EHASD approach. Also detailed algorithm is mentioned below.

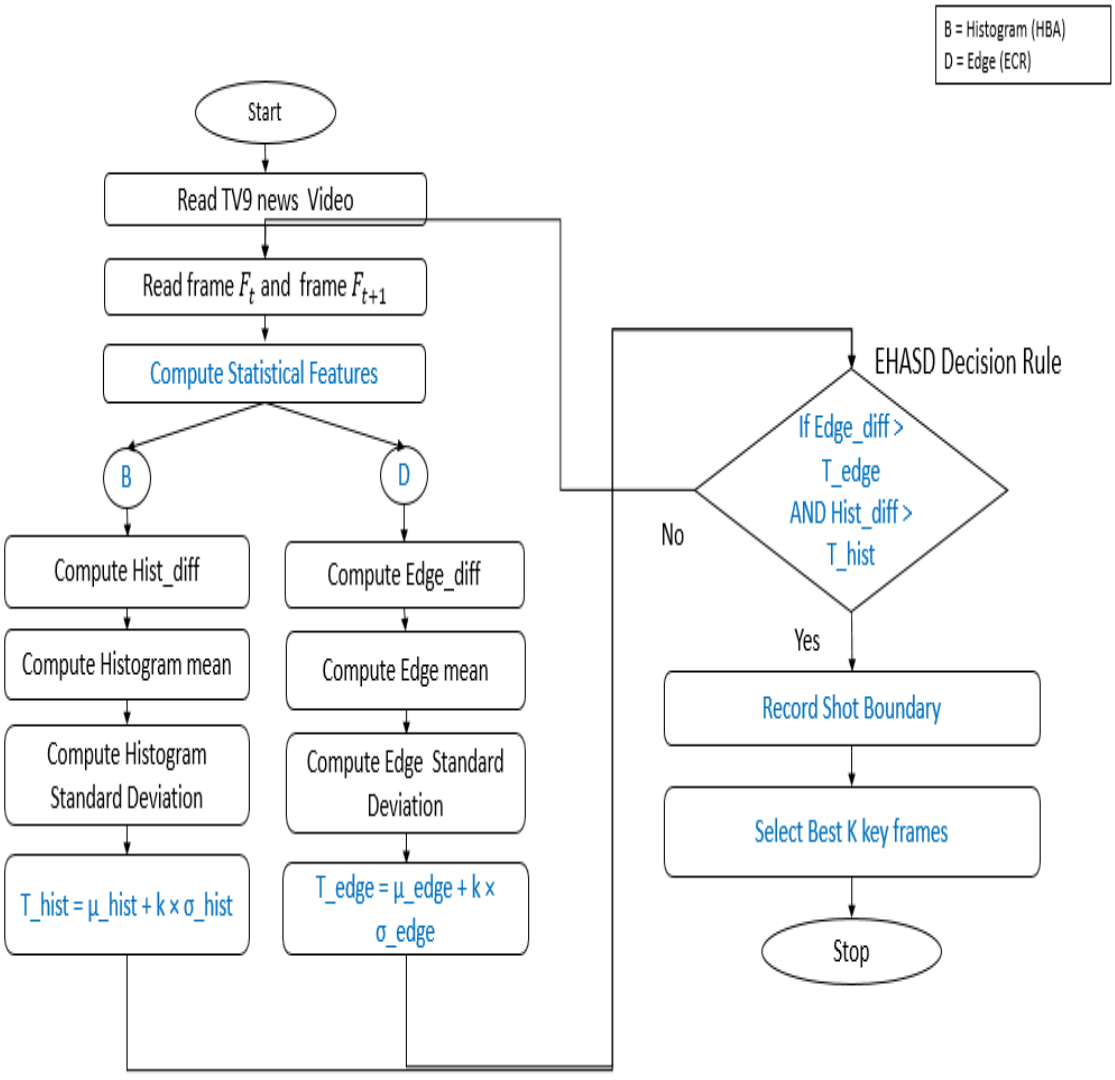


Figure 4. Comparative Workflow of Conventional SBD and Proposed EHASD Framework

Algorithm for Proposed EHASD Framework

Input: Gujarati news video V

Output: List of detected shot boundaries SB

1: Initialize edge_differences ← empty list

2: Initialize histogram_differences ← empty list

3: Load video V and extract FPS, width, height

4: previous_frame ← NULL

5: frame_index ← 0

while V has frames do

7: Read current_frame

8: if current_frame is NULL then

9: break

10: end if

11: if previous_frame ≠ NULL then

12: edge_diff ← compute_edges(current_frame, previous_frame) 10)

13: hist_diff ← compute_histogram_diff(current_frame, previous_frame) (11)

14: Append edge_diff to edge_differences

15: Append hist_diff to histogram_differences

16: end if

17: previous_frame ← current_frame

18: frame_index ← frame_index + 1

19: end while

20: Compute μ_edge, σ_edge from edge_differences

21: Compute μ_hist, σ_hist from histogram_difference

22: T_edge ← μ_edge + 2 × σ_edge (12)

23: T_hist ← μ_hist + 2 × σ_hist (13)

24: for each frame i do

25:if edge_differences[i] > T_edge AND

histogram_differences[i] > T_hist then

26: Mark frame i as shot boundary

27: Append i to SB

28: end if

29: end for

30: return SB

The performance analysis in terms of the number of shots detected using a Global threshold and existing SBD, the number of shots detected with a statistical threshold and existing SBD, and lastly, the number of shots detected with the proposed EHASD will be discussed in the results section. The samples shots from each category is mentioned below with Gujarati text present which mapping successfully with the truth table mentioned in Table 3. Likewise, we have recorded the shots for every category at every stages as shown in below Figure 5.

3.5. Uniqueness of Propsoed EHASD

1. Statistically adaptive threshold for Gujarati language: The Proposed EHASD introduces a statistically adaptive threshold instead of relying on Global or manually set thresholds with edge and histogram fusion features to detect the shot boundaries from News channel videos where repetitive studio layouts are followed by abrupt transitions to field reports or graphics, requiring dynamic sensitivity rather than static thresholds.
2. Edge and Histogram features-based approach Gujarati broadcasts: Unlike pixel, color, or any other frame features, the proposed EHASD only focuses on edge and histogram features, which are used to differentiate news-specific elements such as tickers, graphics, and studio backgrounds.
3. Mean variance-based sensitivity adaptation for different news categories: The proposed approach dynamically adjusts its sensitivity using mean variance using edge differences, which adapt automatically to different news categories like weather, cricket, and political debates without parameter reconfiguration.
4. Design allied with Gujarati script and OCR requirements: Gujarati script contains curved characters, conjuncts, and matras, which makes OCR very dependent on constant and visually clean key frames. The proposed EHASD improves key-frame immovability, ensuring the reservation of valid Gujarati text and storing relevant boundaries for Gujarati text extraction.

Weather-1(n=2996) Politics-1(n=13336) Cricket-1 (n=7106)

Figure 5. The shots detected using the proposed approach, n=no of frames

5. Training-free approach suitable for Gujarati datasets: In contrast to deep learning based SBD models like CNNs, 3D CNNs, and Transformers [25, 27, 29], the proposed EHASD is training-free and fully automatic without manual settings, able to detect shots with meaningful Gujarati text.

Overall, EHASD represents a lightweight and very good alternative to state of art shot detection techniques with robustness, adaptability and explainability without the complexity of deep learning models.

4. Results and Discussion

This section presents the experimental results of SBD-GT, SBD-AT, and the proposed EHASD approach for shot detection and future key frame selection. Standard metrics, such as accuracy, precision, recall, and F1-score, were used to measure performance. This section has two subsections, namely Quantitative analysis of existing SBD-GT, SBD-AT, and EHASD approach for key frames selection.

4.1 Quantitative Performance Comparison between Existing SBD Techniques and the Proposed EHASD Method:

In this research paper, we have conducted experiments on all existing Shot Boundary Detection (SBD) methods on the TV9 news channel video dataset to identify key frames. The experiments were conducted on three different levels in order to identify the best key frames:

4.1.1 Analysis of Shot Boundary Detection with Global Threshold (SBD-GT) Configurations

This table 4 illustrates the quantity of shots identified through several established Shot Boundary Detection (SBD) techniques: Pixel Difference Approach (PDA), Edge Change Ratio (ECR), Color Difference (CD), Gabor-based SBD, and Histogram-Based Approach (HBA) across various video categories labeled with Gujarati words. The findings emphasize the differences in shot detection numbers based on consistent threshold settings.

Feature Normalization Strategy

1. Pixel and Color based Features (PDM, CD): Values are normalized by the total number of pixels:

$$D_{norm} = \frac{D}{W \times H} \quad (14)$$

2. Edge-based Features (ECR): Edge features are binarized using the Sobel operators and ratio is computed as:

$$ECR = \frac{E_{in}+E_{out}}{E_{total}} \quad (15)$$

Values will be in range of [0, 1].
Histogram Features (HBA):

Table 4. Analysis of SBD using different Frame Features

Ground Truth Gujarati Words	Video Label	PDM (Threshold =50000)	ECR (Threshold = 0.5)	CD (Threshold = 10,000,00)	Gabor (Threshold = 50,000)	HBA (Threshold = 0.5)
વરસાદ (Rain), ગરમી (Garmi), બરફ (ICE), વરસાદી (Rainy)	Weather-1	2995	98	491	490	15
	Weather 2	4616	499	491	491	7
રમત (play) , ક્રિકેટ (Cricket),ઇન્ડિયા (India),ઇન્ડિયન (Indian), વર્લ્ડ કપ	Cricket-1	5000	497	475	474	10
	Cricket-2	5000	500	481	481	12
રાજકારણ (Rajkarn) , ભાજપ (Bhajap) ,વડાપ્રધાન (Vadapradhan) , મોદી (Modi) , નરેન્દ્ર મોદી (Narendra Modi) , કોંગ્રેસ (Congress)	Politics-1	5000	500	482	480	5
	Politics-2	5000	500	480	481	26
Total	36077					

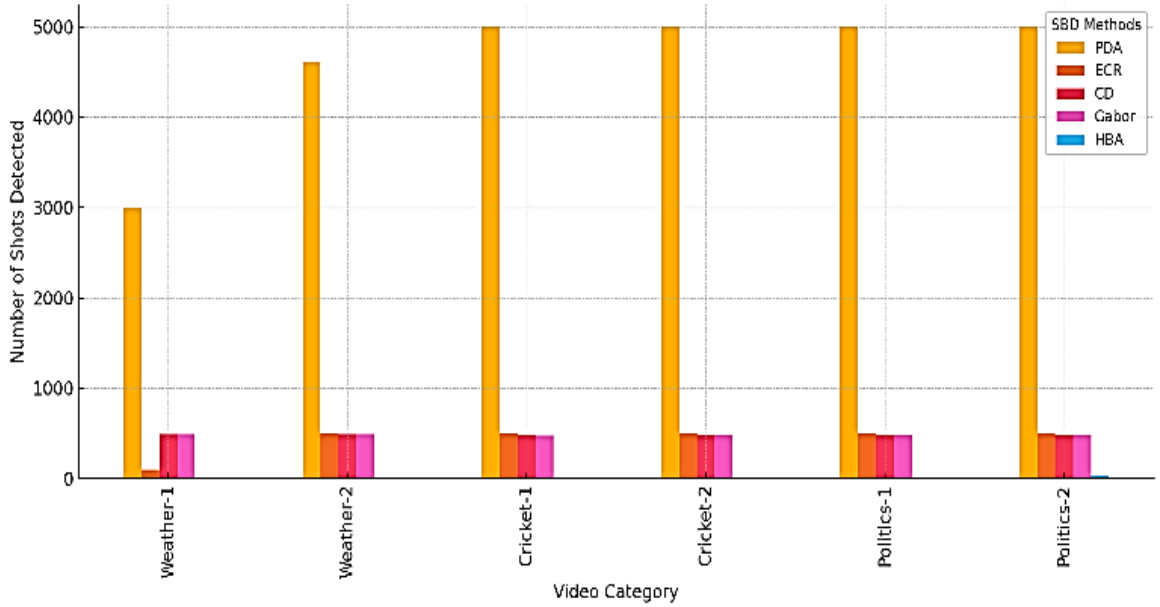


Figure 6. Comparison of Shot Detection Counts Across Methods and Categories

3. Histogram features: These features are computed as L1-normalized means bin sums=1, which enables thresholding independent of illumination changes.

4. Texture Features (Gabor): Gabor energy responses are normalized by the maximum filter response per frame pair

The Global threshold experiments show that PDM, Gabor and CD significantly over-detect shot boundaries, often giving negligible visual changes as new shots. ECR performs better but is less stable, while HBA produces the lowest counts and sometimes misses real transitions. The results confirm that static thresholds are not reliable for Gujarati news content, where motion, tickers, and graphics heavily distort frame differences.

4.1.2 Analysis of Shot Boundary Detection with Adaptive Threshold (SBD-AT) Configurations

The Global / global threshold need to adjust every time for the different video category. As discussed above in the 4.1.1 section SBD-GT approach generated a larger number of shots with more no of false positives and negatives. The resultant shots have been presented in Figure 6. To overcome these challenges of a Global threshold, the adaptive threshold was successfully implemented on the TV9 News channel dataset [35]. The table 5 presents the results of SBD-AT for different video categories like Weather, Cricket and Politics integrating five different features as discussed earlier. The edge mean and standard deviations of all the features have been calculated to check the statistical significant. The results illustrate variations in sensitivity and consistency across pixel, edge, color, texture and histogram-based approaches.

The figure 7 shows the statistical evaluation of the existing SBD approaches with adaptive threshold mean & deviation in a form of histogram for the Pixel Difference Method. In a similar way, CDM, ECR, Gabor and HBA approaches can also be presented on Histograms. The statistical threshold result is computed across all the frames of all the categories of video.

Significance of the Different Lines in a Histogram: Each histogram shows the frequency of pixel difference values between consecutive frames, helping identify shot transitions. Table 6 lists Interpretation and Role of Threshold Lines in Pixel Difference Histograms.

The outcomes of the SBD-AT approach are listed below:

1. To analyse the degree to which Global and adaptive thresholds align with the actual distribution.
2. To showcase the effectiveness of a statistically adaptive threshold compared to a one-size-fits-all Global value.
3. To support your assertion that Global thresholds can either over-segment or miss boundaries, while the adaptive technique aligns with the natural peaks and outliers found in the distribution.

The number of shots is lower than in SBD-GT, leading to over-segmentation. Both approaches, SBD-GT and SBD-AT, highlight the need for adaptive frameworks, even after focusing on low-level, mid-level, and high-level features, as they still cannot generate accurate and robust shots. The no of shots in SBD-GT is 36077, which reduces to 7280.

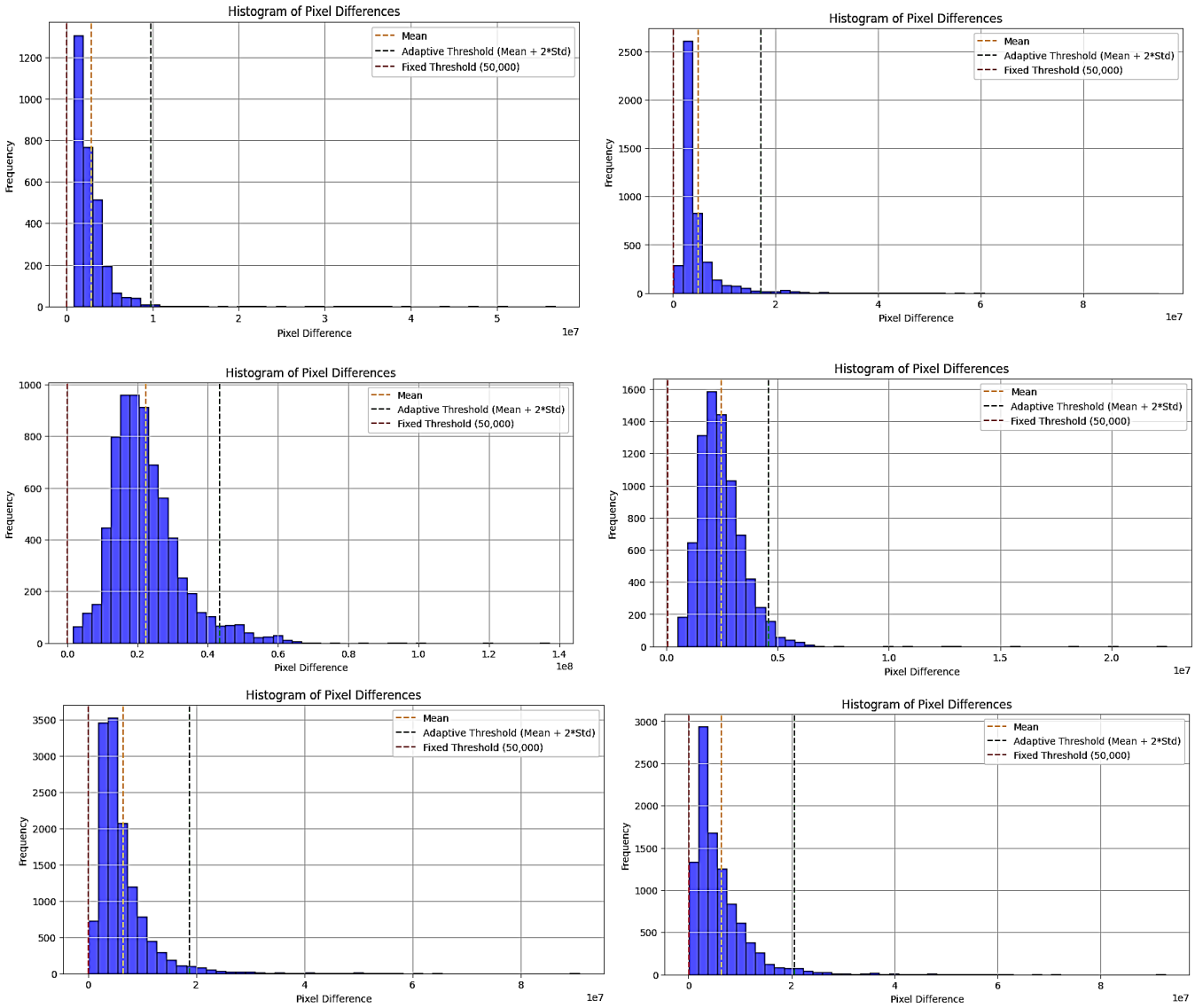


Figure 7. Adaptive Threshold Visualization in Pixel Difference Histograms for Various Gujarati News Videos

4.1.3 Analysis of Proposed Edge Histogram Adaptive Shot Detection (EHASD) Approach:

The proposed approach successfully detects fewer shots with relevant text. For future work, these shots will be passed to the keyframe selection pipeline to select the best keyframes with relevant text. Furthermore, the text from those frames will be extracted for video retrieval purposes. As discussed in the algorithm, the edge mean, histogram mean, and statistical evaluation, i.e., the mean, have been performed for all the video categories from TV9 news channel videos. The mean and standard deviation details of the proposed EHASD approach with tuning factor k=2.0 is presented in table 7.

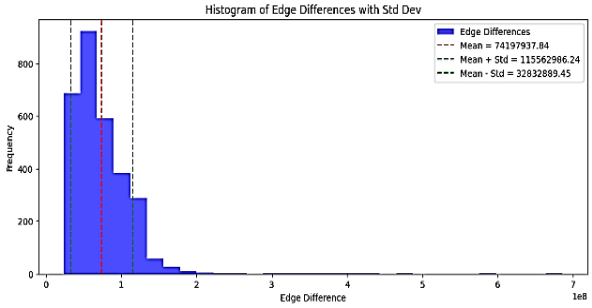
The table 7 shows that the proposed method produces a much smaller and orderly number of shots across all video categories. The Edge Mean values vary widely across videos, indicating that edge activity is highly content dependent. Cricket-1 shows the highest mean (671.94×10^6) due to rapid motion and camera transitions, while Weather-1 shows the lowest (74.2×10^6) due to stable scenes and slow movements. In contrast, the Histogram Mean gives low and stable (0.02–0.03) range for all the categories, verifying that color distribution changes far less as compared to edge intensity, even when scene motion is high. The changes in edge differences are more robust than histogram changes in broadcast news, where studio backgrounds and lower third tickers maintain a similar color pattern. Cricket and Politics videos have high edge scales; the final shot count increased due to statistical thresholding rather than a global threshold. This confirms that the framework captures a structural transition and making the selected shots both meaningful and relevant for text retrieval. The edge mean and Histogram mean also calculated for the proposed approach. The Figure 8 shows histogram analysis of the proposed EHASD approach. Left side histogram indicates the edge differences with std dev. and right side of the histogram represents Histogram difference std dev of all the 6 videos.

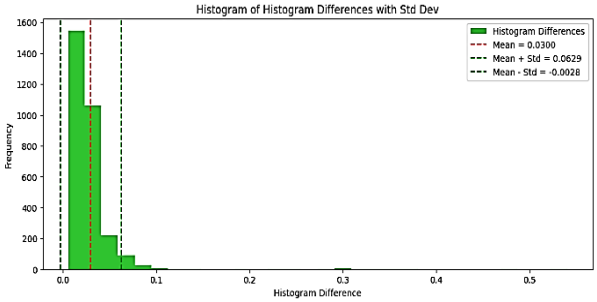
Table 5. Quantitative Results of SBD Techniques on TV9 Gujarati News Dataset (k = 2.0)

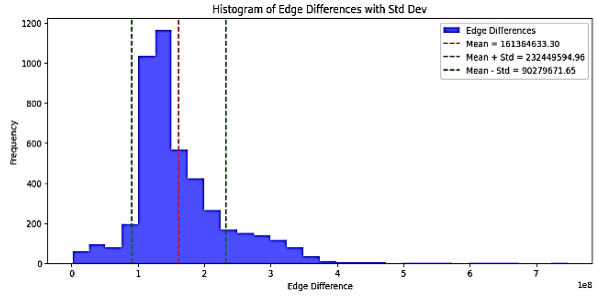
Video Category	Frame Level Features	Edge Mean ($\times 10^6$)	Standard Deviation ($\times 10^6$)	No. of Shots (k = 2.0)
Weather-1	PDM	2.9	3.46	50
	ECR	0.25	0.13	81
	CD	8.99	9.84	44
	Gabor SBD	2.9	3.46	133
	HBA	0.03	0.03	38
Weather-2	PDM	4.95	6.09	177
	ECR	0.38	0.15	330
	CD	13.41	18.23	150
	Gabor SBD	2.64	2.57	74
	HBA	0.03	0.05	168
Cricket-1	PDM	22.49	10.49	339
	ECR	0.57	0.09	140
	CD	69.87	29.95	317
	Gabor SBD	12.25	3.73	288
	HBA	0.02	0.01	245
Cricket-2	PDM	2.45	1.06	250
	ECR	0.51	0.1	54
	CD	6.09	2.64	232
	Gabor SBD	1.41	0.68	332
	HBA	0.03	0.02	333
Politics-1	PDM	6.47	6.15	470
	ECR	0.43	0.16	143
	CD	17.51	19.49	492
	Gabor SBD	8.95	3.31	562
	HBA	0.02	0.03	305
Politics-2	PDM	6.36	7.11	356
	ECR	0.39	0.19	111
	CD	16.67	15.91	329
	Gabor SBD	3.51	2.72	533
	HBA	0.11	0.04	204
Total	-	-	-	7280

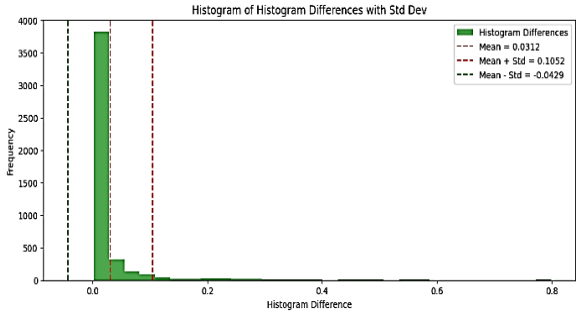
Table 6. Interpretation and Role of Threshold Lines in Pixel Difference Histogram

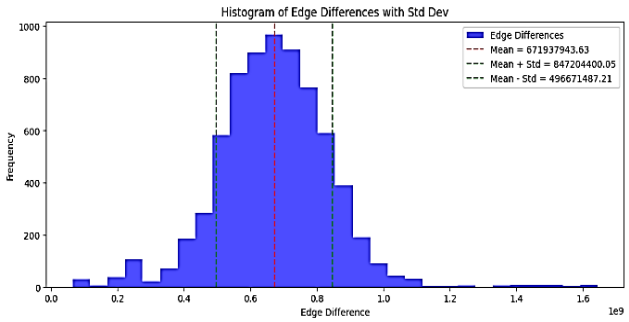
Line Type	Signifies	Usage	Limitations / Advantages
Red Line – Global Threshold	A manually set threshold value	If the pixel difference exceeds the threshold value, then recorded as a shot boundary	May not perform optimally across videos that have varying dynamics
Orange Line – Mean	The mean of all pixel difference values for the video	Displays the central tendency of frame difference values.	Assists in the understanding of the distribution skew and normal activity level
Green Line – Statistical Threshold (Mean + 2*Std Dev)	Shows the adaptive threshold	Alters the cutoff for each video dynamically, based on the attributes of the content.	Takes into account content diversity; strengthens robustness across various types of videos (weather, politics, cricket).

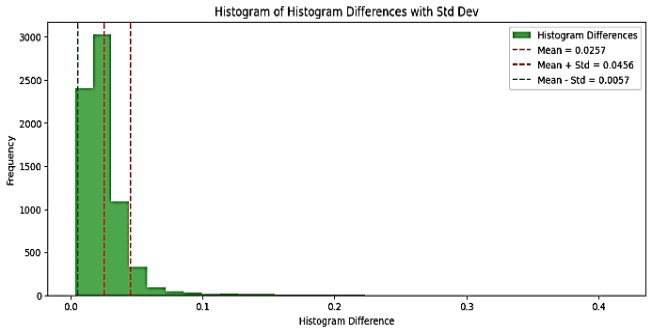

Histogram of Edge Differences with Std Dev
Edge Differences
Mean = 74197937.84
Mean + Std = 115562906.24
Mean - Std = 32832889.45
Frequency
Edge Difference
1e8

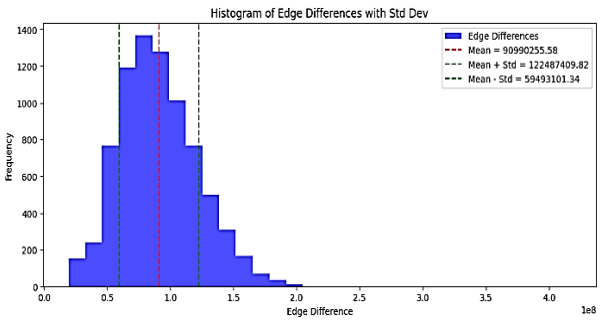

Histogram of Histogram Differences with Std Dev
Histogram Differences
Mean = 0.0300
Mean + Std = 0.0629
Mean - Std = -0.0028
Frequency
Histogram Difference

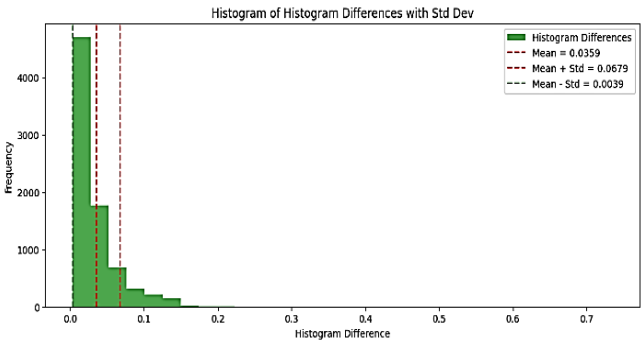

Histogram of Edge Differences with Std Dev
Edge Differences
Mean = 161364633.30
Mean + Std = 232449594.96
Mean - Std = 90279671.65
Frequency
Edge Difference
1e8

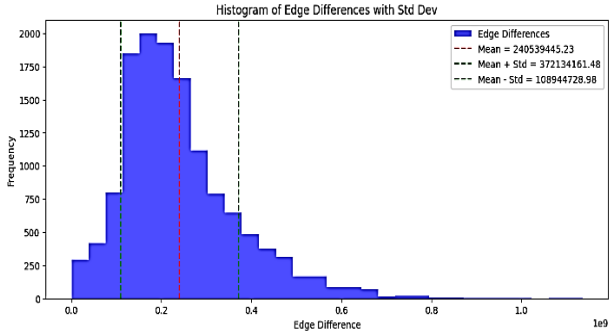

Histogram of Histogram Differences with Std Dev
Histogram Differences
Mean = 0.0312
Mean + Std = 0.1052
Mean - Std = -0.0429
Frequency
Histogram Difference

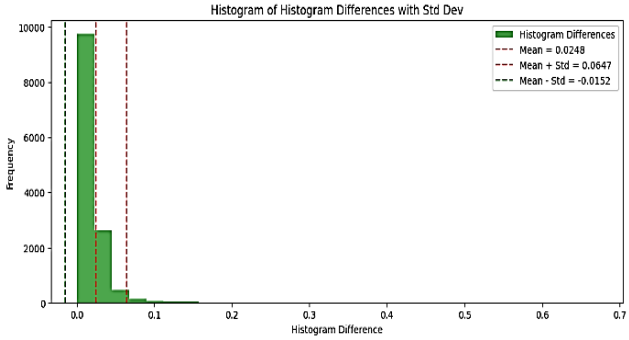

Histogram of Edge Differences with Std Dev
Edge Differences
Mean = 671937943.63
Mean + Std = 847204400.05
Mean - Std = 496671487.21
Frequency
Edge Difference
1e9


Histogram of Histogram Differences with Std Dev
Histogram Differences
Mean = 0.0257
Mean + Std = 0.0456
Mean - Std = 0.0057
Frequency
Histogram Difference


Histogram of Edge Differences with Std Dev
Edge Differences
Mean = 90990255.58
Mean + Std = 122487409.82
Mean - Std = 59493101.34
Frequency
Edge Difference
1e8


Histogram of Histogram Differences with Std Dev
Histogram Differences
Mean = 0.0359
Mean + Std = 0.0679
Mean - Std = 0.0039
Frequency
Histogram Difference


Histogram of Edge Differences with Std Dev
Edge Differences
Mean = 240539445.23
Mean + Std = 372134161.48
Mean - Std = 108944728.98
Frequency
Edge Difference
1e9


Histogram of Histogram Differences with Std Dev
Histogram Differences
Mean = 0.0248
Mean + Std = 0.0647
Mean - Std = -0.0152
Frequency
Histogram Difference

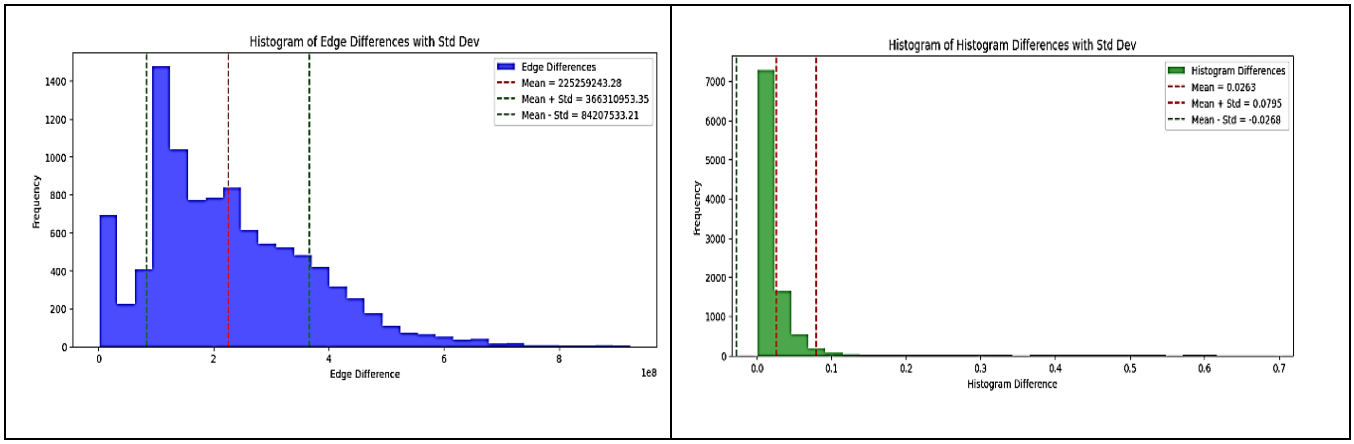


Figure 8. Histogram Analysis of Proposed EHASD Approach

Table 7. Results of EHASD Feature Analysis at k = 2.0

Video Dataset	Total No. of Shots (k = 2.0)	Edge Mean (×10⁶)	Histogram Mean
Weather-1	18	74.2	0.03
Weather-2	113	161.36	0.0312
Cricket-1	44	671.94	0.0257
Cricket-2	39	90.99	0.0359
Politics-1	108	240.54	0.0248
Politics-2	133	225.26	0.0263
Total	455	-	-

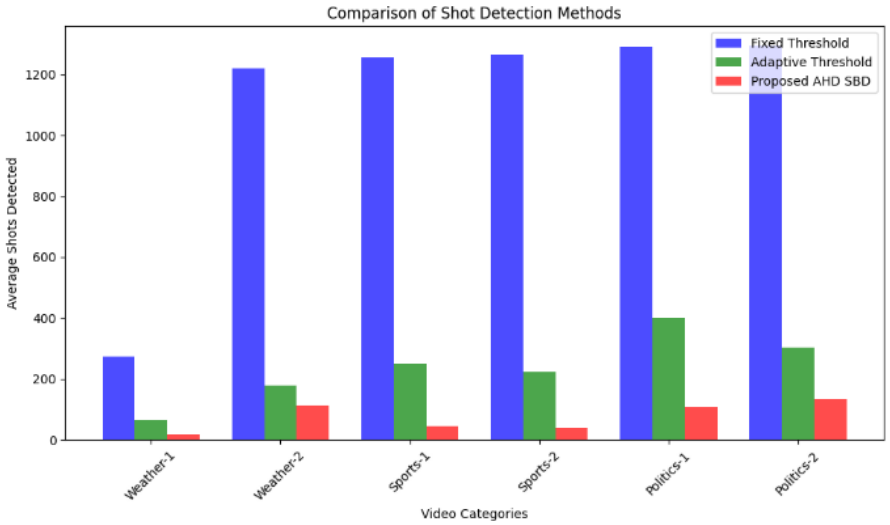


Figure 9. Comparison of shot detection methods across different video categories

The figure 9 shows the comparative study of shot detection approaches across various video categories. The global threshold approach shown in blue constantly identifies a higher number of shots, which indicates higher sensitivity, although with an increase in false positives. The Adaptive Threshold approach, represented in green, achieves a lower shot count while maintaining a reasonable level of accuracy. The Proposed EHASD method depicted in red records the least number of shot, suggesting enhanced precision and fewer false positives. This demonstrates the proposed EHASD approach effectiveness for detecting shots for video retrieval application.

The figure determines the average number of shots detected using three thresholding strategies like Global Threshold, Adaptive Threshold and the proposed approach. The results show that the proposed method

attains a steady detection rate over segmentation in Global thresholding and under segmentation in adaptive thresholding, thereby ensuring more accurate boundary identification across various datasets, including weather, Cricket and political videos.

4.1.4 Performance Analysis of Proposed EHASD Approach

In order to conclude the effectiveness of shot detection for the retrieval of Gujarati text, accuracy is measured using Precision and Recall which are derived at the frame level by comparing detected key-frames to manually confirmed ground-truth text frames. The precision and recall is calculated as:

$$precision = \frac{\text{Correctly identified text frames}}{\text{the total number of frames detected}} \quad (16)$$

$$Recall = \frac{\text{Correctly identified text frames}}{\text{Total actual relevant text frames as per ground truth}} \quad (17)$$

%PCC (Percentage of Correct Classification) represents the overall consistency of correct frame selection with respect to all assessed frames. The performance analysis of the proposed EHASD framework for detecting the shots having Gujarati text based on the ground truth table have been identified, evaluated and presented in figure 10

The Combined ROC and Precision–Recall plots for Gujarati text detection across Weather, Cricket and News video categories. Each curve is generated from a single operating point. The plots provide a comparison between true positive rate, false positive rate, precision, and recall for each video. The overall performance of the proposed approach across all categories is presented in the table 8.

The figure 11 represents an analysis of the performance of the proposed EHASD across various video categories Weather, Cricket and News using different metrics such as PCC, Precision and Recall.

Weather videos show the highest level of accuracy up to 93.25%. Both the Cricket and politics categories also performed well with 90% accuracy. The strong relationship between Precision and Recall across all categories highlights the model's ability to reduce false positives and false negatives. The robustness and balanced performance of the proposed EHASD framework is shown figure 12 which represents the precision of correctly identified frames with Gujarati text.

The best shots have Gujarati text, which was further mapped to the ground truth table as discussed earlier. Table 9 presents an evaluation of the existing video segmentation techniques from the literature survey [21-24] with the proposed EHASD framework

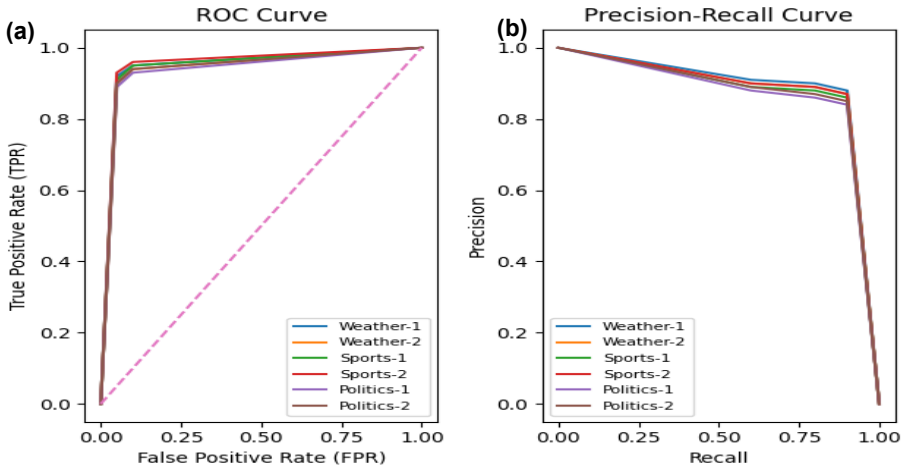


Figure10 (a). Combined ROC Curve **(b)** Precision –Recall Curve AUC TV9 news Channel Video Dataset

Table 8. performance analysis of proposed EHASD

Video Annotation	%PCC
weather-1	0.9325
weather-2	0.9112
Cricket-1	0.902
Cricket-2	0.901
politics-1	0.9111
politics-2	0.91

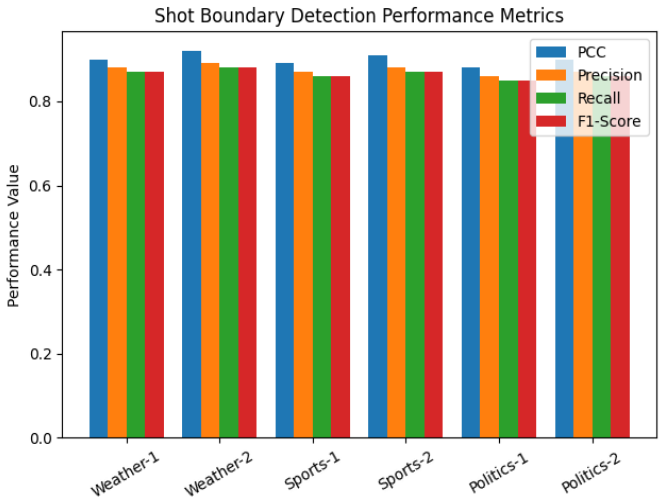


Figure 11. Performance analysis of different categories of video dataset (Source=TV9 Gujarati news channel)

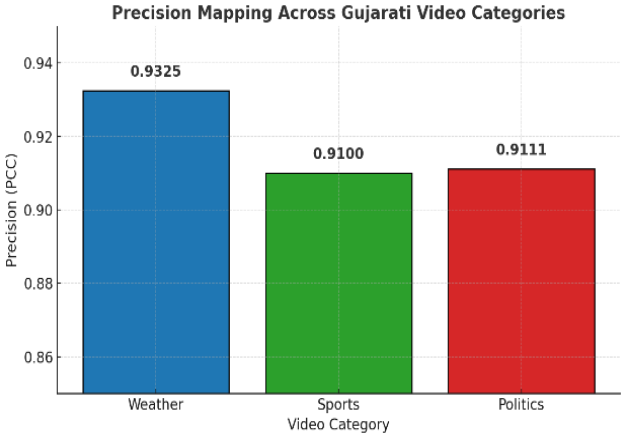


Figure 12. Precision Mapping Across Gujarati Video Categories

Table 9. comparison of existing [13, 15, 16, 20],vs proposed approaches

Dataset	Threshold Types		Transition Types	
	Threshold	Adaptive Threshold	HT	Precision
TRECVID [13]	√	-	√	0.8487
TRECVID 2007 [15]	-	√	√	0.87
[16]	√	-	√	0.824
TRECVID [20]	-	√	√	0.89
SBD HBA Weather-2	√	-	√	0.89
EHASD Weather-1	-	√	√	**0.9325**

The previous TRECVID-based approaches achieved the precision values ranging from 0.82 to 0.89 for detecting both abrupt and gradual transitions. On the other hand, the proposed EHASD achieved a precision of 0.9325 on the weather-1 video from the dataset.

4.2 Quantitative Validation of Proposed Method through Z-Test Analysis:

To evaluate whether the proposed EHASD framework yields a statistically significant difference in the average number of shots detected compared with traditional SBD approaches (table 10), a Z-test was conducted. This test intended to identify whether the proposed EHASD framework offered a significant improvement in shot detection performance compared to existing SBD techniques.

1. ***Null Hypothesis (H_0):*** There is no substantial variation between the average number of shots identified by the proposed EHASD and the traditional SBD approaches.

2. ***Alternative Hypothesis (H_1):*** The proposed EHASD identifies a significantly different count of shots in relation to existing SBD methods.

Step 1: Extract Data for Comparison

The evaluation compared the average number of detected shots across six categories of Gujarati videos, applying a 95% confidence level.

- Mean (X_1) = [69.2, 179.8, 265.8, 240.2, 394.4, 306.6].
- Proposed EHASD Mean (X_2) = [18, 113, 44, 39, 108, 133].

Step 2: Compute Averages and Standard Deviations for sample sizes: $n_1 = n_2 = 6$

$\overline{X1} = 242.66$ and σ1=112.93

$\overline{X2}$ =75.83 and σ2=46.84

Step 3: Apply Z-Test Formula

$$Z = \frac{\overline{X1} - \overline{X2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (18)$$

$$Z = \frac{242.66 - 75.83}{\sqrt{\frac{112.63^2}{6} + \frac{46.84^2}{6}}} \quad (19)$$

$$Z = \frac{166.83}{\sqrt{(2126.1 + 365.7)}} \quad (20)$$

$$Z = \frac{166.83}{48.99} = 3.40 \quad (21)$$

Below figure 13 shows the computed Z-value (3.40) is significantly higher than the critical limit (±1.96), confirming that the performance improvement of EHASD over existing SBD methods is statistically significant.

The calculated Z-value (3.40) exceeded the critical value of ±1.96, leading to rejection of the null hypothesis. This confirms that the proposed EHASD approach significantly reduces redundant or false recognitions while providing better boundary precision than traditional Global- or adaptive-threshold-based SBD methods.

4.2.1 The Paired T-Test

From the paired t-test, we find a statistically significant difference in the number of shots detected between the existing SBD methods and the proposed EHASD framework ($t = 4.47$, $df = 5$, $p < 0.01$). This confirms that EHASD significantly reduces redundant or false shot detections while preserving meaningful shot boundaries across all Gujarati news video categories.

Initially, a Z-test was conducted to examine differences in average shot counts; however, due to the small sample size ($n = 6$) and unknown population variance, this test was found to be statistically inappropriate. Therefore, a paired t-test was implemented, as it is more appropriate for comparing two related samples within the same video categories and provides an effective assessment of statistical significance with limited samples.

The Z-test results show that the computed Z value in Figure 14 is 3.40 exceeds the critical threshold (±1.96), indicating statistically significant performance improvement. To further validate sample-wise consistency, a paired t-test was conducted, yielding $t(5) = 4.47$ with $p < 0.01$. These results jointly confirm the robustness and statistical reliability of the proposed approach.

4.3. Ablation Study: Contribution of Edge and Histogram Components

The ablation results indicate that both Sobel-based edge features and Bhattacharyya histogram distance individually contributed to shot boundary detection performance, their fusion edge & histogram together with adaptive thresholding gives the highest precision, recall and F1-score (Table 11). This confirms that edge information captures structural transitions, whereas histogram distance detects the global appearance of the respective frames which leads to more robust detection.

Table 10. Comparison of data extracted

Video Category	Mean of Existing SBD Shots	Proposed EHASD Shots
Weather-1	(50 + 81 + 44 + 133 + 38)/5 = 69.2	18
Weather-2	(177 + 330 + 150 + 74 + 168)/5 = 179.8	113
Cricket-1	(339 + 140 + 317 + 288 + 245)/5 = 265.8	44
Cricket-2	(250 + 54 + 232 + 332 + 333)/5 = 240.2	39
Politics-1	(470 + 143 + 492 + 562 + 305)/5 = 394.4	108
Politics-2	(356 + 111 + 329 + 533 + 204)/5 = 306.6	133

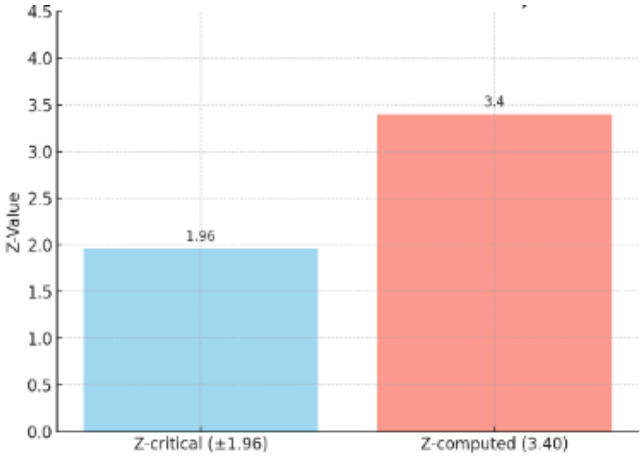


Figure 13. Z-Test Statistical Validation for Shot Boundary Detection

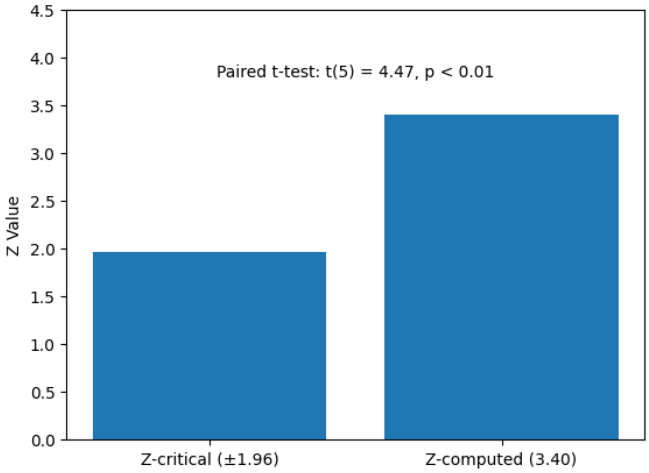


Figure 14. Z-test with Paired-t Test Validation

Table 11. Contribution of Edge and Histogram Components

Features Method	Edge Feature	Histogram Feature	Precision	Recall	F1-Score
Edge features	✓	--	0.86	0.84	0.85
Histogram feature	--	✓	0.88	0.86	0.87
EHASD (Fusion)	✓	✓	0.91	0.89	0.90

Table 12. Comparative Analysis of Learning-Based and Proposed EHASD Methods

Approach	Key idea	Strength
NAS over 3D ConvNets + Transformers [4]	Searches strong SBD architectures; introduces a public short-video SBD dataset	Strong F1 score as evaluated on multiple public sets
3D CNN + attention [25]	Spatiotemporal features, attention for transition detection	Better robustness for complex videos but needs labeled
TransNet V2 (3D CNN) [28]	Fast shot transition detection network	Fast Computation
Proposed EHASD	Multi-feature fusion + adaptive thresholding	No training labels Explainable suitable for regional language Gujarati

To contextualize the experimental findings, Table 12 presents a qualitative comparison between recent deep learning-based shot boundary detection approaches and the proposed EHASD framework

Recent deep learning approaches to detecting shot boundaries include 3D CNNs with attention, transformer models and TransNet V2, which achieve good performance but depend heavily on labeled data and are computationally expensive. In contrast, the proposed EHASD works on multi-feature fusion with adaptive thresholding, requiring no training data for regional language like Gujarati.

5. Conclusion and Future Scope

Our experimental process began by analyzing video content from the TV9 Gujarati news channel and subsequently constructing a ground truth table organized by category. In the initial stage we have implemented SBD techniques: Pixel Difference Approach, ECR, Color Difference, Gabor and HBA approaches for different frame features and calculate the difference between the consecutive frames. Based on the Global threshold we have recorded high number of false positive shots. HBA outperformed other methods with fewer shots and lower false positives still required the improvement. However, challenges continued, including high false positive rates and the absence of SBD techniques for regional languages like Gujarati. To overcome these issues, we have applied adaptive threshold with k tuning factor which was later evaluated against five established SBD techniques: Pixel Difference Approach, ECR, Color Difference, Gabor and HBA. The performance is evaluated based on detected shots and mapped with ground truth table. This accuracy was better than the Global threshold value over existing SBD approaches.

On the other hand, the proposed EHASD + SEHT approach consistently achieves the less no of shots with relevant Gujarati text detections across all categories validating its capability to successfully capture only significant shot boundaries. As a result, EHASD + SEHT maintains a balanced performance between precision and recall, ensuring constant, content retrieval detection which is suitable for structured broadcast content. The EHASD method uses the SEHT approach, which confirmed higher performance, mainly in the Weather video category, where it achieved a shot detection accuracy of 0.9325 with fewer false positives. To ensure that the improvement achieved by the proposed EHASD framework. A Z-test was also carried out to check the statistical significance of the performance against existing Shot Boundary Detection (SBD) methods. The test is compared the mean of number of shots detected across all video categories, with results showing a computed Z-value of 3.40, beyond the critical value of ±1.96 at a 95% confidence level. This statistical result confirms that the difference between the proposed and existing methods is significant. The EHASD framework, therefore, determines an assessable and reliable enhancement in shot detection accuracy, effectively reducing false positives and generating stable segmentation across various Gujarati news videos.

Future work could focus on extending the EHASD framework to support multi-modal video analysis by integrating more features like audio cues / voice, written pattern and motion vectors for enhanced segmentation accuracy. Additionally, combining EHASD with deep learning-based sequence models could improve detection performance for highly gradual or visually ambiguous transitions. Incorporating OCR-based feedback loops could further refine key frame selection in text-intensive videos. Finally, applying and adapting this framework to other low-resource language datasets would offer valuable insights into its cross-lingual generalizability and scalability. For the future multi-channel regional news datasets and public benchmark dataset can be used for this experiment purpose.

This study does not include a direct experimental comparison with recent deep learning-based SBD models, such as TransNet V2 or 3D CNN–based approaches, as these methods require large, domain-specific annotated datasets that are currently unavailable for Gujarati news broadcasts. Future work will focus on benchmarking EHASD against such publicly available models after domain adaptation and annotation, and exploring hybrid lightweight–deep SBD frameworks

Future work could focus on extending the proposed EHASD framework toward multi-modal video analysis by integrating complementary features such as audio and voice-based cues, which have validated effectiveness in characterizing temporal signal variations. Integrating written pattern and structural analysis of textual regions, including layout persistence and caption behavior, may further enhance segmentation accuracy in text-intensive news videos. In addition, motion based descriptors and multi-feature fusion strategies, inspired by biometric and behavioral pattern analysis, could improve robustness against complex scene dynamics. Speech-based temporal similarity measures, such as MFCC with DTW, may also support the detection of highly gradual or visually ambiguous transitions. Future research will explore OCR-driven feedback loops for adaptive key-frame selection and evaluate the generalizability of EHASD across other low-resource language datasets and multi-channel regional news benchmarks.

References

[1] N. Spolaor, H.D. Lee, W.S.R. Takaki, L.A. Ensina, C.S.R. Coy, F.C. Wu, A Systematic

Review on Content-Based Video Retrieval. Engineering Applications of Artificial Intelligence, 90, (2020) 103557. https://doi.org/10.1016/j.engappai.2020.103557

[2] Office of the Registrar General and Census Commissioner, India, (2018) Census of India 2011, C-16 Population by Mother Tongue: Distribution of the 22 Scheduled Languages.

[3] A. Bhuva, D. Mishra, Gujarati Optical Character Recognition using Efficient Text Feature Extraction Approaches. Informatica, 49, (2025). https://doi.org/10.31449/inf.v49i28.8341

[4] A. Chabukswar, Deepa Shenoy, Dasharath S.M, Venugopal K.R, Deceptive News Content Detection using a Hybrid Transformer-based and Deep Learning Model with Explainability. International Research Journal of Multidisciplinary Technovation, (2025) 212–234. https://doi.org/10.54392/irjmt25613

[5] B.T. Truong, S. Venkatesh, Video Abstraction: A Systematic Review and Classification. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 3(1), (2007) 3-es. https://doi.org/10.1145/1198302.1198305

[6] Y. Zhai, M. Shah, Visual Attention Detection in Video Sequences using Spatiotemporal Cues. In Proceedings of the 14th ACM International Conference on Multimedia, (2005) 815-824. https://doi.org/10.1145/1180639.1180824

[7] C.B.A. Sai Ram, S. Sharma, Video segmentation: A Systematic Review on Recent Advances, Techniques, and Classification. International Conference on Advances in Computing, Communication Control and Networking (ICACCCN2018), (2018) 615–620, https://doi.org/10.1109/ICACCCN.2018.8748578

[8] T. Kar, P. Kanungo, S.N. Mohanty, S. Groppe, J. Groppe, Video Shot-Boundary Detection: Issues, Challenges and Solutions. Artificial Intelligence Review, 57(4), (2024) 104.https://doi.org/10.1007/s10462-024-10742-1

[9] U. Gargi, R. Kasturi, S. Strayer, Performance Characterization of Video-Shot-Change Detection Methods. IEEE Transactions on Circuits and Systems for Video Technology, 10(1), (2000) 1-13. https://doi.org/10.1109/76.825852

[10] Y. Bendraou, F. Essannouni, D. Aboutajdine, A. Salam, Video Shot Boundary Detection Method using Histogram Differences and Local Image Descriptor. In 2014 Second world conference on complex systems (WCCS) IEEE, Morocco. https://doi.org/10.1109/icocs.2014.7060883

[11] J. Mas, G. Fernandez, (2003). Video Shot Boundary Detection Based on Color Histogram. In TRECVID. https://www-nlpir.nist.gov/projects/tvpubs/tvpapers03/ramonl ull.paper.pdf

[12] H. Zhang, A. Kankanhalli, S.W. Smoliar, Automatic Partitioning of Full-Motion Video. Multimedia Systems, 1(1), (1993) 10-28. https://doi.org/10.1007/bf01210504

[13] N. J. Janwe and K. K. Bhoyar, Video Shot Boundary Detection based on JND Color Histogram. IEEE Second International Conference on Image Information Processing (ICIIP), Shimla, India, (2013) pp. 476–480. https://doi.org/10.1109/ICIIP.2013.6707637

[14] A. Adnan, M. Ali, Shot boundary detection using sorted color histogram polynomial curve. Life Science Journal, 10(4), (2013) 1965–1972.

[15] M. Verma, B. Raman, A Hierarchical Shot Boundary Detection Algorithm Using Global and Local Features. In: Raman, B., Kumar, S., Roy, P., Sen, D. (eds) Proceedings of International Conference on Computer Vision and Image Processing. Advances in Intelligent Systems and Computing, Springer, Singapore, 460, (2017). https://doi.org/10.1007/978-981-10-2107-7_35

[16] Y. Tewari, P. Soni, S. Singh, M.S Turlapati, A. Bhuva, Real Time Sign Language Recognition Framework for Two Way Communication. In 2021 International Conference on Communication information and Computing Technology (ICCICT) IEEE, (2021) 1-6. https://doi.org/10.1109/ICCICT50803.2021.9510094

[17] N. Ibrahim, Z. Abduljabbar, Video Shot Boundary Detection based on Frames Objects Comparison and Scale-Invariant Feature Transform Technique. Computer Science and Information Technologies, 5(2), (2024) 130-139. https://doi.org/10.11591/csit.v5i2.p130-139

[18] S. Dhiman, R. Chawla, S. Gupta, A Novel Video Shot Boundary Detection Framework Employing DCT and Pattern Matching. Multimedia Tools and Applications, 78(24), (2019) 34707-34723. https://doi.org/10.1007/s11042-019-08170-3

[19] B.A. Halim, T. Faiza, (2019) Shot Boundary Detection: Fundamental Concepts and Survey. Conference on Innovative Trends in Computer Science.

[20] P.K. Sahoo, S. Soltani, A.K.C. Wong, A Survey of Thresholding Techniques. Computer Vision, Graphics, and Image Processing, 41(2), (1988) 233-260. https://doi.org/10.1016/0734-189x(88)90022-9

[21] S.N. Kumar, V.S.K. Reddy, L.K. Balivada, (2024). Content-based video retrieval using Deep Learning Algorithms. Research square. https://doi.org/10.21203/rs.3.rs-4331245/v1

[22] M.J. Esteve Brotons, F.J. Lucendo, R.J. Javier, J. García-Rodriguez, Shot Boundary Detection with 3D Depthwise Convolutions and Visual Attention. Sensors, 23(16), (2023) 7022.

https://doi.org/10.3390/s23167022

[23] T. Soucek, J. Lokoc, TransNet V2: An Effective Deep Network Architecture for Fast Shot Transition Detection. In Proceedings of the 32nd ACM International Conference on Multimedia, (2024) 11218-11221). https://doi.org/10.1145/3664647.3685517

[24] W. Zhu, Y. Huang, X. Xie, W. Liu, J. Deng, D. Zhang, Z. Wang, J. Liu, Autoshot: A Short Video Dataset and State-of-the-Art Shot Boundary Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, (2023) 2238-2247). https://doi.org/10.1109/cvprw59228.2023.00218

[25] S. Tang, L. Feng, Z. Kuang, Y. Chen, W. Zhang, Fast Video Shot Transition Localization with Deep Structured Models. In: Jawahar, C., Li, H., Mori, G., Schindler, K. (eds) Computer Vision – ACCV 2018. ACCV 2018. Lecture Notes in Computer Science, Springer, Cham. (2019) 11361. https://doi.org/10.1007/978-3-030-20887-5_36

[26] J.T. Jose, S. Rajkumar, M.R. Ghalib, A. Shankar, P. Sharma, M.R. Khosravi, Efficient Shot Boundary Detection with Multiple Visual Representations. Mobile Information Systems, 2022(1), (2022) 4195905. https://doi.org/10.1155/2022/4195905

[27] Z. Yang, L. Tian, C. Li, December. A Fast Video Shot Boundary Detection Employing OTSU's Method and Dual Pauta Criterion. In 2017 IEEE International Symposium on Multimedia (ISM) IEEE, (2017) 583-586. https://doi.org/10.1109/ism.2017.114

[28] W. Tong, L. Song, X. Yang, H. Qu, R. Xie, CNN-based Shot Boundary Detection and Video Annotation. In 2015 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting IEEE, (2015) 1-5. https://doi.org/10.1109/bmsb.2015.7177222

[29] R. Liang, Q. Zhu, H. Wei, S. Liao, A Video Shot Boundary Detection Approach based on CNN Feature. In 2017 IEEE International Symposium on Multimedia (ISM) IEEE, (2017) 89-494). https://doi.org/10.1109/ism.2017.97

[30] R.M. Bommisetty, P. Palanisamy, A. Khare, Content Based Video Retrieval—Methods, Techniques and Applications. In: Dash, S., Pani, S.K., Abraham, A., Liang, Y. (eds) Advanced Soft Computing Techniques in Data Science, IoT and Cloud Computing. Studies in Big Data, Springer, Cham, 89, (2021). https://doi.org/10.1007/978-3-030-75657-4_4

[31] E.M. Saoudi, S. Jai-Andaloussi, A distributed Content-Based Video Retrieval system for large datasets. Journal of Big Data, 8(87), (2021). https://doi.org/10.1186/s40537-021-00479-x

[32] S.H. Abdulhussain, A.R. Ramli, M.I. Saripan, B.M. Mahmmod, S.A.R. Al-Haddad, W.A. Jassim, Methods and Challenges in Shot Boundary Detection: a review. Entropy, 20(4), (2018) 214. https://doi.org/10.3390/e20040214

[33] Z. Li, X. Liu, S. Zhang, Shot Boundary Detection based on Multilevel Difference of Colour Histograms. In 2016 First International Conference on Multimedia and Image Processing (ICMIP) IEEE, (2016) 15-22. https://doi.org/10.1109/icmip.2016.24

[34] V.L. Narla, G. Suresh, M.K. Singh, V. Kumar M, Speech Signal Splicing Detection System based on MFCC and DTW. International Research Journal Multidisciplinary Technovation, 6(6), (2024) 170–181. https://doi.org/10.54392/irjmt24613

[35] TV9 Gujarati News Channel, Gujarati news video dataset. (2024). [Online]. Available: https://www.tv9gujarati.com

Authors Contribution Statement

Avani Bhuva: Conceptualization, Methodology, Writing - Original Draft. Dhirendra Mishra: Validation, Writing - Review & Editing. Both Authors Read and Approved the Final Version of the Manuscript.

Funding

The authors declare that no funds, grants or any other support were received during the preparation of this manuscript.

Competing Interests

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Data Availability

The data supporting the findings of this study can be obtained from the corresponding author upon reasonable request.

Has this article screened for similarity?

Yes