



# Content-Based Video Retrieval System Using Dwt-Gabor Hybrid Approach

<sup>1</sup>D. Asha, <sup>2</sup>Y. Madhatee Latha

<sup>1</sup>Research Scholar (H.T. No.: 1503PH0832), <sup>2</sup>Professor

<sup>1</sup>Department of Electronics and Communication Engineering,

<sup>1</sup> Jawaharlal Nehru Technological University Hyderabad, Hyderabad, Telangana, India

<sup>2</sup>Department of Electronics and Communication Engineering,

<sup>2</sup> Malla Reddy Engineering College for Women affiliated to JNT University, Hyderabad, Telangana, India

**Abstract:** Content-Based Video Retrieval (CBVR) has emerged as an essential research area due to the exponential growth of multimedia data and the need for efficient indexing, searching, and retrieval. Traditional retrieval methods relying on textual annotations are often inadequate, subjective, and time-consuming. To address this, we propose a hybrid CBVR framework that leverages the complementary strengths of Discrete Wavelet Transform (DWT) and Gabor filters for effective feature extraction. The integration of these methods yields a robust feature descriptor capable of discriminating between visually similar video frames. Similarity measurement is performed using distance metrics such as Minkowski distance to ensure accurate retrieval. Experimental analysis on UCF video datasets demonstrates that the DWT-Gabor hybrid approach achieves improved average precision of 85.71%.

**Index Terms** - Content-Based Video Retrieval (CBVR), Discrete Wavelet Transform (DWT), Gabor Filter, Video Shot Boundary Detection (CBVR)

## I. INTRODUCTION

The rapid growth of multimedia data, particularly digital video, has created an urgent need for efficient techniques to organize, search, and retrieve video content. Traditional retrieval systems rely heavily on manual annotation, where textual keywords or tags are associated with video sequences. However, manual labelling is time-consuming, subjective, and often inadequate for capturing the rich and complex semantics of video data. This has led to the emergence of Content-Based Video Retrieval (CBVR), where retrieval is performed using the intrinsic visual and audio content of videos rather than textual metadata.

CBVR systems typically extract and analyse low-level features such as colour, texture, shape, and motion, which are then mapped to higher-level semantic concepts [1]. Among these, motion features play a crucial role in differentiating video content, since video is inherently a temporal sequence of frames [1]. Techniques such as optical flow, motion histograms, and spatio-temporal descriptors have been widely studied for capturing dynamic aspects of video. Complementary to motion, features like colour histograms, Gabor texture descriptors, and wavelet transforms provide additional discriminatory power for robust retrieval. The literature survey on CBVR system is tabulated in Table.1. The paper outline is as follows: in section 2 we discussed about DWT-Gabor feature extraction. In section 3 our proposed CBVR system is illustrated. In section 4 and 5 we presented our results and conclusion respectively.

Table. 1 Literature Survey of CBVR Systems

Paper Title	Year of Publication	Features	Method
SFD: Similar Frame Dataset for Content-Based Video Retrieval [2]	2024	Faster R-CNN) and a Multimodal Large Language Model (BLIP2)	Adjacent Frames Contrastive Learning (AFCL)
Doppel Search: A Novel Approach to Content-Based Video Retrieval for AI Challenge HCMC 2023 [3]	2023	ViT-b/32 model	CLIP (Contrastive Language-Image Pre-training) model
Content based video retrieval using discrete cosine transform [4]	2021	DCT coefficients	Discrete cosine transforms
Content Based Video Retrieval System Using Multiple Features [5]	2018	Color, Texture and motion features	Color distributions, LBP, SAD
Temporal Aggregation of Visual Features for Large-Scale Image-to-Video Retrieval [6]	2018	Temporal encoding tracking binary vectors through time.	kd-tree for faster NN search
Multi-Input-Multi-Output Interface Video Retrieval Method [7]	2019	Color histogram, LBP, Edges	MIMO
Content-Based Video Big Data Retrieval with Extensive Features and Deep Learning [8]	2022	main features, including subtitles, speeches, and objects in video frames,	optical character recognition (OCR), automatic speech recognition (ASR), and object identification with deep learning techniques

## II. DWT-GABOR FEATURE EXTRACTION

The combination of Discrete Wavelet Transform (DWT) [9] and Gabor filters [10] is widely used in image and video analysis for extracting robust texture and frequency features. DWT decomposes a frame into four sub-bands LL, LH, HL and HH. A Gabor filter is a band-pass filter that models the human visual system's sensitivity to orientation and spatial frequency. It captures edge, texture, and directional information similar to human perception. DWT provides multi-resolution analysis and Gabor filter extracts texture details invariant to orientation and scale. The hybrid approach balances global frequency structure and local directional texture, improving retrieval accuracy.

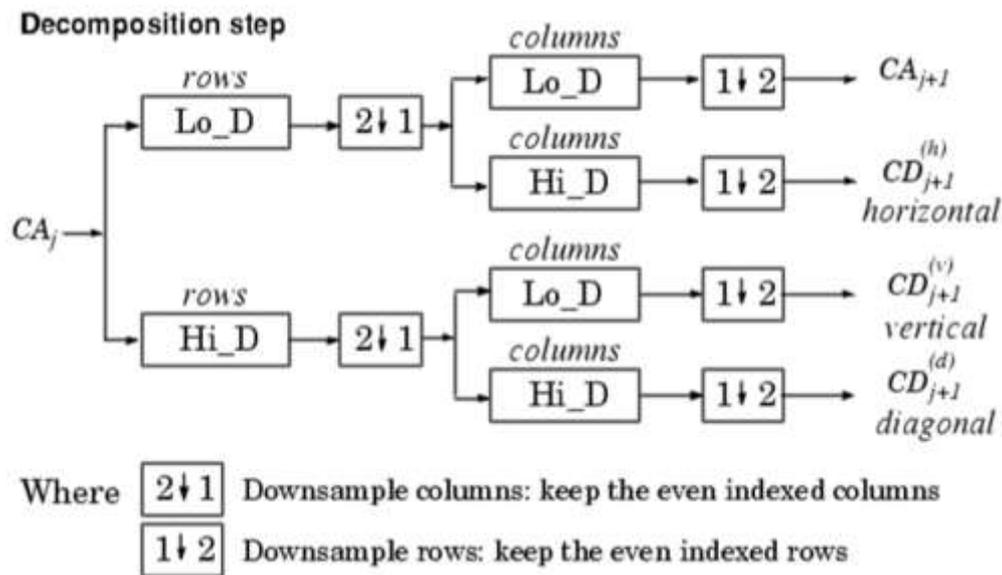


Fig. 1 DWT decomposition steps for an image

The Daubechies db4 wavelets are compactly supported orthogonal wavelets. db4 means it has 4 filter coefficients (length = 4). It provides a balance between time resolution and frequency resolution. It is smoother than the Haar wavelet and captures both edges and texture more effectively. The Algorithm steps for DWT-Gabor feature extraction are given below:

Step 1: Read a video frame and convert RGB to Gray frame.

Step 2: Apply db4 DWT on frame to decompose LL, LH, HL and HH equations (1)-(6).

$$LL(x, y) = \sum_{m=0}^3 \sum_{n=0}^3 h(m)h(n)f(2x - m, 2y - n) \quad (1)$$

$$LH(x, y) = \sum_{m=0}^3 \sum_{n=0}^3 h(m)g(n)f(2x - m, 2y - n) \quad (2)$$

$$HL(x, y) = \sum_{m=0}^3 \sum_{n=0}^3 g(m)h(n)f(2x - m, 2y - n) \quad (3)$$

$$HH(x, y) = \sum_{m=0}^3 \sum_{n=0}^3 g(m)g(n)f(2x - m, 2y - n) \quad (4)$$

Where  $f(x,y)$  is input frame,  $(x,y)$  are spatial coordinates of frame,  $h(m)$ ,  $g(n)$  are db4 filter coefficients and  $m, n=0,1,2,3$

$$h(0) = \frac{1+\sqrt{3}}{4\sqrt{2}}, h(1) = \frac{3+\sqrt{3}}{4\sqrt{2}}, h(2) = \frac{3-\sqrt{3}}{4\sqrt{2}}, h(3) = \frac{1-\sqrt{3}}{4\sqrt{2}} \quad (5)$$

$$g(n) = (-1)^n h(3 - n) \quad (6)$$

Step 3: Define Gabor filter parameter  $\lambda=4$ ,  $\gamma = 0.5$ ,  $\theta=0^0, 45^0, 90^0$ , and  $135^0$ ,  $\varphi=0$  for even symmetric and  $\pi/2$  for odd symmetric. Construct Gabor Kernel using equation (7).

$$Gabor\ Kernel(x, y) = \exp\left(\frac{-x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cdot \cos\left(2\pi \frac{x'}{\lambda} + \varphi\right) \quad (7)$$

$$where\ x' = x\cos\theta + y\sin\theta, y' = -x\sin\theta + y\cos\theta$$

$\lambda$  is wavelength,  $\gamma$  is Spatial aspect ratio,  $\theta$  is orientation,  $\varphi$  is phase of fset,

$\sigma$  is standard deviation of Gaussian Envelope

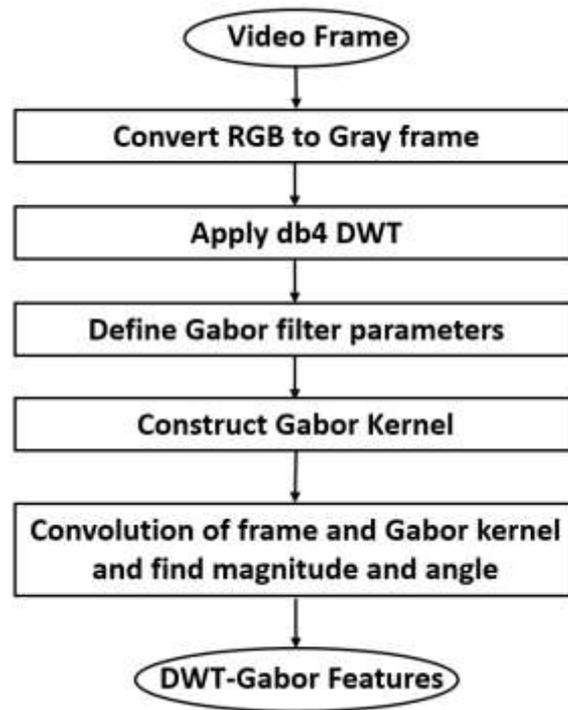


Fig. 2 Flowchart for DWT-Gabor Feature Extraction

### III. PROPOSED CBVR SYSTEM

In the proposed Content-Based Video Retrieval (CBVR) framework as shown in Fig.3, we integrate Discrete Wavelet Transform (DWT) and Gabor filtering to capture both frequency- and texture-oriented information from video frames. System workflow starts with Video Shot Boundary Detection (VSBD) to remove redundancy between frames [11]. Next identify key frames from each shot we considered middle frame from each shot and then DWT- Gabor features are extracted for different orientations. Lastly similarity is measured between feature vectors to retrieve the similar videos. The algorithm steps for VSBD are given below:

Step 1: read video frames sequentially and extract DWT-Gabor features as shown in fig. 2.

Step 2: measure the similarity between feature vectors of sequential frames and construct a continuous signal.

Step 3: detect the least 5 valley points and take the average of them to calculate threshold value.

Step 4: the values below threshold are the shot transition points.

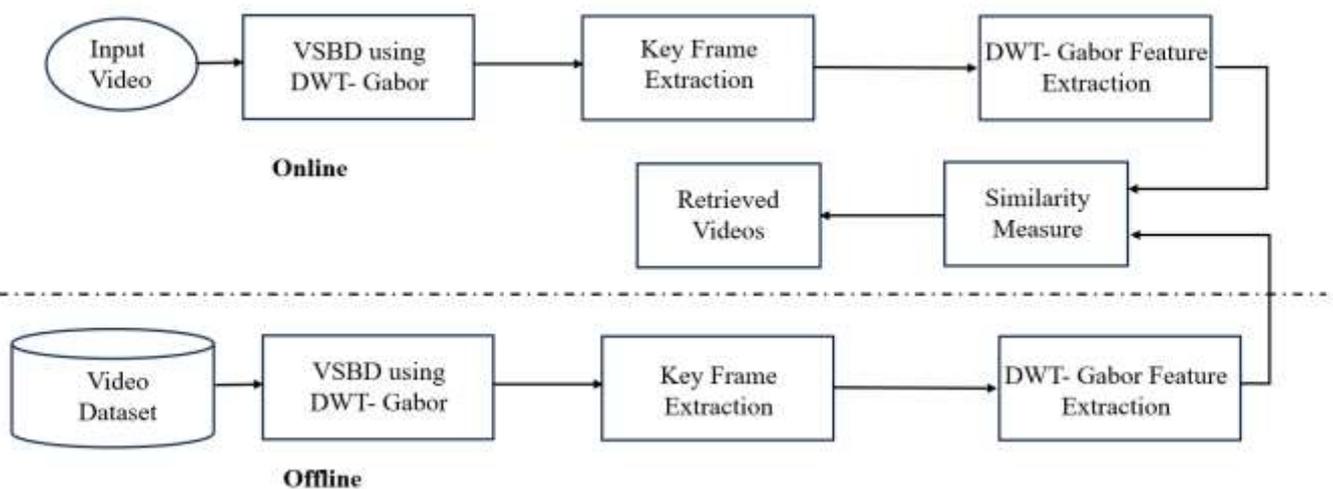


Fig. 3 Block Diagram of proposed CBVR System

In offline process the VSBD technique is applied for shot detection on all videos of database and then key frames are extracted from each video, we choose the middle frame of a shot as key frame. Next DWT-Gabor feature vectors are computed for key frames and stored as database feature vector (DBFV). In online

process the above procedure is repeated for query video and query feature vector(QFV) is computed. The similarity is measured between QFV and DBFV using Minkowski distance as in (8), we considered  $p=4$ .

$$MD(i) = \left( \sum_{j=1}^K |DBFV_{ij} - QFV_j|^p \right)^{1/p} \quad (8)$$

where  $K$  is No. of videos in dataset

#### IV. RESULTS AND DISCUSSIONS

The experiments are simulated using MATLAB 2022b version on an intel core i7 CPU with 16GB RAM and 500 GB disk space. The CBVR system is tested on UCF dataset [12], we evaluated system by computing precision, recall rate and F1 score defined in equation (9)-(11). We considered 15 classes of videos each consisting of 15 relevant videos. From 225 videos 475 key frames are extracted. For an example video stage wise outputs are discussed. The DWT-Gabor features are shown in Fig. 4, VSBD Plot is shown in Fig. 5. The computed threshold value is 0.9941, it will divide the video into 2 shots [1 to 166] and [167 to 176]. From each shot middle key frame is extracted as shown in Fig. 6. On key frames we apply Gabor kernel for 4 orientations  $[0^\circ 45^\circ 90^\circ 135^\circ]$  and magnitude and phase for different orientations is shown in Fig. 7. The retrieval results for few example videos are shown in Table. 2. The average precision, recall rate and F1 score for proposed CBVR system are compared with other CBVR systems are shown in Table.3 and graphs are shown in Fig. 8.

$$Precision(P) = \frac{TP}{TP+FP} * 100 \quad (9)$$

$$Recall\ rate\ (R) = \frac{TP}{TP+FN} * 100 \quad (10)$$

$$F1\ Score(\%) = \frac{2*P*R}{P+R} \quad (11)$$

Where TP- True Positive, FP- False Negative and FN-False Negative.

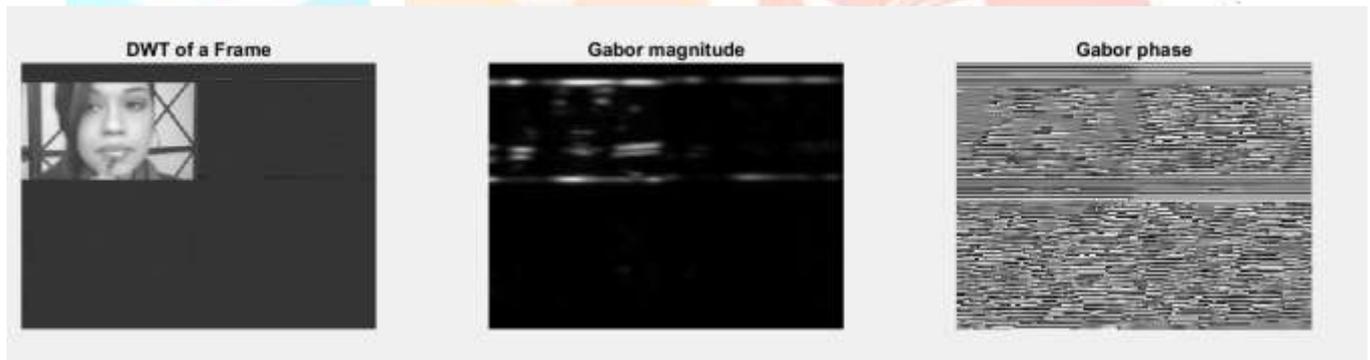


Fig. 4 DWT-Gabor Feature Extraction

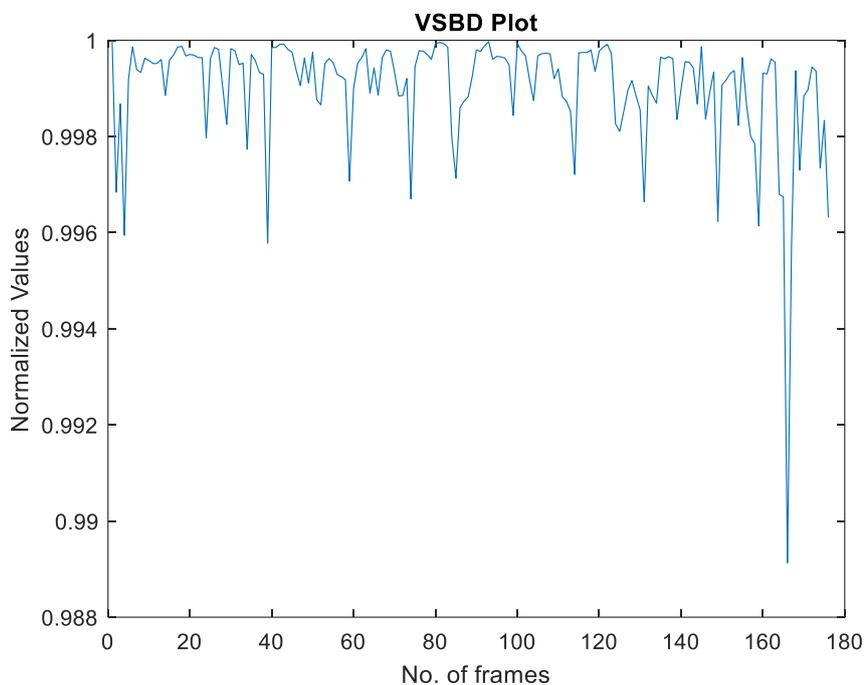


Fig. 5 VSBD Plot

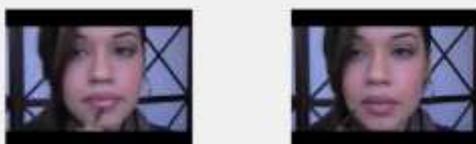


Fig. 6 Extracted Key Frames

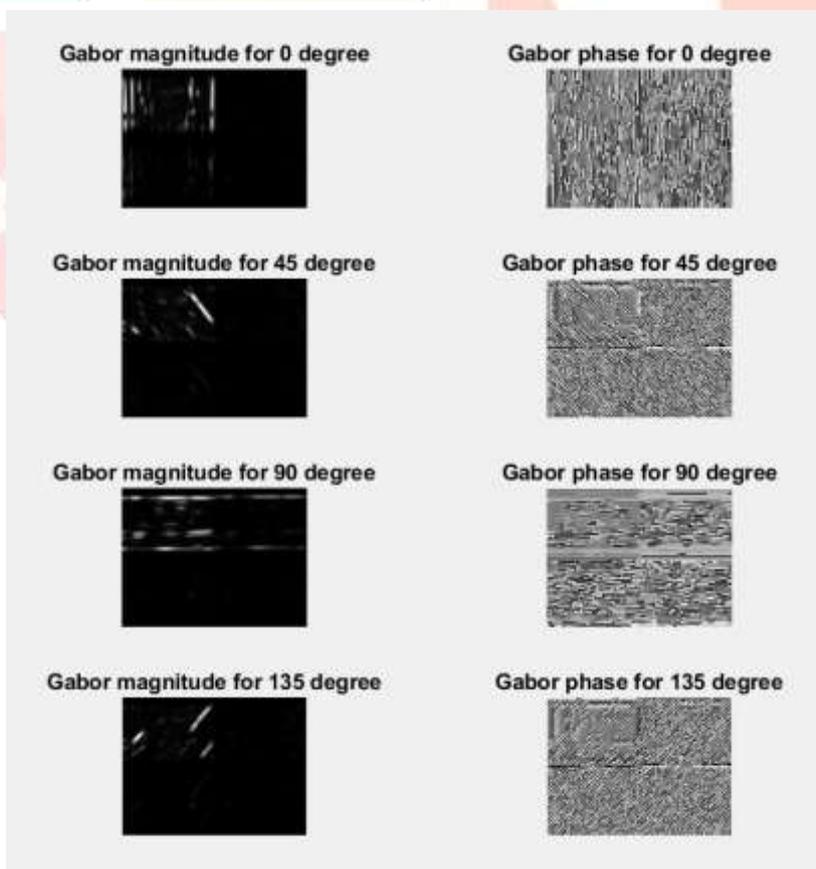


Fig. 7 Gabor Magnitude and phase for  $[0^0 45^0 90^0 135^0]$  Orientations

Table. 2 Retrieval results for few examples videos

Query Video	Retrieved top 10 videos	P (%)	R (%)	F1 (%)
		100	86.66	92.85
		70	66.66	68.28
		60	60	60
		90	86.66	88.29
		100	93.33	96.54
		80	73.33	76.51
		100	93.33	96.54
Average		85.71	79.99	82.71

Table. 3 Comparison of proposed CBVR system with other works

CBVR Systems	Average Precision (%)
Proposed CBVR system	85.71
CBVR using multiple features [5]	80
CBVR using discrete cosine transform [4]	64.75

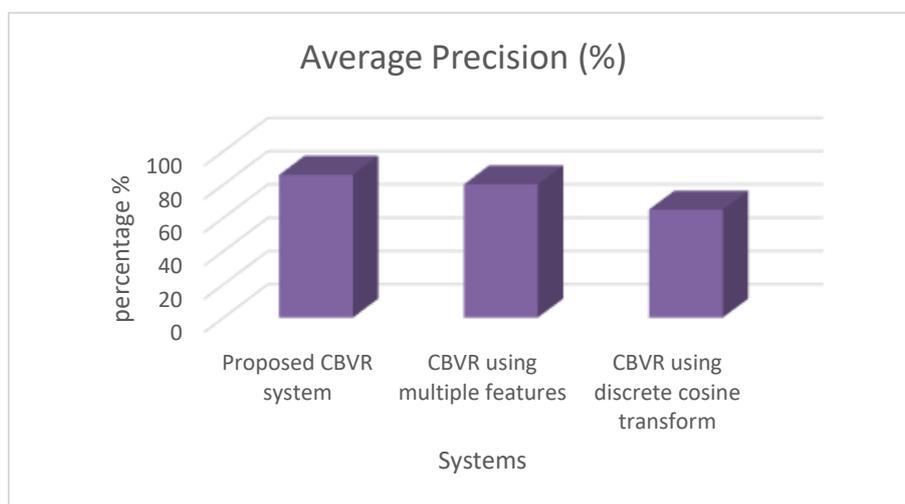


Fig. 8 Comparison graph of different CBVR systems

## V. CONCLUSION

This work presented a hybrid Content-Based Video Retrieval (CBVR) framework that integrates Discrete Wavelet Transform (DWT) and Gabor filters to exploit both multi-resolution and directional texture features. By employing Minkowski distance for similarity measurement, the system effectively discriminates between visually similar video frames. Experimental evaluation on the UCF dataset demonstrates that the proposed DWT–Gabor approach achieves an average precision of 85.71%. Future research may extend this work by incorporating motion-based features.

## REFERENCES

- [1] A Sedky Adly, M. S. Abdelwahab, I. Hegazy and T. Elarif, "Issues and Challenges for Content-Based Video Search Engines A Survey," 2020 21st International Arab Conference on Information Technology (ACIT), Giza, Egypt, 2020, pp. 1-18, doi: 10.1109/ACIT50332.2020.9300062.
- [2] C. Han, G. Meng and C. Huo, "SFD: Similar Frame Dataset for Content-Based Video Retrieval," 2024 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 2024, pp. 2403-2409, doi: 10.1109/ICIP51287.2024.10647446.
- [3] Phong Phan Nguyen Huu, Khoa Tran Dinh, Ngan Tran Kim Ngoc, Luong Tran Duc, Phuc Le Hoang, Quyen Nguyen Huu, Duy Phan The, and Van-Hau Pham. 2023. DoppelSearch: A Novel Approach to Content-Based Video Retrieval for AI Challenge HCCM 2023. In Proceedings of the 12th International Symposium on Information and Communication Technology (SOICT '23). Association for Computing Machinery, New York, NY, USA, 916–922. <https://doi.org/10.1145/3628797.3628950>.
- [4] Farhan, Ahmeed Suliman. "Content based video retrieval using discrete cosine transform", Indonesian Journal of Electrical Engineering and Computer Science, Vol. 21, No. 2, February 2021, pp. 839-845, ISSN: 2502-4752, DOI:10.11591/ijeecs.v21.i2.pp839-845.
- [5] D. Asha, Y. Madhavee Latha, V. S. K. Reddy, "Content Based Video Retrieval system using Multiple Features", International Journal of Pure and Applied Mathematics (IJPAM), Vol. 118, No.14, February 2018, pp 287-294, <https://acadpubl.eu/jsi/2018-118-14-15/articles/14/42.pdf>.
- [6] Noa Garcia. 2018. Temporal Aggregation of Visual Features for Large-Scale Image-to-Video Retrieval. In Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval (ICMR '18). Association for Computing Machinery, New York, NY, USA, 489–492. <https://doi.org/10.1145/3206025.3206083>
- [7] S. Mitsui and S. Wada, "Multi-Input-Multi-Output Interface Video Retrieval Method," 2019 34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), JeJu, Korea (South), 2019, pp. 1-4, doi: 10.1109/ITC-CSCC.2019.8793368.
- [8] Phan, T. C., Phan, A. C., Cao, H. P., & Trieu, T. N. (2022). Content-Based Video Big Data Retrieval with Extensive Features and Deep Learning. Applied Sciences (Switzerland), 12(13). <https://doi.org/10.3390/app12136753>.
- [9] Refael C. Gonzalez, Richard E. Woods, Digital Image processing, Second edition, Pearson.
- [10] D. Srivastava, S. S. Singh, B. Rajitha, M. Verma, M. Kaur and H. -N. Lee, "Content-Based Image Retrieval: A Survey on Local and Global Features Selection, Extraction, Representation, and Evaluation Parameters," in IEEE Access, vol. 11, pp. 95410-95431, 2023, doi: 10.1109/ACCESS.2023.3308911.
- [11] D. Asha, Y. Madhavee Latha, "CONTENT-BASED VIDEO SHOT BOUNDARY DETECTION USING CANNY EDGE FEATURES ", International Journal of Emerging Technologies and Innovative Research (www.jetir.org | UGC and ISSN Approved), ISSN:2349-5162, Vol.12, Issue 8, page no. ppd120-d126, August-2025, Available at : <http://www.jetir.org/papers/JETIR2508315.pdf>.
- [12] K.Soomro, A.R.Zamir and M.Shah, "UCF101: A dataset of 101 human actions classes from videos in the wild," CoRR, Dec.2012, <https://www.crcv.ucf.edu/datasets/human-actions/ucf101/UCF101.rar>.