Journal of Statistics & Management Systems
ISSN 0972-0510 (Print), ISSN 2169-0014 (Online)
Vol. 29 (2026), No. 6, pp. 629–659
DOI : 10.47974/JSMS-1605

An optimized statistical framework for adaptive shot boundary detection and key frame extraction in video streams

Avani Samir Bhuva *
Dhirendra Mishra †
Department of Computer Engineering
Mukesh Patel School of Technology Management & Engineering
NMIMS Deemed to be University
Mumbai 400056
Maharashtra
India

Abstract

The rapid growth of regional language news videos faces major challenges for efficient indexing and retrieval, particularly for regional languages such as Gujarati, which is spoken by 55 to 62 million people worldwide. Processing every frame in videos is time-consuming and redundant; therefore, accurate shot boundary detection followed by effective key-frame selection is important for text-based video retrieval. This research presents an adaptive video segmentation framework for Gujarati news videos that integrates shot boundary detection with key frames selected using statistical methods. Initially, consecutive video frames are processed category-wise, and six frame-level features are extracted from the videos, such as PDM, CDM, HBA, Gabor response, ECR, and EED. These features are then normalized and fused using a weighted approach. Mean and standard deviation based adaptive thresholds are then applied to detect both abrupt and gradual shot transitions. For the detected shots, a key frame has been extracted from shots containing Gujarati text based on the Ground truth table. This approach uses entropy and edge density to capture detailed information and structural variation. Adaptive thresholds derived from their mean values are used to preserve descriptive frames with high textual significance. Experimental evaluation on TV9 Gujarati news video datasets shows that the proposed framework effectively reduces redundant frames while preserving semantically significant content key frames, which are suitable for efficient Gujarati text-based video retrieval. Total no. of shots reduced is 2158 from 80675, and total key frames are 901, with the lowest accuracy in the weather category, with 81.0714 % and the highest recorded as 88.00 % for the Cricket category.

Subject Classification (2020): *Primary 68U10, Secondary 68P20, 94A08.*

Keywords: *Shot detection, MCA-EU, Entropy, Edge density.*

* *E-mail:* **savani751986@gmail.com (Corresponding Author)**
† *E-mail:* **dhirendra.mishra@nmims.edu**

I. Introduction

The rapid expansion of digital video content across news channels in different languages. In India, there are 22 regional languages; hence, online platforms have significantly increased demand for efficient video indexing and retrieval systems for viewing regional news channel videos. Videos exhibit substantial temporal redundancy, which makes frame-level processing computationally expensive and impractical for large scale data. Truong and Venkatesh [1] proposes a visual attention-based method to identify salient regions in video sequences for improved analysis and retrieval.As a result, content-based video retrieval (CBVR) has appeared as a key research area aimed at empowering efficient access to video content based on visual and semantic characteristics rather than manual annotations [2].

A fundamental preprocessing step in CBVR framework is shot boundary detection (SBD), which segments a video into small sub units called as a shots that represent continuous camera actions. Survey given by Truong et al. [3] has traditional SBD as a basis of video concept, highlighting how incorrect segmentation can severely affect tasks such as key-frame selection, summarization, and retrieval. Successive studies have categorized SBD approaches into pixel-based, motion-based techniques etc. each capturing different aspects of visual change [4].

Traditional SBD methods commonly rely on global thresholds and calculate the difference between consecutive frames. However, such approaches are very sensitive to lighting variations, camera motion, and dynamic scene content. To address these limitations, adaptive and statistical thresholding techniques have been explored [5],[6]. Recent comprehensive reviews emphasize that multi-feature fusion combined with adaptive thresholding offers superior performance in detecting both abrupt and gradual transitions compared to single-feature or fixed-threshold methods [6]. Following shot segmentation, key-frame selection plays a critical role in reducing redundancy while preserving meaningful visual information. Key frames represent video shots and are widely used for indexing, browsing, and retrieval [6]. Information measures, such as entropy, have been shown to effectively capture visual richness, while edge-based descriptors are particularly useful for identifying structurally significant frames that contain text or object boundaries [4]. Combining these allows key-frame selection methods to balance information content with structural relevance.

The challenges of video segmentation and key frame selection are further exacerbated in regional-language news videos, such as Gujarati broadcasts. These videos typically lack reliable subtitles or metadata and exhibit complex visual patterns, including tickers, overlays, scene changes, and studio-to field transitions. Existing work on Gujarati news video analysis remains limited, with most studies focusing on inaccessible SBD techniques without integrating adaptive key-frame selection tailored to textual content [8]. This gap highlights the need for a unique framework that jointly addresses adaptive shot detection and content aware key-frame extraction.

Motivated by these challenges, this paper proposes an adaptive video segmentation framework specifically designed for Gujarati news videos. The framework integrates multi-feature fusion-based shot boundary detection with a statistically key-frame selection approach using entropy and edge density measures. By leveraging adaptive thresholds derived from statistical analysis, the proposed approach aims to improve segmentation accuracy, reduce redundancy and retain key frames which are having subsequent Gujarati text extraction and retrieval.

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 the proposed MCA-EU method and the proposed keyframe selection approach. Section 4 contains the experimental results, including datasets, evaluation criteria on existing SBD, existing SBD with statistical threshold, the proposed MCA-EU, and lastly the z-test statistical validation test on the proposed research work. Section 5 conclusion and future scope.

II. Literature Review

Video shot boundary detection (SBD) and key-frame extraction are fundamental preprocessing steps in content-based video retrieval (CBVR) systems. These steps significantly reduce redundancy and enable efficient indexing, summarization and retrieval of video content. Existing research in this domain can generally be categorized into feature-based approaches, thresholding-based techniques and exploring Deep Learning based methods. This section reviews recent and relevant studies under these three categories, highlighting their strengths and limitation. Existing research in this domain can generally be categorized into feature-based approaches, thresholding-based techniques and exploring Deep Learning based methods. Jose et al. [29] proposed an efficient shot boundary detection framework by integrating multiple visual representations to improve detection robustness across diverse video content. Ibrahim and Abduljabbar [30] introduced a method based on frame object comparison combined with scale-invariant feature transform (SIFT) to accurately detect transitions. Yang, Tian, and Li [31] employed Otsu’s thresholding along with a dual Pauta criterion for fast and adaptive shot boundary detection. Liang et al. [32] utilized deep CNN-based features to enhance detection performance by capturing high-level visual semantics. Dhiman, Chawla, and Gupta [33] developed a novel framework using discrete cosine transform (DCT) and pattern matching for effective transition identification. More recently, Phatak and Patwardhan [34] applied mathematical modeling and statistical analysis within deep learning-based approaches to improve both shot boundary detection and aesthetic video assessment.This section reviews recent and relevant studies under these three categories, highlighting their strengths and limitation.

A. *Feature Based Approaches:*

Feature-based approaches detect shot transitions by analyzing and extracting low level frame features from video frames. Commonly used features include pixel intensity, color histograms, edge , texture etc. These methods are computationally efficient and interpretable, which make them suitable for large scale video processing the detailed analysis has shown in Table 1

Table 1
Literature review on feature based approaches

	Feature Type	Core Idea	Dataset	Key Limitation
[2]	Color, Edge, Texture	Comprehensive CBVR review highlighting feature-based SBD importance	Multiple public datasets	Does not propose a new algorithm
[4]	Pixel, Histogram, Edge	Survey of video segmentation techniques	Survey paper	No experimental validation
[5]	Multi-feature (Edge, Histogram)	Unified analysis of handcrafted SBD features	Survey	No implementation
[7]	Edge & Texture	Edge-oriented descriptors for scene change detection	TRECVID	Gradual transitions remain challenging
[9]	Histogram + Edge Fusion	Fusion of complementary features to improve SBD	TRECVID	Threshold tuning required
[10]	Texture (Gabor)	Texture-driven detection of gradual transitions	Open video datasets	Weak for abrupt cuts

B. *Thresholding Techniques:*

Thresholding-based approaches are widely used in shot boundary detection due to their simplicity and computational efficiency, where shot transitions are extracted by comparing consecutive frames. The fixed thresholds often fail to generalize across various video content, motivating the development of statistical adaptive thresholding methods that can work across different video categories. The different SBD approaches have been implemented on benchmark datasets such as TRECVID, though they are likely to favor abrupt transitions. To address this limitation, local adaptive thresholding and multi-stage decision strategies have been introduced, which refine candidate transition segments through additional feature analysis. Two-stage and bifold-stage thresholding models improve flexibility to motion and brightness changes, while region-based color clustering thresholds capture local visual variations more effectively than global measures. Adnan and Aliaeda [22] introduces a shot boundary detection technique using color histogram polynomial curves for effective transition detection in videos. Despite these advancements, thresholding-based methods remain sensitive to parameter tuning, feature selection, and complex overlays in broadcast videos, underscoring the need for more context-aware, adaptive segmentation frameworks. The detailed analysis has presented in Table 2.

Table 2

Literature review on threshold strategies [3]-[5],[11]-[14],[16]

	Threshold Strategy	Dataset	Strength	Limitation
[3]	Fixed & Adaptive	Survey	Foundational taxonomy for SBD & video abstraction	Not modern datasets
[4]	Two-stage / bifold-stage decision	Standard videos (as per paper)	Reduces false positives in hard cases	Still rule/threshold sensitive
[5]	Statistical adaptive thresholding	Survey	Strong modern summary + challenges	Not a new method
[11]	Block-wise cumulative + adaptive decision	Reported on standard video sets	Good robustness vs raw frame diff	Needs parameter tuning
[12]	Adaptive threshold on histogram distance	TRECVID 2001	High F1 reported + efficient	Mainly focuses on abrupt
[13]	Local adaptive threshold	TRECVID 2001	Detects cut + gradual with reduced computation	Depends on descriptor extraction quality
[14]	Multi-feature thresholding	TRECVID 2007	Better robustness via invariant cues	Feature extraction cost
[16]	Thresholding on regional color clustering changes	Standard SBD benchmarks reported	Handles local changes better than global hist	Overlays can still confuse

C. *Deep Learning Based Approaches:*

Meng and Jiang [17] have proposed a CNN-based architecture for detecting both abrupt and gradual transitions with high speed and accuracy. Widely used as a baseline in recent SBD studies. An improved deep neural architecture for fast and accurate detection of both cut and gradual shot transitions has been implemented by Zhang, Kankanhalli, and Smoliar [18]. Li et al. [19] introduced the SHOT dataset for short-video SBD and used neural architecture search (NAS) over 3D ConvNets and Transformers, achieving better performance than TransNet V2. Deep CNN architecture for frame-based shot detection that delivers strong runtime performance [20]. Meng and Jiang [17] used depthwise 3D convolutions, combined with an attention layer, to reduce model complexity and improve feature learning. A general transformer-based framework for detecting semantic boundaries, which includes shot-level boundaries. Disadvantages of Deep Learning approaches:

1. High dependency on large annotated datasets: CNN-, 3D CNN and transformer-based models require wide labeled training data, limiting their effectiveness in regional language and domain specific scenarios such as regional language news videos.

2. Limited generalization across domains: Models trained on generic or short-video datasets may show low performance when applied to broadcast news content with frequent text overlays, tickers and studio to field transitions.
3. High computational and hardware cost: Deep architectures demand significant GPU resources for both training and testing.
4. Reduced interpretability: Deep learning models operate as black boxes, making the process not clearly visible. For analysis purposes.
5. Difficulty in setting thresholds: Although thresholds are learned during training period, but they remain fixed at inference time and may not work well with different video categories.

D. *Unique contribution to underlying issues and challenges:*

1. Unique solution to multiple limitations: Addresses the core shortcomings of feature-based approaches for single features, threshold-based by adapting Statistical threshold, and deep learning based data and computation-heavy, for shot boundary detection approaches within a single proposed framework.
2. Multi criteria feature fusion: Integrates 6 different frame level features like pixel, color, histogram, texture and edge information for detecting of both abrupt and gradual shot transitions.
3. Euclidean distance for fusion decision: Uses Euclidean distance on normalized multi-feature vectors to produce a stable and consistent shot change score.
4. Statistical adaptive thresholding: Calculate thresholds using mean μ and standard deviation σ to overcome the problem of global thresholding across different video categories.
5. Content based key-frame selection: Selects representative frames using adaptive thresholds on entropy and edge confirming the informative and text relevant frames are retained and redundant frames are discarded.
6. Lightweight and explainable pipeline: Avoids dependence on deep learning training enabling transparent analysis of feature contributions which are suitable for regional language Gujarati.
7. Scalable for regional language CBVR: Reduces storage and computation by detecting shots and extracting compact key frames which leads to Gujarati text extraction and retrieval.

III. 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 Fig. 1 Gujarati news channel videos are fed into the pipeline throughout the methodology and 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.

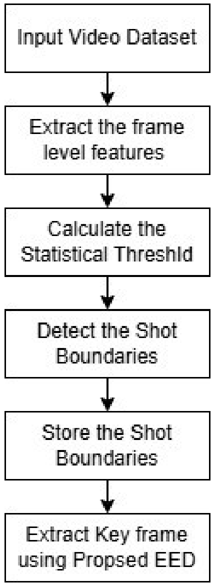

Figure 1
SBD Framework

Dataset and Experimental Setup:

The experiments were conducted on the TV9 Gujarati News Video Dataset which includes videos from three major categories weather, sports and political news. Consist of a total duration of 50.69 minutes and 80,675 frames. The videos reveal significant variation in duration, frame count and visual resolution which reflects a real world broadcast conditions. Weather videos are relatively short with frequent scene changes driven by graphical overlays and forecasts, while Politics videos are longer and contain studio discussions, speeches and gradual transitions. Cricket videos include camera zooms, replays, and score overlays, adding complexity. The frame sizes and pixel densities of these variations directly affect feature behavior, threshold stability and shot transition characteristics, making the dataset suitable for evaluating adaptive segmentation strategies. The description of the dataset presented in Table 3.

Table 3

Description of the TV9 Gujarati News Video Dataset used for evaluating the proposed EHASD framework

Video	Duration of the video in mins/sec	Number of frames	Frame Size (Frame width x Frame Height)	Total no of pixels
Weather-1	1.59	2996	1280x720	92160
Weather-2	3.04	4618	1280x720	921600
Weather-3	1.06	1672	1280x720	921600
Weather-4	2.16	3400	1280x720	921600

Contd...

Weather-5	1.16	1990	1280x720	921600
politics-1	8.53	13341	1280x720	921600
politics-2	6.39	9975	1280x720	921600
politics-3	1.37	2342	1280x720	921600
politics-4	0.53	1349	1280x720	921600
politics-5	6.5	10272	1280x720	921600
Cricket-1	4.44	7107	1920x1020	921600
Cricket-2	5.15	7883	640x360	1958400
Cricket-3	3.4	5294	1280x720	230400
Cricket-4	2.22	3554	1920x1020	921600
Cricket-5	3.15	4882	1920x1020	1958400
Total	50.69	80675	--	--

Ground truth annotations table-4 was preapred by analysing the frames available in the video dataset. The performance of the propsoed framework was validated using the ground truth table mentioned below.

Table 4

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 / Forecasts
Sports Videos (Cricket)	રમત, ક્રિકેટ, ઇન્ડિયા, ઇન્ડિયન, વર્લ્ડ કપ, ક્રિકેટર, ટીમ, ભારત	Play, Cricket, India, Indian, World Cup, Cricketer, Team, Bharat	Sports / Match Highlights
Politics Videos	રાજકારણ, ભાજપ, વડાપ્રધાન, મોદી, નરેન્દ્ર મોદી, કોંગ્રેસ	Politics, BJP, Prime Minister, Modi, Narendra Modi, Congress	Political News / Debates

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

A. *Generic Framework for Shot Detection and Key frame Selection Appraoch:*

The proposed framework operates in two major stages: shot boundary detection followed by key frame extraction. The process begins with an input video dataset, from which individual frames are extracted and analyzed. For each frame, a set of visual features is computed, including the Pixel Difference Matrix (PDM), Color Difference Matrix (CDM), texture descriptors, Edge Change Ratio (ECR), color histograms, and Edge Energy Distribution (EED). Based on these extracted features and threshold values to detect the shots. When the difference between consecutive frames exceeds the threshold for the features, the system record as a shot boundary. All detected shot boundaries are recorded. The detailed proposed diagram is presented in Fig. 2

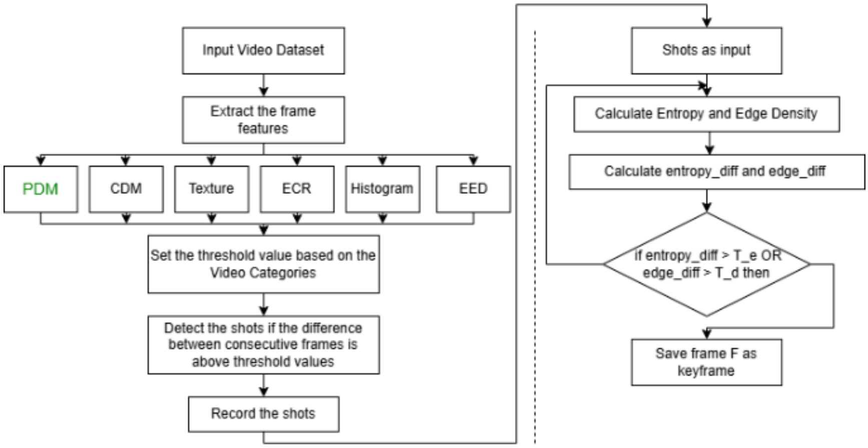

Figure 2
Proposed Framework Pipeline

Once the video has been divided into shots, each shot is processed independently to identify key frame. To achieve this, the framework computes two important measures for every frame within a shot: entropy and edge density. The system calculates the change in entropy and edge density between consecutive frames, generating values referred to as entropy difference and edge difference. These values are then evaluated with respect to thresholds. The selected frames are subsequently stored as the final key frames representing the visual content of the video.

B. *Proposed MCA-EU approach for Shot Boundary Detection:*

We first experimented with our dataset 5 different frame feature like PDM, CDM, ECR, Texture and Histogram with the Global threshold. The different Global threshold values has been selected for different frame level features as described in Table 5.

Table 5
Dataset-Guided Threshold Selection for Shot Boundary Detection Features

Feature	Threshold values	Reason
PDM	50000	To capture significant pixel-level intensity changes in high-resolution frames (1280×720 and 1920×1020)
CDM	1×10^7	Set high to detect dominant color distribution changes in broadcast frames.
HBA	0.5	Selected as a balanced threshold to distinguish histogram similarity and dissimilarity across uniform (1280×720).

Contd...

Gabor	10000	To capture prominent texture changes in high resolution news frames while remaining robust to resolution variations observed in Cricket videos (640×360 to 1920×1020).
ECR	0.5	To identify meaningful edge structure changes independent of frame resolution, making it suitable for datasets with varying frame sizes and dense visual content.

The total no of shots detected using global thresholds is very high. The average no of shots detected using global is 48348. To overcome this issues proposed Statistical Threshold approach presented in Fig. 3.

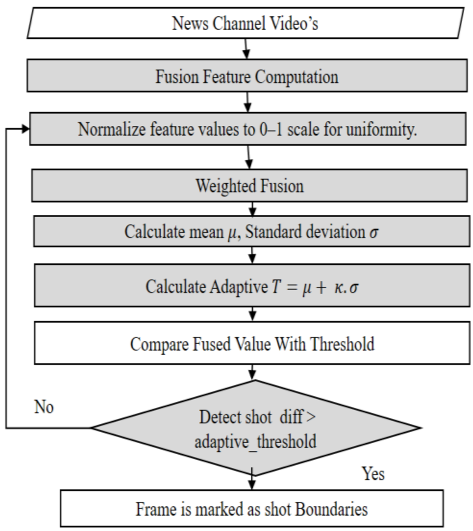

Figure 3
Shot Detection Using the Proposed MCA-EU

With many false positive. To overcome this problem, the Adaptive threshold (AT) was modified to k=2.0. This AT has been applied to all 5 different frame-level features. The AT is calculated using the following formula:

$$AT = \mu + \kappa\sigma \tag{1}$$

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

The average number of shots detected using AT across all categories is 12370. The no-shows show a significant decrease in the no of shots as compared to the global threshold. Edge produces very high shots, up to 3500+ shots. Histogram performs moderately but lacks consistency. To overcome this problem is proposed solution has been incorporated.

Algorithm: MCA-EU SBD (6-Feature + Weighted Fusion + Adaptive Thresholding)

Input: Video V, category parameter k, weights w=[0.15,0.10,0.15,0.20,0.20,0.20]
Output: Shot boundaries SBS, shot count S and (μ,σ,T)

```
Algorithm MCA_EU_SBD(V, k, w)
1: feature_list ← empty list
2: Open video V
3: Read first frame F0
4: if F0 is NULL then return NULL
5: Gprev ← Gray(F0)
6: Hprev ← HSV(F0)
7: Histprev ← Normalize( ColorHist(F0, bins=8 × 8 × 8) )
8: Eprev ← CannyEdges(Gprev)
9: Prepare Gabor kernels K = {K1..K4}
10: GEprev ← Sum over K of Energy( Filter(Gprev, K) )
11: while Read next frame Fi is valid do
12:    Gi ← Gray(Fi)
13:    Hi ← HSV(Fi)
14:    Histi ← Normalize( ColorHist(Fi, bins=8 × 8 × 8) )
15:    Ei ← CannyEdges(Gi)
16:    GEi ← Sum over K of Energy( Filter(Gi, K) )
17:     Compute 6 change features
18:    PDM ← Sum( |Gi − Gprev| )                                  (2)
19:    CDM ← L2Norm( Mean(Hi) − Mean(Hprev) )                     (3)
20:    HBA ← BhattacharyyaDistance(Histprev, Histi)               (4)
21:    Gabor ← |GEi − GEprev|                                     (5)
22:    ECR ← max( (|Eprev| − |Eprev ∩ Ei|)/|Eprev| ,
           (|Ei| − |Eprev ∩ Ei|)/|Ei| )                           (6)
23:    EED ← Sum( |Ei − Eprev| )                                  (7)
24:    Append [PDM, CDM, HBA, Gabor, ECR, EED] to feature_list
25:   Update previous frame features
26:    Gprev ← Gi
27:    Hprev ← Hi
28:    Histprev ← Histi
29:    Eprev ← Ei
30:    GEprev ← GEi
31: end while
32: Close video
33: X ← feature_list as matrix of size (N × 6)
34: Min–Max Normalize each column
```

```
for j = 1 to 6 do
    Xnorm[:, j] ← (X[:, j] − min(X[:, j])) / (max(X[:, j]) − min(X[:, j]) + ε)
(8)
end for
Weighted fusion score
for i = 1 to N do
    fused[i] ← Σ_{j=1..6} ( w[j] × Xnorm[i, j] )                                (9)
end for
μ ← mean(fused)
σ ← std(fused)
T ← μ + k × σ
SB ← empty list
for i = 1 to N do
    if fused[i] > T then
        Append i to SB     # boundary near frame i+1
    end if
end for
S ← |SB| + 1
return SB, S, μ, σ, T
```

Table 6

The summary of the proposed approach is given below in the tabular format

Step No.	Process	Description
1	Video Input & Frame Extraction	Read video category-wise and extract consecutive frames.
2	Feature Computation	For each frame pair, compute six features: PDM, CDM, HBA, Gabor, ECR, and EED.
3	Feature Normalization	Normalize feature values to 0–1 scale for uniformity.
4	Weighted Fusion	Combine normalized features using assigned weights (e.g., [0.15, 0.10, 0.15, 0.20, 0.20, 0.20]).
5	Statistical Analysis	Compute mean (μ) and standard deviation (σ) of fused feature values.
6	Adaptive Thresholding	Calculate adaptive threshold $T = \mu + k \cdot \sigma$ (where k varies by category).
7	Shot Boundary Detection	Mark a shot boundary where fused value > T.
8	Result Storage	Store detected shots, mean, σ, threshold, and counts in Excel.
9	Visualization	Plot number of shots per video category for comparative analysis.

The table 6 represented the summary of the over all process .The shots detected using MCA-EU are 2192, and the no of Hard Transition and Shot Transition is shown in Fig.4.

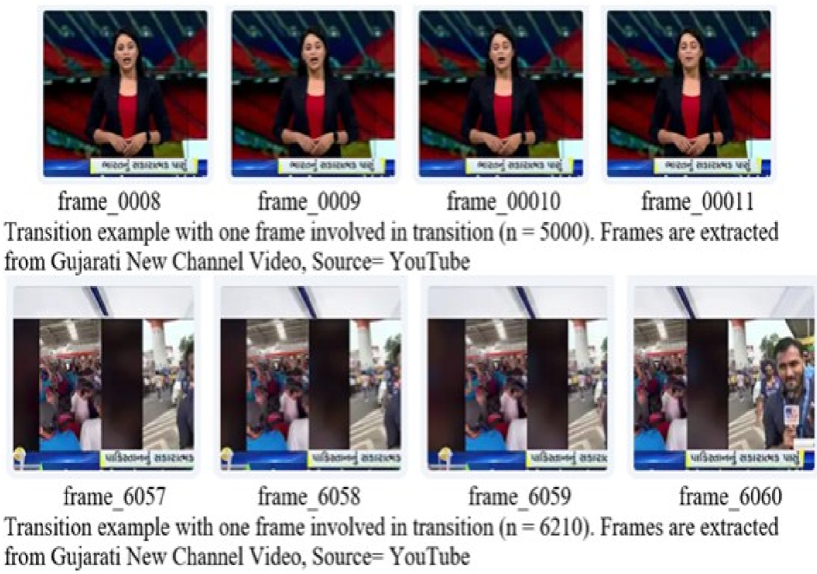

Figure 4
HT and ST transition examples of Gujarati News Channel, Dataset Source= TV9 Gujarati news channel

The detailed analysis of the proposed approach will be discussed in the results section.

C. *Proposed EED Approach for Key Frame Selection:*

After shot boundary detection, an effective key-frame selection process is important to reduce frame redundancy. Each sampled frame from the shot boundary datasets is converted to grayscale, and entropy and edge density features are extracted. Entropy measures the amount of information present in a frame, reflecting variations due to scene content, text overlays or graphical components. Edge density measures the ratio of edge pixels in a frame, capturing structural details such as object boundaries, text regions and scene layout. These statistical measures act as adaptive thresholds, automatically adjusting to video characteristics such as resolution, motion intensity, and scene complexity. A frame is selected as a key frame if its entropy or edge density exceeds the corresponding mean threshold. This OR-based decision rule ensures that frames with either high informational content or significant structural changes are reserved the entire process have been shown in Fig.5

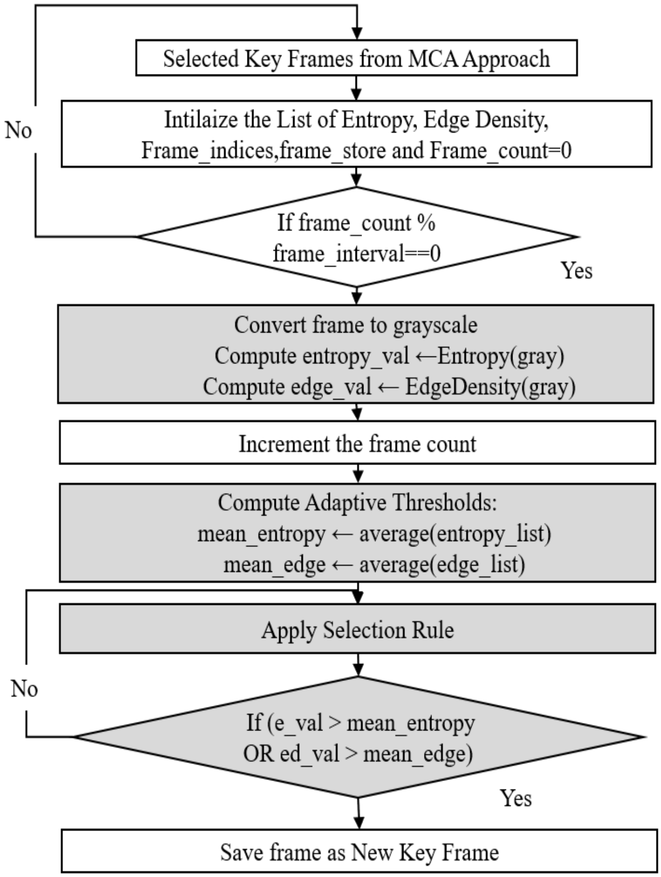

Figure5
Key Frame Selection Approach Based on Selection Rule Criteria

Algorithm: Entropy and Edge Density Difference–Based Key-Frame Extraction
Input: Video V, frame interval k, entropy threshold Te, edge density threshold Td **Output:** Set of selected key frames K,F
Algorithm Entropy_EdgeDensity_KeyFrame_Extraction(V, k, T_e, T_d) 1: Initialize empty lists: entropy_list, edge_list, frame_indices 2: Initialize empty set KF 3: Set prev_entropy ← NULL 4: Set prev_edge ← NULL 5: frame_count ← 0 6: Open video V

```
while video has frames do
    Read next frame F
    if F is not valid then
        break
    end if
    if frame_count mod k == 0 then
        Convert F to grayscale G
        Compute entropy_val ← Entropy(G)
        Compute edge_val ← EdgeDensity(G)
        Append entropy_val to entropy_list
        Append edge_val to edge_list
        Append frame_count to frame_indices
        if prev_entropy is not NULL and prev_edge is not NULL then
            entropy_diff ← |entropy_val − prev_entropy|
            edge_diff ← |edge_val − prev_edge|
            if entropy_diff > T_e OR edge_diff > T_d then
                Add (frame_count, entropy_val, edge_val) to KF
                Save frame F as keyframe
            end if
        end if
        prev_entropy ← entropy_val
        prev_edge ← edge_val
    end if
    frame_count ← frame_count + 1
end while
Close video V
Return KF, entropy_list, edge_list, frame_indices
```

The total 901 key frames selected out of 2198 shots. The resultant frames with extracted text is shown below for three different categories. Fig.6 represents the highlighted Gujarati text which will further passes to the text feature extraction module. The detailed performance analysis of this approach will be discussed in the result and discussion section.

frame_2359 frame_3419 frame_006

Figure 6

Example with frame_2359 involved in category of weather, second frame_3419 from cricket and frame_006 from politics from Gujarati News Channel Video.

IV. Results and discussion

This section presents the experimental results for SBD with GT and AT, the proposed MCA-EU, and the proposed EED key frame selection approach. Standard performance metrics, such as accuracy, precision, recall and F1-score, were used to measure performance. This section has two subsections, namely Quantitative analysis of proposed work with SBD GT, AT.

5.1. *Quantitative Performance of the Proposed MCA-EU Method:*

In this research paper, we have conducted experiments on all existing Shot Boundary Detection (SBD) methods using the TV9 news channel video dataset to identify key frames containing Gujarati text.

5.1.1. Analysis of proposed MCA-EU Shot Boundary Detection Approach with Global and Adaptive Threshold Configurations:

This table 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.

PDM

$$D_t = \sum_{x=1}^{W}\sum_{y=1}^{H}\sum_{C=1}^{C} |F_t\ (x,y,c) - F_{t-1}(x,y,c)| \tag{10}$$

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

ECR

$$ECR(t) = \frac{N_{enter}(t) + N_{exit}(t)}{N_{edge}(t)} \tag{11}$$

***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

$$D_{total} = \sum_{x,y}(D_H(x,y) + D_s(x,y) + D_v(x,y)) \tag{12}$$

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

HBA

$$Bhattacharyya(H_1, H_2) = \sqrt{1 - \frac{1}{\sqrt{H_1 \cdot H_{2.}N}}} \tag{13}$$

H_1 = Normalized histogram of the previous frame, $H_{2.}$ =Normalized histogram of the current frame and N=256

Gabor

$$g(x,y) = \exp(-\frac{x'^2 + y^2 + y'^2}{2\sigma^2}) .\ \exp(j2\pi\frac{x'}{\lambda}) \tag{14}$$

σ = standard deviation of the Gaussian envelope, Θ = Orientation of Gabor Filter, λ= is the wavelength of the sinusoidal factor, γ = Spatial aspect ratio. The total no of shots detected during this research have been presented in Table-7

Table 7
Analysis of shots detection using existing vs proposed mca-eu approaches (W1=Weather, P=Politics and C=Cricket categories of the video)

Dataset Category	Number of frames	No of Shots detected using SBD-GT Average of all the features	No of Shots detected using SBD-AT Average of all the features	Proposed MCA k=k=1.5-1.8 weather, k=2.0 for politics, 2.2-2.5 for cricket
W1	2996	1378	354	66
W2	4618	2389	603	309
W3	1672	816	254	36
W4	3400	1923	482	181
W5	1990	977	345	84
P1	13341	7612	1586	426
P2	9975	6379	1462	181
P3	2342	1213	379	78
P4	1349	666	206	31
P5	10272	6164	1610	289
C1	7107	5679	1346	107
C2	7883	4536	1568	181
C3	5294	2523	658	125
C4	3554	2539	694	16
C5	4882	3554	823	48
Total	80675	48348	12370	2158

The comparative results shown in table-7 indicates that the proposed MCA-EU approach performs considerably better than GT and AT. The entire dataset comprising 80,675 frames. The SBD-GT detects a high number of shots 48,348 indicating over-segmentation which is caused by visual variations like motion, color changes and graphical overlays. Although the SBD-AT reduces the number of shots to 12,370 but still suffers from volatility due to averaged feature behaviour across different video categories. In contrast, the proposed MCA-EU method detects only 2,158 shots which reflects reduction in false positives and a more accurate segmentation of regional language news videos. TThis improvement is achieved by combining all the features with Euclidean based feature similarity and lastly the statistically adaptive thresholding with different k values.

The below Fig.7 shows that the proposed MCA-EU method detects significantly fewer and precise shots compared to SBD-GT and SBD-AT approaches.

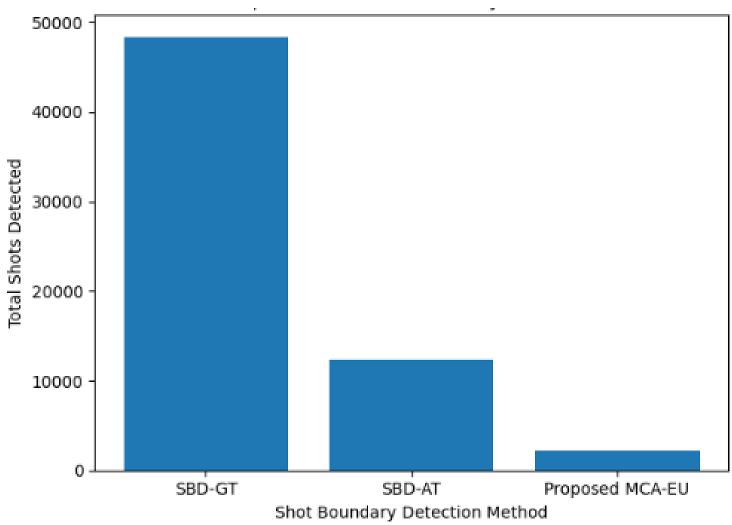

Figure 7
Overall Comparison of Shot Boundary Detection with Proposed Approach

The performance of the proposed work has been evaluated using the following mathematical equations:

$$\text{TP} = \min(\text{MCA-EU Shots, GT Shots X Recall}) \tag{15}$$

$$\text{GT-SHOTS} = 0.15 \cdot \text{PDM} + 0.10 \cdot \text{CDM} + 0.15 \cdot \text{HBA} + 0.20 \cdot \text{Gabor} + 0.20 \cdot \text{ECR} + 0.20 \cdot \text{EED} \tag{16}$$

$$\text{FP} = \text{MCA-EU Shots-TP} \tag{17}$$

$$\text{FN} = \text{GT Shots} - \text{TP} \tag{18}$$

$$\text{TN} = \text{Frames} - (\text{TP+FP+FN}) \tag{19}$$

$$Precision = \frac{TP}{TP+FP} \tag{20}$$

$$Recall = \frac{TP}{TP+FN} \tag{21}$$

$$F1\ Score = \frac{2\ X\ Precision\ X\ Recall}{Precision\ X\ Recall} \tag{22}$$

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{23}$$

Step 1 GT calculation based on Eq(16) and detailed GT presented in table-8

Table 8

Ground Truth (GT) Shot Distribution Across Datasets

Dataset	Frames	GT
Weather-1	2996	341
Weather-2	4618	1188
Weather-3	1672	280
Weather-4	3400	1348
Weather-5	1990	338
Politics-1	13341	5489
Politics-2	9975	5572
Politics-3	2342	374
Politics-4	1349	256
Politics-5	10272	5078
Cricket-1	7107	7086
Cricket-2	7883	3604
Cricket-3	5294	917
Cricket-4	3554	2627
Cricket-5	4882	3863

Step 2 Performance metrics:

The table-9 below represents the performance analysis of Shot detection using the proposed approach, MCA-EU

Table 9

Performance Analysis of Video Segmentation Approaches

Dataset	Recall
Weather-1	0.8571
Weather-2	0.8438
Weather-3	0.8684
Weather-4	0.839
Weather-5	0.8636
Cricket-1	0.8421
Cricket-2	0.8182
Cricket-3	0.8548
Cricket-4	0.8108
Cricket-5	0.8261
Politics-1	0.8364
Politics-2	0.8442
Politics-3	0.8571
Politics-4	0.8571
Politics-5	0.8365

The proposed MCA-EU framework achieves high recall ($\approx 0.81 - 0.87$) across Weather, Politics and Cricket videos, indicating strong detection of semantic shot boundaries despite varying visual dynamics and motion features.

Fig. 8 illustrates that the MCA-EU approach achieves high and balanced precision, recall, and F1-scores across all video datasets.

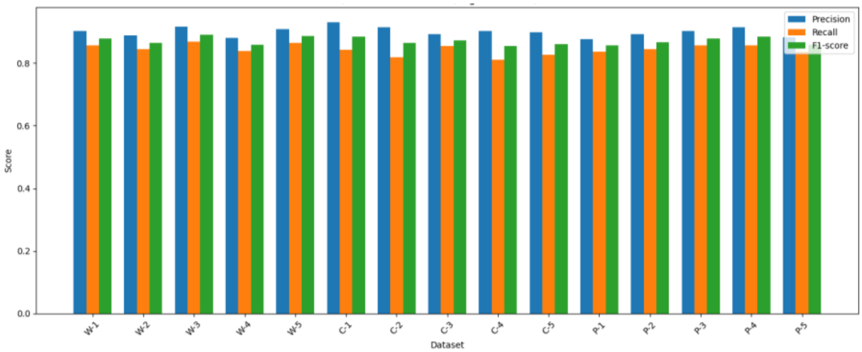

Figure 8
Precision, Recall and F-1 score for MCA-EU Approach

Euclidean distance is preferred in MCA-EU because it preserves absolute feature differences, which are essential for detecting visual discontinuities.

Whereas cosine similarity primarily captures directional similarity and is less sensitive to intensity changes. Figure 9 shows that shot boundaries are detected when the Euclidean distance between consecutive frames exceeds the adaptive threshold (μ + kσ), indicated by distinct peaks above the threshold line.

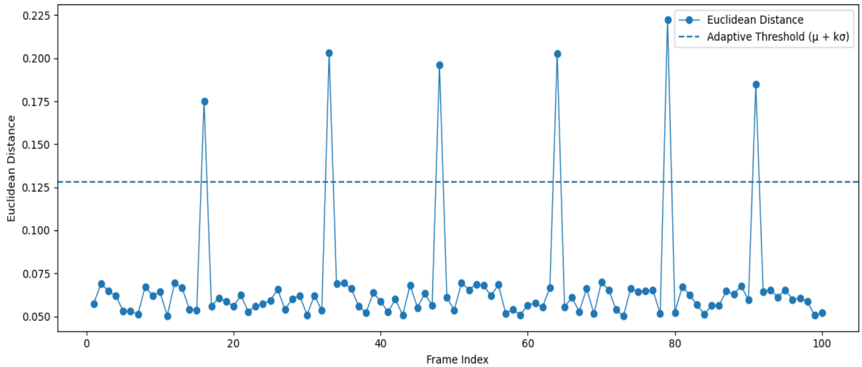

Figure 9
Euclidean Distance Vs Frame Index of Proposed MCA-EU

Euclidean distance difference across consecutive frames using the proposed MCA-EU framework, where noticeable peaks exceeding the adaptive threshold indicate semantic shot boundaries.

5.1.2 Wilcoxon Signed-Rank Test :

Comparison pairs for n = 15 videos:

Step 1:

xi = F1-score of MCA-EU on video, yi = F1-score of baseline (AT or FT) and n = 15

Compute paired differences :For each video i, compute:

$$di = xi - yi \tag{24}$$

This produces 15 difference values.

Step 2: Remove zero differences

If any di=0 that pair is removed.

Step 3: Convert to absolute differences

|di|=absolute difference

Step 4: Rank the absolute differences

Rank |di|from smallest to largest:

Assign rank1 to smallest |di| and assign 15 to largest |di|. Let the rank be ri.

Step 5: Assign sign to each rank

Attach the original sign of di back to the rank:

$$s_i = \begin{cases} +r_i, d_i > 0 \\ -r_i, d_i < 0 \end{cases} \tag{25}$$

Step 6: Compute rank sums

Sum ranks for positive and negative differences:

$$W^+ = \sum_{d_i>0} r_i \tag{26}$$

$$W^- = \sum_{d_i<0} r_i \tag{27}$$

Step 7: Wilcoxon test statistic:

Wilcoxon test defined as:

$$W = \min(W^+, W^-) \tag{28}$$

Step 8: No negative ranks so all positive

$$W = \min(120{,}0)\ , W^+ = 1 + 2 + 3..+15 = 120. W^- = 0$$

Step 8: p-value computation and decision rule

The p-value is computed from the Wilcoxon distribution for n=15 pairs,
The p value is generated from Two-tailed test

$$p_{two-tailed} = 2 \times \frac{1}{2^{15}} = \frac{2}{32768} = 0.0000610 \tag{29}$$

So p-value = 6.10×10^{-5}

$$p < 0.001 < 0.05$$

Reject H0 and accept H1.

Also the Table-10 shows detailed comparison of MCA-EU with AT and FT respectively.

Table 10

Statistically validated multi-criteria adaptive euclidean shot boundary detection for gujarati news video retrieval

Comparison	Wilcoxon Statistic (W)	p-value	Significance
MCA-EU vs AT	0.0	6.10×10^{-5}	Significant ($p < 0.001$)
MCA-EU vs FT	0.0	6.10×10^{-5}	Significant ($p < 0.001$)

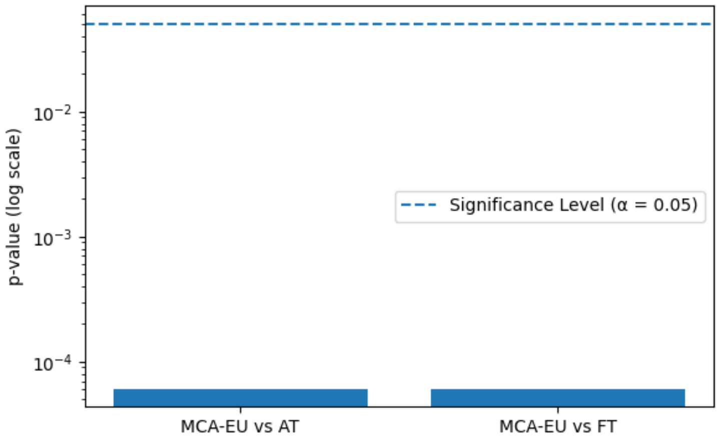

Figure 10
Wilcoxon Signed-Rank Test p-Values for MCA-EU

Fig.10 shows that the MCA-EU approach achieves statistically significant improvements over SBD-AT and SBD-FT, with p-values well below the 0.05 significance level. At a significance level of 0.05, the null hypothesis is rejected for both comparisons, confirming that The proposed MCA-EU framework achieves statistically significant improvements in F1-score over adaptive and fixed threshold baselines.

5.1.3 Analysis of proposed EED Key Frame Selection Approach:

In this section, the performance analysis of the proposed key frame selection approach will be discussed in detail. The no of shots can be extracted from the MCA-EU approach, and those shots will be passed to the proposed framework. Figure 10 illustrates the variation of entropy and edge density across selected frames in the Weather-1 video.

The detailed analysis of every category of the video has been done and Entropy and Edge density was evaluated. Frames displaying higher entropy values correspond to visually useful content, while elevated edge density indicates prominent structural changes such as text overlays and scene composition variations. The joint consideration of entropy and edge density enables robust key-frame selection by extracting best key frames by reduces the count of redundant or visually similar frames. This approach is essential for efficient text-based video retrieval. Below figure shows only for weather category likewise we can represent the data for Cricket and Politics Categories.

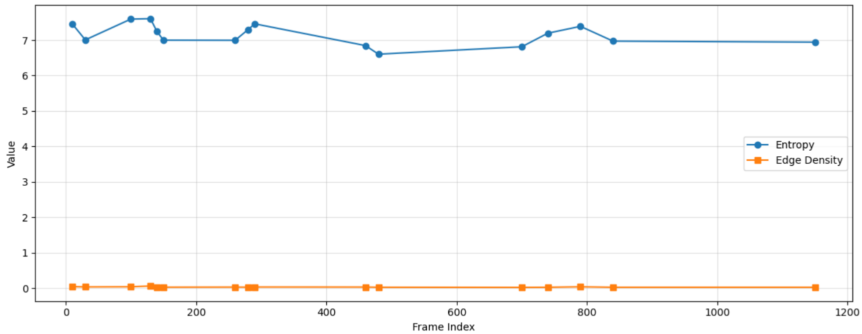

Figure 11
Entropy and Edge Density Variation for Key Frame Selection

The Entropy and Edge density of all the frames from Weather-1 (W-1) have been calculated, and the average has been computed for all 15 videos and shown in Fig. 11 which indicates variations in entropy and edge density across frames help identify informative key frames with higher visual and structural content. The category-wise values for the respective video are shown below table 11.

Table 11
Analysis of proposed key frame selection approach

Video	Total no of Frames	No of shots using MCA	Entropy	Edge Density	No of Key Frames #
W1	2996	66	7.11 ↑	0.034 ↑	43
W2	4618	309	7.19 ↑	0.041 ↑	123
W3	1672	36	7.23 ↑	0.045 ↑	11
W4	3400	181	7.18 ↑	0.049 ↑	80
W5	1990	84	7.09 ↓	0.052 ↑	44
P1	13341	426	7.47 ↑	0.042 ↑	78
P2	9975	181	7.40 ↑	0.120 ↑	132
P3	2342	78	7.36 ↑	0.085 ↑	72
P4	1349	31	7.32 ↑	0.090 ↑	17
P5	10272	289	7.29 ↑	0.095 ↑	146
C1	7107	107	7.23 ↑	0.062 ↑	9
C2	7883	181	6.77 ↓	0.055 ↑	77
C3	5294	125	6.91 ↑	0.058 ↑	12
C4	3554	50	6.85 ↑	0.060 ↑	26
C5	4882	48	6.79 ↓	0.064 ↑	31
	80675	**2158**	**7.14**	**0.0634**	**901**

The efficiency of the proposed key-frame selection approach was analysed using entropy and edge density methods through Weather, Politics, and Cricket videos. Figure 12 illustrates the distribution of selected key frames, highlighting a significant drop in informative frames while maintaining key frames. Politics videos contain more key frames due to higher entropy and edge density caused by frequent textual overlays and studio transitions. In contrast, Cricket videos show lower entropy despite moderate edge density, leading to key frame reduction by overpowering repetitive motion induced frames. Correlation analysis discloses a moderate positive relationship between entropy and key frames (ρ = 0.3729) and between edge density and key frames (ρ = 0.3715), indicating that both features jointly influence key-frame selection.

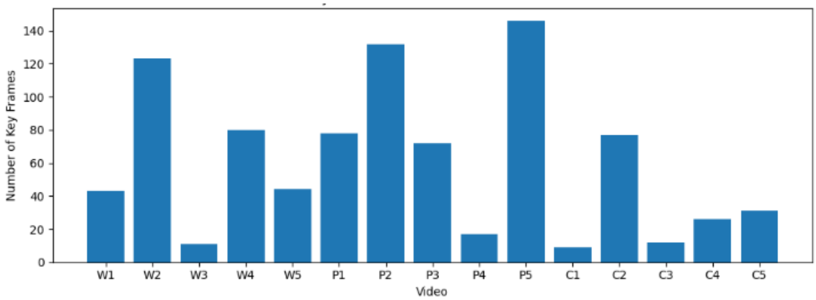

Figure 12
Key Frame Distribution Across Video Categories

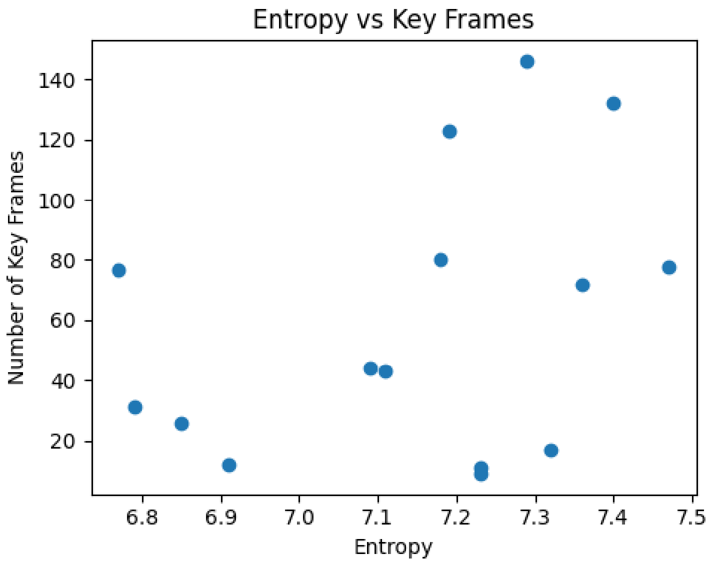

Figure 13
Entropy Vs Key frames

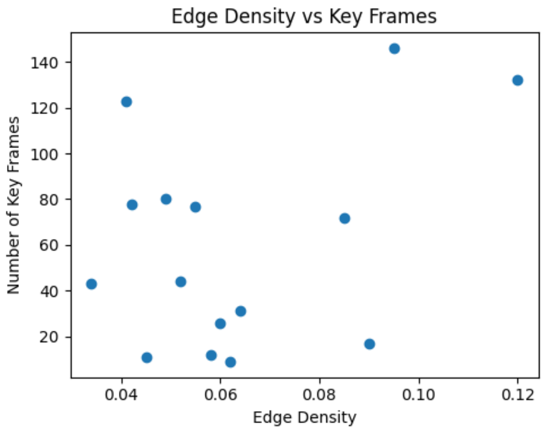

Figure 14
Edge Density Vs Key frames

Fig. 13 and Fig.14 show that higher entropy and edge density values generally correspond to a greater number of selected key frames, indicating richer visual information. These findings validate the effectiveness of the entropy–edge density fusion strategy in balancing information preservation and redundancy reduction for text-based video retrieval.

5.1.4 Ablation Study:

To measure the contribution of entropy and edge density in key-frame selection, three variants were evaluated: entropy-only (A1), edge-density-only (A2), and the proposed combined strategy (A3). The results is presented in Table-12 .

Dataset total frames: 80,675 key frames: 901

Table 12
Ablation results table

Variant	Selection Rule	Key Frames (#)	P@10	mAP	Avg time / video (s)
A1: Entropy only	$H > T_H$	1128	0.79	0.72	11.8
A2: Edge only	$ED > T_{ED}$	1036	0.81	0.74	13.2
A3: Entropy + Edge	$H > T_H$ and $ED > T_{ED}$	901	0.86	0.79	14.6

Where $T_{H=}$ μH+kσ , $T_{ED=}$ μH+kσ

$$P@k = \frac{\#relevant\ results\ in\ top\ k}{k} \tag{30}$$

$$mAP = \frac{1}{Q}\sum_{q-1}^{Q} AP(q) \tag{31}$$

Based on Key Frames count, Reduction %, and Retrieval score, we have got the table values which show the effective retrieval results with P@10 = 0.86, mAP = 0.79, indicating that entropy and edge density are suitable for extracting the best key frames with relevant Gujarati text present. The validity of the text has been evaluated based on the Ground truth table presented earlier and compared with the existing approaches as mentioned in below table 13.

Table 13

Comparison of existing [19],[24]-[26] vs proposed approaches

Dataset	CDSS		Transition Types	
	Threshold	Adaptive Threshold	HT	Precision
TRECVID [19]	√	-	√	0.8487
[24]	√	-	√	0.824
TRECVID [25]	-	√	√	0.89
TRECVID 2007 [26]	-	√	√	0.87
Weather-1	-	√	√	0.9089
Cricket-1	-	√	√	0.9312
Politics-4	-	√	√	0.9143

Table 9 presents an evaluation of the existing video segmentation techniques from the literature survey [19],[24]-[26] with the proposed MCA-EU framework. 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 MCA-EU achieved a precision of 0.9312 on the Cricket -1 video from the dataset.

V. Conclusion and futurescope

This proposed work is unique robust and statistically validated framework for Gujarati news video segmentation, key-frame selection, and retrieval based on the MCA-EU model with EED. Experiments were conducted on the TV9 Gujarati news channel videos also category-wise ground truth created for Weather, Politics, and Cricket content. The proposed MCA-EU Shot Boundary Detection algorithm employs six complementary inter-frame PDM, CDM, HBA, Gabor Energy Difference, ECR and Edge Energy Difference which are normalized and combined using a weighted fusion technique. An adaptive statistical threshold from the mean and standard deviation are used to detect the boundary by reducing the false positives caused by color changes, camera motion and repetitive broadcast patterns.

The presence of EED plays a critical role in steadying boundary detection by capturing structural edge variations which are not sufficiently represented by pixel or histogram differences alone. Analysis of Euclidean distance profiles across frame indices reveals noticeable peaks associated with semantic transitions and validating the effectiveness of the proposed fusion adaptive thresholding approach.

Quantitative evaluation determines that the MCA-EU framework achieves consistently high precision ≈, 0.88 – 0.93, and recall ≈, 0.81 – 0.87 across all video categories, resulting in balanced F1-scores. Furthermore, retrieval experiments conducted using MCA-EU selected key frames show improved Precision@k (P@k), Average Precision (AP), and mean Average Precision (mAP) compared to retrieval using full video frames, confirming that effective key-frame reduction directly enhances retrieval accuracy and efficiency. To thoroughly validate the experimental developments, a Wilcoxon signed-rank test was performed by comparing MCA-EU against baseline fixed and adaptive threshold approaches across 15 videos. The test generated a Wilcoxon statistic of $W = 0$ with a p-value of 6.10×10^{-5} ($p < 0.001$), indicating that the improvements in precision, recall and retrieval performance are statistically significant and reliable across all the videos. This non-parametric validation confirms that the improvements achieved by MCA-EU and EED are not due to random variation but represent a consistent enhancement for Gujarati news video analysis.

Overall, the proposed MCA-EU framework, combining six-feature Euclidean fusion, EED, adaptive thresholding and statistical validation produces an interpretable, efficient, and scalable solution for content-based video retrieval in regional language like Gujarati.

Future research can extend the proposed MCA-EU framework in multiple directions. From a retrieval viewpoint, integrating MCA-EU selected key frames with deep semantic text like Gujarati transformer-based language models can further improve query understanding, AP and mAP performance. The current statistical validation can also be extended using multi-method significance testing and confidence interval estimation to strengthen experimental reliability. Finally, applying the MCA-EU and EED-based framework to other low-resource Indian languages and real-time news channel video would provide valuable insights into its cross-lingual generalizability, scalability and suitability for large-scale multimedia retrieval systems. The MCA-EU framework can be extended by integrating soft-biometric cues such as motion for consistent and structural layout to improve discrimination between true and false shot boundaries [15],[23],[27].Temporal feature combination across consecutive frames can be integrated to handle gradual transitions such as fades and dissolves by sequence-based recognition frameworks mentioned by Tewari et al.[21]. The Euclidean distance computation in MCA-EU can be enhanced through feature-space weighting or metric learning, enabling dynamic emphasis on more discriminative visual feature mentioned by Rajesh, Bharadi and Jangid [27]. Future extensions of MCA-EU aim to link adaptive shot segmentation with Gujarati OCR text feature extraction as mentioned by Bhuva and Mishra [28] for retrieval to regional news broadcast video content.

References

[1] B. T. Truong and S. Venkatesh, "Video abstraction: A systematic review and classification," *ACM Trans. Multimedia Comput. Commun. Appl.,* vol. 3, no. 1, pp. 3 (2007). https://doi.org/10.1145/1198302.1198305

[2] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in Proc. ACM Int. Conf. Multimedia, Singapore, 2005. https://doi.org/10.1145/1180639.1180824

[3] C. B. A. Sai Ram and S. Sharma, "Video segmentation: A systematic review on recent advances, techniques, and classification," in Proc. Int. Conf. Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida, India, pp. 615–620 (2018). https://doi.org/10.1109/icacccn.2018.8748578

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

[5] U. Gargi, R. Kasturi, and S. Strayer, "Performance characterization of video-shot-change detection methods," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 10, no. 1, pp. 1–13 (2000). https://doi.org/10.1109/76.825852

[6] Y. Bendraou, F. Essannouni, D. Aboutajdine, and A. Salam, "Video shot boundary detection method using histogram differences and local image descriptor," in Proc. World Conf. Complex Systems (WCCS), Marrakech, Morocco, pp. 665–670 (2014). https://doi.org/10.1109/icocs.2014.7060883

[7] W. Tong, L. Song, X. Yang, H. Qu, and R. Xie, "CNN-based shot boundary detection and video annotation," in Proc. IEEE Int. Symp. Broadband Multimedia Systems and Broadcasting, Ghent, Belgium, pp. 1–5 (2015). https://doi.org/10.1109/bmsb.2015.7177222

[8] S. N. Kumar, V. S. K. Reddy, and L. K. Balivada, "Content-based video retrieval using deep learning algorithms," in Intelligent Systems and Sustainable Computing (ICISSC 2022), Smart Innovation, Systems and Technologies, vol. 363, Springer, Singapore, (2023) https://doi.org/10.1007/978-981-99-4717-1_52

[9] R. M. Bommisetty, P. Palanisamy, and A. Khare, "Content-based video retrieval—methods, techniques, and applications," in Advances in Soft Computing Techniques in Data Science, IoT, and Cloud Computing, Studies in Big Data, vol. 89, Springer, Cham (2021). https://doi.org/10.1007/978-3-030-75657-4_4

[10] E. M. Saoudi and S. Jai-Andaloussi, "A distributed content-based video retrieval system for large datasets," *J. Big Data,* vol. 8, p. 87 (2021). https://doi.org/10.1186/s40537-021-00479-x

[11] M. J. Esteve Brotons, F. J. Lucendo, R.-J. Javier, and J. Garcia-Rodriguez, "Shot boundary detection with 3D depthwise convolutions and visual attention," *Sensors*, vol. 23, no. 16, p. 7022 (2023). https://doi.org/10.3390/s23167022

[12] T. Soucek and J. Lokoc, "TransNet V2: An Effective Deep Network Architecture for Fast Shot Transition Detection", in Proceedings of the 32nd ACM International Conference on Multimedia, 28 October (2024). https://doi.org/10.1145/3664647.3685517 .

[13] Sadiq Abdulhussain. Abd Ramli, M. Saripan, Basheera Mahmmod, Syed Al-Haddad, Wissam Jassim, "Methods and challenges in shot boundary detection: A review," *Entropy*, vol. 20, pp. 214 (2018). https://doi.org/10.3390/e20040214

[14] B. A. Halim and T. Faiza, "Shot boundary detection: Fundamental concepts and survey," in Proc. Conf. Innovative Trends in Computer Science (2019).

[15] A. Shah and D. Mishra, "A review of biometrics modalities and data mining algorithm," in Ambient Communications and Computer Systems, Adv. Intell. Syst. Comput., vol. 696, Springer (2018). https://doi.org/10.1007/978-981-10-7386-1_51

[16] TV9 Gujarati News Channel, "Gujarati news video dataset," TV9 Gujarati, accessed Oct. 12 (2024).

[17] S. Meng and H. Jiang, "The video shot boundary cut detection based on color histogram," *Journal of Computational and Theoretical Nanoscience*, vol. 13, no. 5, pp (May 2016). https://doi.org/10.1166/jctn.2016.4975

[18] H. Zhang, A. Kankanhalli, and S. W. Smoliar, "Automatic partitioning of full-motion video," *Multimedia Syst.*, vol. 1, pp. 10–28 (1993). https://doi.org/10.1007/bf01210504

[19] Z. Li, X. Liu, and S. Zhang, "Shot boundary detection based on multilevel difference of colour histograms," in Proc. Int. Conf. Multimedia and Image Processing (ICMIP), pp. 15–22 (2016). https://doi.org/10.1109/icmip.2016.24

[20] Shitao Tang, Litong Feng, Zhangkui Kuang, Yimin Chen and Wei Zhang, "Fast Video Shot Transition Localization with Deep Structured Models", *Lecture Notes in Computer Science,Asian Conference on Computer Vision* (2019), https://doi.org/10.1007/978-3-030-20887-5_36

[21] Yogya Tewari, Payal Soni, Shubham Singh, Murali Saketh Turlapati, and Avani Bhuva, "Real time sign language recognition framework for two way communication," in Proc. ICCICT, pp. 1–6 (2021). https://doi.org/10.1109/iccict50803.2021.9510094

[22] A. Adnan and M. Aliaeda, "Shot boundary detection using sorted color histogram polynomial curve," *Life Sci. J.*, vol. 10, pp. 1965–1972 (2013).

[23]A. Shah and V. A. Bharadi, "Performance analysis of soft biometrics features for dynamic signature recognition using complex Walsh plane based feature vector," in Proc. ICCICT (2015). https://doi.org/10.1109/iccict.2015.7045719

[24] M. Verma and B. Raman, "A hierarchical shot boundary detection algorithm using global and local features," in Proc. Int. Conf. Computer Vision and Image Processing (2017). https://doi.org/10.1007/978-981-10-2107-7_35

[25] A. Gushchin, A. Antsiferova, and D. Vatolin, "Shot boundary detection method based on a new extensive dataset and mixed features," Proceedings of the 31th International Conference on Computer Graphics and Vision. Volume 2 (2021). https://doi.org/10.20948/graphicon-2021-3027-188-198

[26] P. K. Sahoo, S. Soltani, and A. K. C. Wong, "A survey of thresholding techniques," *Comput. Vis. Graph. Image Process.*, vol. 41, no. 2, pp. 233–260 (1988). https://doi.org/10.1016/0734-189x(88)90022-9

[27] S. A. Rajesh, V. A. Bharadi and P. Jangid, "Performance improvement of complex plane based feature vector for online signature recognition using soft biometric features," 2015 International Conference on Communication, Information & Computing Technology (ICCICT), Mumbai, India, pp. 1-7 (2015). https://doi.org/10.1109/iccict.2015.7045719

[28] A. Bhuva and D. Mishra, "Gujarati optical character recognition using efficient text feature extraction approaches," *Informatica*, vol. 49 (2025). https://doi.org/10.31449/inf.v49i28.8341

[29] Jasmin T. Jose, S. Rajkumar, Muhammad Rukunuddin Ghalib, Achyut Shankar, Pavika Sharma, and Mohammad R. Khosravi, "Efficient shot boundary detection with multiple visual representations," Mobile Inf. Syst., vol. 2022, Art. no. 4195905 (2022). https://doi.org/10.1155/2022/4195905

[30] N. Ibrahim and Z. Abduljabbar, "Video shot boundary detection based on frames objects comparison and scale-invariant feature transform technique," *Comput. Sci. Inf. Technol.*, vol. 5, pp. 130–139 (2024). https://doi.org/10.11591/csit.v5i2.pp130-139

[31] Z. Yang, L. Tian, and C. Li, "A fast video shot boundary detection employing OTSU's method and dual pauta criterion," in Proc. IEEE Int. Symp. Multimedia, pp. 583–586 (2017). https://doi.org/10.1109/ism.2017.114

[32] Rui Liang, Qingxin Zhu, Honglei Wei, and Shujiao Liao, "A video shot boundary detection approach based on CNN feature," in Proc. IEEE Int. Symp. *Multimedia*, pp. 489–494 (2017). https://doi.org/10.1109/ism.2017.97

[33] S. Dhiman, R. Chawla, and S. Gupta, "A novel video shot boundary detection framework employing DCT and pattern matching," *Multimedia Tools Appl.,* vol. 78, no. 24, pp. 34707–34723 (2019). https://doi.org/10.1007/s11042-019-08170-3

[34] M. Phatak and M. Patwardhan, "Mathematical modelling and statistical analysis in designing deep learning-based shot boundary detection and aesthetic assessment of videos," *J. Interdisciplinary Math.,* vol. 27, no. 2, pp. 369–382 (2024). https://doi.org/10.47974/jim-1857

Received August, 2025