Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

An Optimized Key Frame Extraction Technique for Content-Based Video Retrieval Using Hybrid PSO-BOA

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

[1]Research Scholar, Department of Electronics and Communication Engineering, Jawaharlal Nehru Technological University Hyderabad, Telangana, India.
[1]Associate Professor, Department of Electronics and Communication Engineering, Malla Reddy College of Engineering and Technology, Hyderabad, Telangana, India.
[2]Professor, Department of Electronics and Communication Engineering, Malla Reddy Engineering College for Women, affiliated to JNT University, Hyderabad, Telangana, India.

E-mail: [1]ashamanickrao@gmail.com, [2]madhaveelatha2009@gmail.com

Abstract

Extraction of key frames is an essential and significant stage in any video analysis application, aimed at describing the video content precisely by removing the redundancy. An optimized key frame extraction technique based on the Binary Butterfly Optimization Algorithm (BBOA) is introduced. Here, the extraction of Color Coherent Vector (CCV) features from each frame of the video and the application of the BBOA algorithm for minimising redundancy and maximising diversity between the frames is carried out. The process is applied iteratively until the precise number of frames is selected. Further, the system is extended by proposing a Content-Based Video Retrieval (CBVR) system using the selected key frames, extracting multiple features from Grey Level Run Length Matrix (GLRLM) texture features and Dual Tree- Complex Wavelet Transform (DT-CWT) shape descriptors along with CCV features. Due to the multiple features, the feature vector size is huge, so to reduce its dimension, a hybrid Binary Particle Swarm Optimization-Butterfly Optimization Algorithm (PSO-BOA) feature selection method is applied. The experiment was conducted on the UCF101 dataset, and our proposed system outperformed with a compression rate, precision, recall rate, F1 score, and FRR of 0.985, 0.913, 0.78, 0.836, and 0.965, respectively, demonstrating the effectiveness of using the hybrid Optimization algorithm in improving the efficiency of the CBVR system.

Keywords: Content-Based Video Retrieval (CBVR), Binary Butterfly Optimization Algorithm (BBOA), Particle Swarm Optimization-Butterfly Optimization Algorithm (PSO-BOA), Colour Coherence Vector (CCV), Grey Level Run Length Matrix (GLRLM), Dual Tree- Complex Wavelet Transform (DT-CWT).

1. Introduction

The rapid use of multimedia on web platforms has created a huge demand for video analysis techniques like video classification, video surveillance systems, video indexing, and retrieval systems, making it a significant research area with quite challenging aspects. Content-

DOI: https://doi.org/10.36548/jiip.2025.4.017
Received: 13.10.2025, received in revised form: 11.11.2025, accepted: 25.11.2025, published: 05.12.2025

based retrieval methods have gained importance, as they help improve accuracy and efficient retrieval according to the given user query, rather than relying on the metadata of the video file, which includes human errors and manipulations.

Key frame extraction is the initial stage in all video applications to reduce redundancy in video, as it occupies a large storage space, and is difficult to handle while computing further. Traditionally, video shot boundary detection was used to divide the video into shots and then extract key frames. Detecting shot transitions, which are of various kinds, it was difficult to handle. From a shot, either the middle or first or last frame was chosen as the key frame. To overcome these challenges, a proposed key frame extraction technique using an Optimization algorithm ensures that selected optimum frames are diverse and non-redundant from other frames, thereby summarising the video.

Here, an optimized key frame extraction method is proposed by first extracting Color Coherence Vector (CCV) features from each frame of the video and applying an Optimization algorithm to assess diversity and redundancy between the sequential frames, this is referred to as frame selection, inspired by Hyperspectral band selection methods. This process is iteratively applied until the required number of key frames is extracted from the videos. For key frame extraction, the focus is on a single feature to reduce computational cost. However, for the extracted key frames, further additional features are added, such as GLRLM texture descriptors and DT-CWT shape descriptors. Then, a metaheuristic nature-inspired Optimization algorithm for high-dimensionality reduction is applied, which results in reduced storage space for feature vectors. A hybrid PSO-BOA algorithm is proposed for feature selection. This hybrid model combines the global convergence of BOA with the fast local search of PSO to select features that maximise information diversity and minimise redundancy. PSO explores the solution space through social interaction and velocity updates, while BOA balances exploration and exploitation using sensory fragrance-based search.

Section 2 presents a literature review, followed by an overview of the proposed optimized key frame extraction in Section 3. Section 4 details the CVBR system. Sections 5 and 6 discuss performance evaluation results, conclusions, and future research directions.

2. Literature Review

Various techniques exist for extracting, representing, and retrieving visual data in content-based video retrieval. Our focus has been on literature related to keyframe extraction and feature selection. In this context, hyperspectral band selection methods are often used to reduce dimensionality. While hyperspectral images and videos share similar dimensional structures, hyperspectral images capture data at different wavelengths, whereas videos capture images at different times.

Rajeshwari et al. [1] introduce a spatio-temporal keyframe extraction method using HOG-SVM, VGG-16, and inter-frame difference techniques to efficiently identify human-detected keyframes in CCTV footage. The approach reduces the time and space required to analyse large volumes of surveillance video. Zhong et al. [2] present a keyframe extraction algorithm that combines multi-feature fusion and deep prior information, demonstrating effectiveness for motion videos, especially sports videos that are prone to quality interference. Mallick et al. [3] use motion vector-based keyframe extraction for video summarisation and apply spatial pyramid matching for key frame comparison. Keyframes are divided into finer sub-regions for feature computation and tested on the UCF and VCD datasets. Hussain et al.

[4] propose an efficient content-based video retrieval system that utilises the AlexNet CNN model on keyframes to enhance retrieval from large video collections. Color histogram and k-means clustering techniques are used for selecting keyframe. Kar et al. [5] cover various methods for shot transitions and key frames identification with different evaluation metrics. Thomas et al. [6] present a video summarization method of extracting one keyframe using human perceptual and optimization. Farhan et al. [7] illustrate a DCT transform based method for feature extraction.

Arora et al. [8] introduce the Butterfly Optimization Algorithm, which mimics the food search and mating behavior of butterflies, solving global Optimization problems. Cher et al. [9] present a method that provides a decomposition-based multiobjective Optimization for hyperspectral image feature selection to overcome the inefficient search and inadequate local Optimization in high-dimensional, multi-peak search spaces. Kaya et al. [10] provide a study that develops and introduces a binary Black Widow Optimization algorithm for feature selection, indicating its efficiency in handling complex Optimization problems. Shami et al. [11] discuss PSO in a detailed study and the conversion of continuous PSO variants to binary forms.

Su et al. [12] present a PSO-based system for selecting the optimal number of bands for hyperspectral dimensionality reduction. This approach is applied to improve classification accuracy. Murinto et al. [13] present a new technique for reducing the dimensionality of hyperspectral images using independent component analysis optimized by PSO. Mengjian et al. [14] introduce hybrid Optimization algorithms by combining the metrics of each algorithm to overcome its weaknesses. They present a novel hybrid algorithm called HPSOBOA, which combines the BOA with PSO. Ghaleb et al. [15] highlight the importance of feature selection in improving classification performance and demonstrate the multi-objective grasshopper Optimization algorithm.

The literature survey was carried out on different hyperspectral image band selection methods using various nature-inspired metaheuristic global optimization algorithms. Since video has a similar structure to hyperspectral images, the same methods are applicable here. The survey reveals that for keyframe extraction, deep learning concepts are used, which will increase efficiency at the cost of a complex system and computational expense. Shot detection using different features has been studied. Instead of shot boundary detection, traditional methods require a procedure-based or rule-based technique to detect shot transitions and then extract keyframes from shots, involving many parameters for detection. Thus, an optimization keyframe extraction method is proposed, where an iterative optimization algorithm is applied to select the desired number of frames.

The literature survey highlights the challenges associated with applying Optimization algorithms to reduce frame redundancy and with relying on a single feature for initial keyframe extraction, thereby decreasing computational demands. Implementing a hybrid Optimization approach that incorporates a multi-objective fitness function is anticipated to enhance overall performance.

3. Proposed CBVR System

The proposed CBVR system using Hybrid PSO-BOA is illustrated in Figure 1, where in the offline process, the proposed method is applied to database videos and the feature vectors are extracted. In the online process, the process is applied to a selected query video and the similarity between the database and query feature vectors is measured. Frame redundancy is removed using a BBOA optimization technique for key frame selection. Next, features are extracted to improve the accuracy of retrieval. CCV, GLRLM, and DT-CWT multiple features are extracted. Due to the multiple features, the feature vector storage increases; therefore, a hybrid PSO-BOA Optimization algorithm using a multiobjective fitness function for selecting features is computed to maximize accuracy and minimize the number of selected features. Lastly, similarity is measured using Chi-square distance, and videos are retrieved.

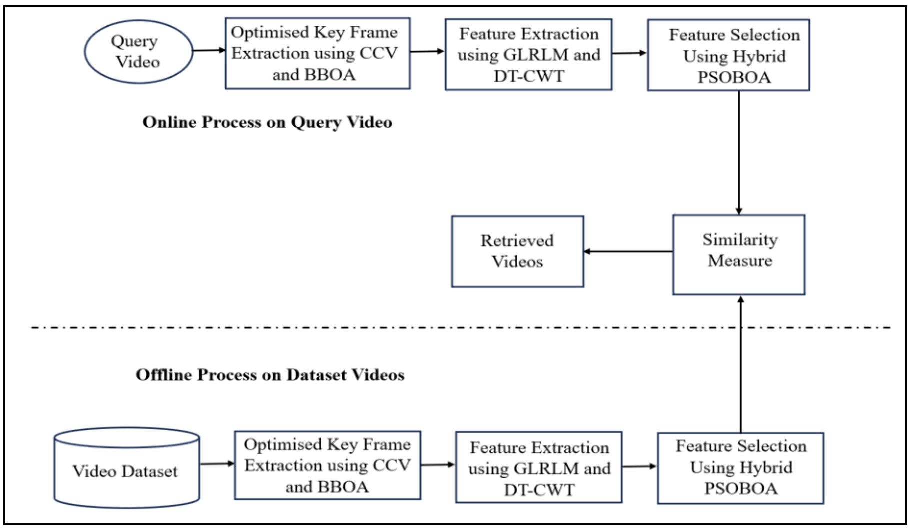


Figure 1. Proposed CBVR System Framework

3.1 Optimized Key Frame Extraction Technique

By analysing the video's content, keyframe extraction generates a concise and meaningful summary of the video. For representation, CCV features are used, and the BBOA algorithm is applied for frame selection. The proposed keyframe extraction algorithm, along with its flowchart, is shown in Figure 2.

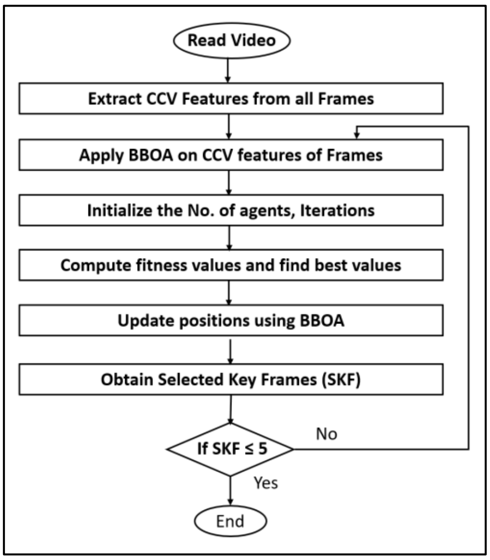


Figure 2. Proposed Key Frame Extraction Flowchart

3.1.1 CCV

The CCV splits the pixels of each color into coherent and incoherent pixels [16]. The coherent pixels belong to a large, connected region of similar color, while incoherent pixels are isolated pixels or small, scattered regions. The CCV captures color information, color distribution, and spatial structure, making it better than other color descriptors like histograms, color moments, and HVS histograms. It also preserves spatial coherence, and is stable under illumination changes. CCV is formulated as shown in (1).

$$CCV(Cp) = \left(\alpha_{Cp}, \beta_{Cp}\right) \tag{1}$$

where α_{Cp} is No. of Coherent pixels of Color Cp, $\beta_{Cp}$is No. of Incoherent pixels of Color Cp. The CCV starts with colour quantisation to reduce colors, and then connected component analysis is performed to identify the connected regions of the same colour coherence. The quantization level of 256 is found optimal for CCV as videos are in RGB format with 8 bits per pixel, therefore 2^8=256, and a threshold is set to determine the region size and categorise the pixels into coherent and incoherent regions.

3.1.2 BBOA

The Butterfly Optimization Algorithm (BOA) is an efficient metaheuristic inspired by butterflies' foraging and mating behaviors. BOA models how butterflies use their sense of smell to locate nectar or mates, guiding the search for optimal solutions. Each butterfly emits a fragrance based on its fitness, attracting others in a process known as global exploration. When no fragrance is detected, butterflies move randomly to explore their surroundings, referred to as local exploration. The switch probability (p) (ranging from 0.6 to 0.8) determines whether global or local exploration is applied. To address the optimization search (global and local) problem, the BBOA parameters are set according to the established literature.

The mathematically expressed as in (2-5).

$$f_i = C.I^a \tag{2}$$

$$C_{t+1} = C_t + \left[\frac{0.025}{(C_t.T_{max})}\right] \tag{3}$$

$$where\ f_i\ is\ fragrance\ of\ i^{th} butterfly, C\ is\ sensory\ modality\ (0.01\ to\ 0.1),$$

$$I\ \ is\ stimulus\ intensity, a\ is\ power\ exponent\ (0.1\ to\ 0.5)$$
$$intially\ C_0 = 0.01, a = 0.1, T_{max}\ is\ Maximum\ No. of\ iteration$$

$$Global\ Exploration: x_i^{t+1} = x_i^t + (r^2.g^* - x_i^t) * f_i \tag{4}$$

$$Local\ Exploration: x_i^{t+1} = x_i^t + \left(r^2.x_j^t - x_k^t\right) * f_i \tag{5}$$

$$where\ x_i^t\ is\ the\ solution\ vector\ for\ i^{th}\ butterfly\ and\ t^{th}\ iteration,$$

$$g^*\ is\ current\ best\ solution, r\ is\ random\ number\ in\ [0,1]$$

$$if\ r > p\ then\ Global\ Exploration$$

$$else\ local\ exploration$$

$$where\ r\ is\ ramdom\ number$$

Proposed Key Frame Extraction Algorithm

Step 1: Read the frames and duration of the video.

Step 2: Extract CCV features

Step 3: Apply the BBOA Optimization Algorithm

Inputs: No. of agents=40, Maximum Iterations=100, a=0.1, p=0.6 and c=0.01. The parameters are set for stabilising the Optimization problem based on the literature we studied.

Step 4: Compute fitness using the fitness function using diversity and redundancy as in (6)

$$FitF = \alpha * \frac{2}{N_F(N_F-1)}\left(\sum_{i=0}^{N_F-1}\sum_{j=i+1}^{N_F}\sqrt{\sum_{k=1}^{P}\left(Fv_{ik}-Fv_{jk}\right)^2} - \beta * \left|\frac{\sum_{i=0}^{P}(Fv_{ik}-\overline{Fv_i})(Fv_{jk}-\overline{Fv_j})}{\sqrt{\sum_{i=0}^{P}(Fv_{ik}-\overline{Fv_i})^2}.\sqrt{\sum_{i=0}^{P}\left(Fv_{jk}-\overline{Fv_j}\right)^2}}\right|\right) \tag{6}$$

$$where\ N_F\ is\ No.of\ Frames, Fv\ is\ the\ feature\ vector, P\ is\ size\ of\ feature\ vector$$

$$\alpha,\beta\ are\ weights, the\ range\ of\ \alpha\ is\ (0.5\ to\ 1.0) and\ \beta\ is\ (0.0\ to\ 0.5)$$

The Euclidean metric is used for diversity and correlation for redundancy, and the weights α=1, β=0.3. The high α encourages more diverse frames and the low β avoids selecting similar frames.

Step 5: Find the best fitness values.

Step 6: Update positions using BBOA and obtain Selected Key Frames

Step 7: If No. of Key frames >= 5 then repeat from Step 3, else end.

The threshold for iteration is chosen based on the ratio of the average number of frames per video to the number of frames per second. The complexity of BBOA-based keyframe extraction increases with the increase in the number of iterations and the number of agents where 100 iterations and 40 agents are considered. As the number of frames increases the runtime increases and the number of loops will increase until the number of selected key frames is less than or equal to 5. The fitness function used for BBOA is a combination of Euclidean distance and correlation for diversity and redundancy, respectively. The Euclidean distance is used to find the absolute difference for checking diversity, and for redundancy.

3.2 Feature Extraction and Selection

Feature extraction and selection are two essential steps in any machine learning or Artificial Intelligence systems to transform the raw visual data into a precise, and meaningful representation of the data that enables accurate outcomes. For feature extraction, CCV, GLRLM and DT-CWT features were considered. They are computed on optimized keyframes. A hybrid Optimization algorithm is used for feature selection.

3.2.1 GLRLM

GLRLM is a statistical texture descriptor that quantifies the length of consecutive, collinear pixels with the same gray level in a specific direction [17]. The matrix rows represent gray levels, columns indicate run lengths, and each entry shows how often a run of a particular gray level occurs in the chosen direction. The following features are calculated from the GLRLM matrix:

Short-run Emphasis (SRE): Measures the presence of fine textures as in (7).

$$SRE = \frac{\sum_{g,r} \frac{GLRLM(g,r)}{r^2}}{Total\ Runs} \tag{7}$$

$$where\ \ GLRLM(g,r)\ is\ g^{th}\ row\ and\ r^{th}\ Column\ element\ in\ GLRLM\ matrix$$

Long-run Emphasis (LRE): Measures the presence of coarse textures as in (8).

$$LRE = \frac{\sum_{g,r} GLRLM(g,r).r^2}{Total\ Runs} \tag{8}$$

Grey Level Non-Uniformity (GLN): Assesses the uniformity of grey levels across runs as in (9).

$$GLN = \frac{1}{Total\ Runs} \sum_{g=0}^{L-1} \left(\sum_{r=1}^{R} GLRLM(g,r)\right)^2 \tag{9}$$

$$where\ L\ is\ Gray\ levels, R\ is\ No. of\ Run\ length$$

Run-Length Non-Uniformity (RLN): Evaluates the uniformity of run lengths across grey levels as in (10).

$$RLN = \frac{\sum_r \left(\sum_g GLRLM(g,r)\right)^2}{Total\ Runs} \tag{10}$$

Run Percentage (RP): Represents the proportion of runs relative to the image size as in (11).

$$RP = \frac{Total\ Runs}{Total\ Pixels} \tag{11}$$

3.2.2 DT-CWT

The DT-CWT is an improved version of the DWT, providing better directional selectivity and shift invariance [18]. It uses two parallel DWT filter trees, a real part tree and an imaginary part tree. It offers a multi-resolution transform-based feature extraction technique that captures six oriented sub-bands per level (-15, +15, -45, +45, -75, +75) with different directions that are suitable for extracting motion changes from each video along with edge features [19]. DT-CWT features are extracted from each key frame of a video, so changes over time in the video are accounted for. The magnitude and angle features of the DT-CWT are useful for motion and shape analysis. The DT-CWT features are defined as in (12-15) [19].

$$W_{j,\theta}(x,y) = f(x,y) * \varphi_{j,\theta}(x,y) \tag{12}$$

$$\varphi_{j,\theta}(x,y) = 2^{-j}\varphi_{\theta}\left(\frac{x}{2j},\frac{y}{2j}\right) \tag{13}$$

$$where\ for\ each\ scale\ j\ the\ 2d\ DT - CWT\ yeilds\ six\ directianal\ subbands\ \varphi_j^{(\pm 15^0)}, \varphi_j^{(\pm 45^0)}, \varphi_j^{(\pm 75^0)}.$$

$$\left|W_{j,\theta}(x,y)\right| = \sqrt{Re\left(W_{j,\theta}\right)^2 + Im\left(W_{j,\theta}\right)^2} \tag{14}$$

$$arg_{j,\theta}(x,y) = tan^{-1}\left(\frac{Im(W_{j,\theta})}{Re(W_{j,\theta})}\right) \tag{15}$$

3.2.3 Hybrid Binary PSO-BOA for Feature Selection

The hybridization will take advantage of both algorithms and improve efficiency, giving rise to a new optimization algorithm. Combining PSO with BOA achieves a balanced search and improves accuracy. PSO excels at global search but lacks local search [10], while BOA effectively identifies local solutions [6], Therefore, these two algorithms form a good pair in evolving a hybrid approach [12]. The Binary PSO (BPSO) is specifically designed for discrete search spaces, addressing binary problems where each particle dimension is either 0 or 1. It applies a transfer function, usually sigmoidal, to map continuous velocity values to probabilities for bit flipping. Error and Feature Reduction Ratio (FRR) are considered in the fitness function, where more importance is given to Error than to FRR by multiplying scaling values of 0.9 and 0.1, respectively. The Hybrid Binary PSO-BOA algorithm is shown below.

Hybrid Binary PSO-BOA Algorithm

Step 1: Inputs: population=feature vectors, parameters c_1=c_2=0.5, r, r_1 ,r_2 are random numbers in [0,1] range, a=0.01, p=0.6, initial c=0.01, max iteration= 50 or 100, training=20% or 40%.

Step 2: Compute the fitness [10] of each butterfly as in (16).

$$Fitness\ function = \gamma * \left(1 - \left[\frac{Correctly\ predicted}{Total\ Valid}\right]\right) + \delta * \frac{|Selected\ Features|}{|Total\ No.\ of\ Features|} \tag{16}$$

$$where\ \gamma\ and\ \delta\ are\ constants\ equal\ to\ 0.9\ and\ 0.1\ respectively.$$

$$where\ error\ is\ given\ more\ importance\ than\ feature\ selection.$$

Step 3: Loop for t=1

Update f_i and find the best f, set r

Step 4: If r<p then apply eq. (4)

Else apply eq. (5)

Step 5: Update velocity as in (17)

$$V_i^{t+1} = w.V_i^t + c_1.r_1 * (P_{best} - x_i^t) + c_2.r_2 * (G_{best} - x_i^t) \tag{17}$$

Step 6: Calculate new fitness values of each butterfly

Step 7: If f_{new} <best f the update as in (18)

$$x_i^{t+1} = x_i^t + V^{t+1} \tag{18}$$

Step 8: Increment t and repeat the loop till t reaches max iterations.

3.3 Similarity Measure

The similarity between query and dataset videos is measured using chi-square distance as in (19). There are many different measures to find similarity, but the chi-square distance is sensitive to relative distance, which makes it a better measure. The smaller the distance, the greater the similarity.

$$ChS(j) = \sum_{j=1}^{K} \frac{1}{2} \sum_{i=1}^{N} \frac{\left(Q_i - DB_{ij}\right)^2}{(Q_i + DB_{ij} + \epsilon)} \tag{19}$$

$where\ Q\ is\ query\ features\ and\ DB\ is\ database\ features,$

$N\ is\ No.of\ features\ and\ K\ is\ No.of\ Videos\ in\ database, \epsilon\ is\ constant$ $=1*10^{-10}$

4. Results and Discussions

This section presents a detailed analysis of the experimental results from the proposed CBVR system simulated on MATLAB 2021a version installed on an 11th GEN Intel® core™ i5 @ 2.50 GHz with 8.0 GB RAM. The system is evaluated on 300 videos from the UCF-101 database, it is an action recognition benchmark database [20]. The description of the video database is shown in Table 1, which includes 30 classes with 10 videos each. The number of video classes is considered to provide a manageable subset and to reduce computational time. The number of video classes is chosen so that the diversity of different human action videos still remains within the subset. The evaluation metrics considered for validation are Compression Ratio (CR), Feature Reduction Rate (FRR), Precision (P), Recall Rate (R), and F1 score as in (20-24).

$$CR = \left(1 - \frac{No.of\ Optimized\ Key\ frames}{Total\ No.of\ frames}\right) \tag{20}$$

$$FRR = \left(1 - \frac{Selected\ features}{Total\ features}\right) \tag{21}$$

$$P = \frac{C}{TR}\ where\ C - correct\ videos, RV - Retrieved\ video \tag{22}$$

$$R = \frac{C}{VC}\ where\ C - correct\ videos, VC - Videos\ in\ Class \tag{23}$$

$$F1\ Score = \frac{2PR}{(P+R)} \tag{24}$$

Table 1. Video Descriptions

Properties	Description
No. of Video Classes	30
No. of Videos	300
Bits per pixel	24
No. of Frames	58494
Frame rate	25 or 30
No. of Optimized Key Frames	825
Compression Ratio	0.9858
Execution time for key frame extraction	52127.3763 Seconds
Feature Vector Size	825 x 58121
Execution time for Feature Extraction	3673.730746 Seconds
Optimized Feature vector size	825 x 1697
Feature Reduction Ratio (FRR)	0.96563
Execution time for feature selection	2430.3808 Seconds
Average Convergence Curve	0.9193

The optimized key frame extraction for an example video is shown in Figure 3. The ‘105.avi’ video has a total of 112 frames. When the BBOA algorithm was applied, it looped 5 times, and in each iteration, the number of frames selected was 64, 38, 22, 11, and 3 frames, respectively. The key frames selected are frames 35, 74, and 85; these frames are shown in Figure 3. The analysis of optimized key frame extraction is presented in Table 2. The execution time of key frame extraction depends on the number of frames in the video and the number of iterations. The feature extraction stage for the ‘1.avi’ video is shown in Figures 4, 5, and 6. The video retrieval results for a few videos are shown in Table 3. The feature selection was implemented using tests for different training percentages and iterations. The details are tabulated in Table 4. The selected features decrease, execution time increases, and the average convergence curve increases as the number of iterations increases from 50 to 100 and the training percentage increases from 20% to 40%. The convergence curves for 50 and 100 iterations are shown in Figures 7 and 8 for 20% and 40% data training. The hybrid PSO-BOA feature selection is robust to varying data ratios as it balances the global and local search by changing the optimization algorithm. The 40% training percentage improves retrieval performance and reduces feature selection. The overall performance of the CVBR system for different video classes is shown in Figure 9. The proposed system is compared with other systems, and the comparison is shown in Table 5. It is observed that our proposed system outperforms other systems with 0.913 precision, a 0.78 recall rate, and a 0.836 F1 score. As the proposed system is based on optimal key frames and optimum features, some are missed, resulting in relevant videos that might not match strongly. On the other hand, with 30 different classes, 78% of videos are still able to be retrieved for their class of videos.

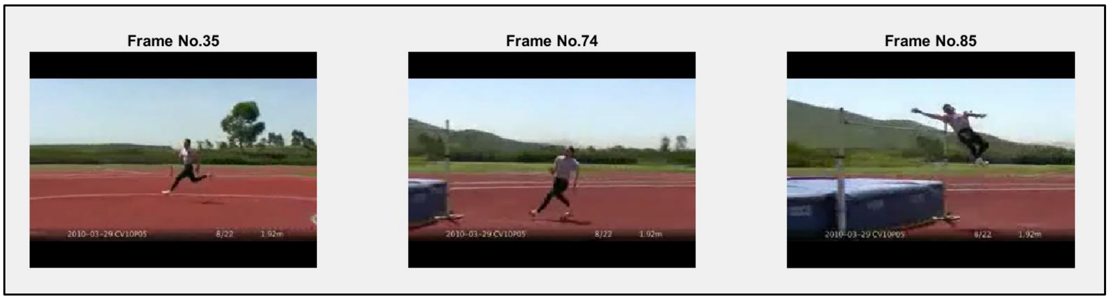
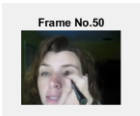
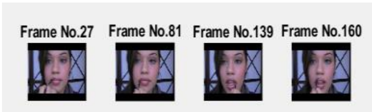
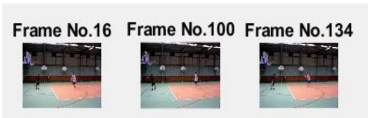
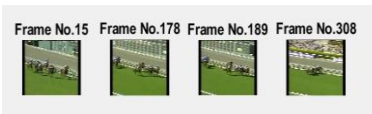
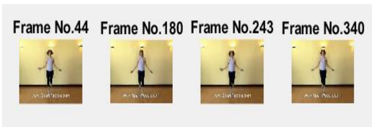


Figure 3. Optimized Key Frames for '105.avi' Video

Table 2. Optimized Key Frame Extraction Analysis on a Few Videos

Video	**No. of Frames**	**No. of Iterations**	**No. of Key Frames**	**Execution Time (seconds)**	**Extracted Key Frames**	**CR**
	112	5	3	22.461		0.97 3
	249	7	3	83.990		0.98 8
	139	6	1	26.906		0.99 3
	92	5	2	16.315		0.97 8
	178	6	4	32.509		0.97 8
	140	5	3	39.438		0.97 9
	361	9	4	200.214		0.98 9
	525	10	4	375.272		0.99 2

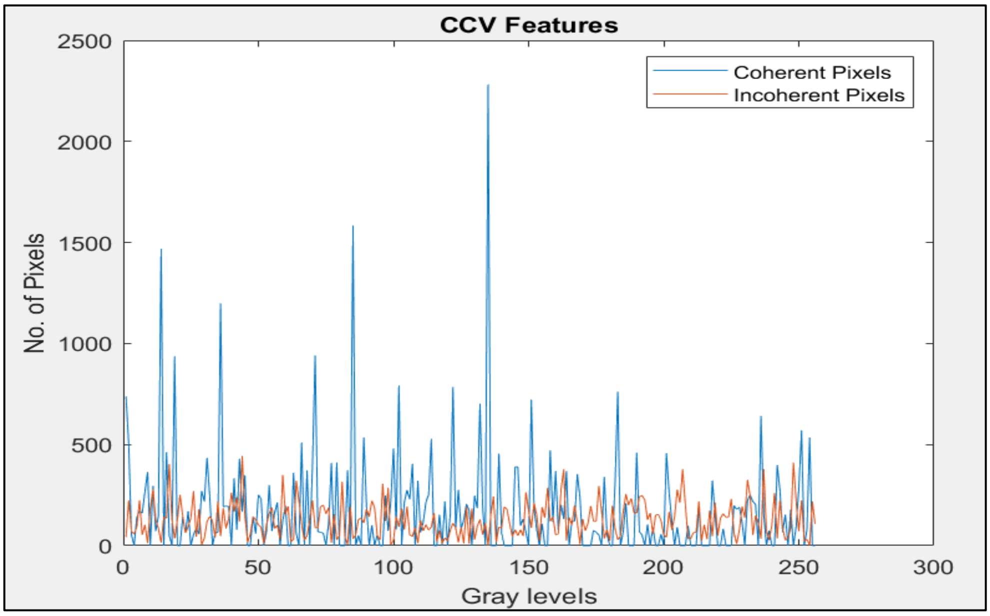


Figure 4. Plot of CCV Features for 1.avi Video

Field	Value
ShortRunEmphasis	0.6428
LongRunEmphasis	30.1610
GrayLevelNonuniform	0.0100
RunLengthNonuniform	0.3823
RunPercentage	1.2207e-05
LowGrayLevelEmphasis	0.0226
HighGrayLevelEmphasis	1.5303e+04
MeanGrayLevel	93.1840
MeanRunLength	2.9619

Figure 5. GLRLM Features for 1.avi Video

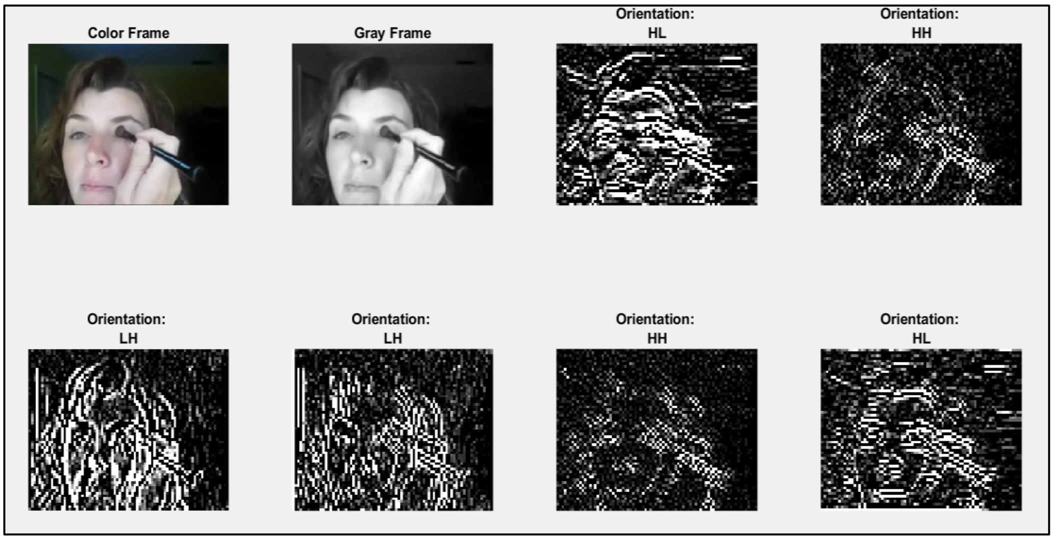


Figure 6. DT-CWT Sub Band Outputs for 1.avi Video

Table 3. Video Retrieval Examples

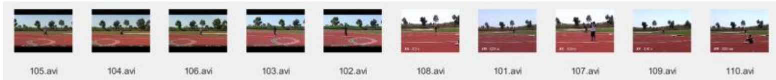
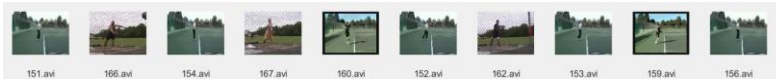
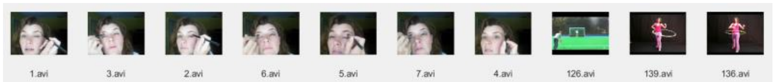
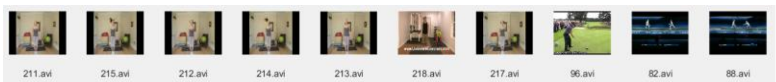
Query	Retrieved Videos	P	R	F1 Score
		1.00	1.00	1.00
		0.60	0.70	0.65
		1.00	1.00	1.00
		1.00	0.90	0.94
		0.80	0.70	0.75
		1.00	1.00	1.00
		1.00	0.70	0.82
		1.00	0.70	0.82
		1.00	0.90	0.94
		1.00	1.00	1.00
		1.00	0.90	0.94
		1.00	0.90	0.94
		1.00	0.90	0.94
		1.00	0.70	0.82

Table 4. Feature Selection Result Analysis

Optimization Algorithm	**Training %**	**Iteration**	**Selected Features**	**Execution time (in Seconds)**	**Average Convergence Curve**	**FRR**	**PR**	**RR**	**F1 Score**
Hybrid Binary PSO-BOA	20	50	2332	2043.381	0.921	0.95988	0.854	0.723	0.783
	40	50	2251	2526.879	0.925	0.96127	0.865	0.754	0.806
	20	100	1710	4841.466	0.908	0.97058	0.896	0.742	0.812
	40	100	1697	5374.247	0.923	0.97080	0.913	0.780	0.841
Average			1998	3696.493	0.919	0.96563	0.882	0.750	0.810

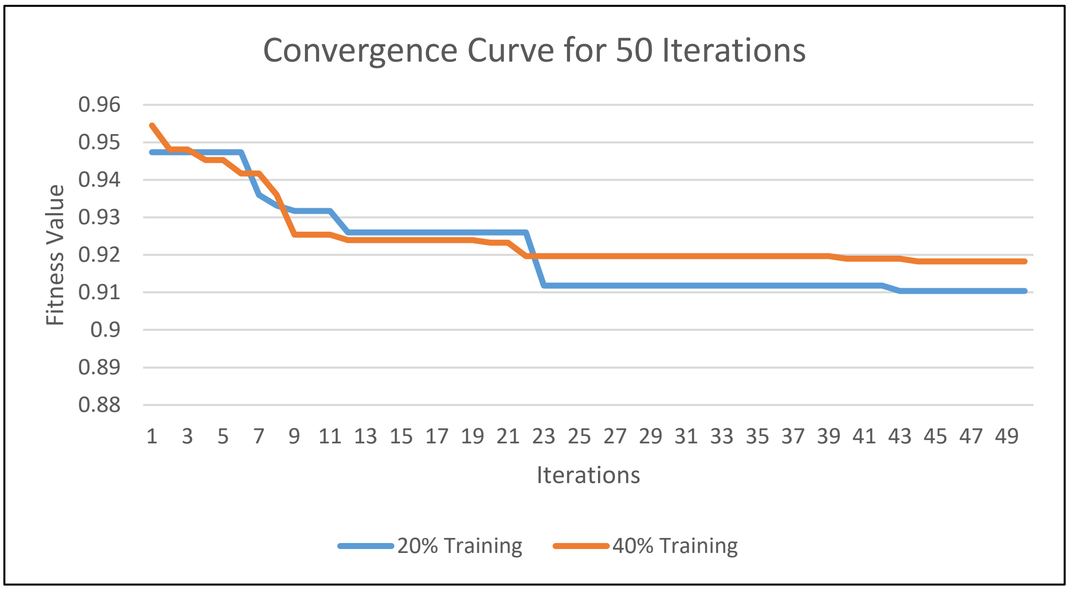


Figure 7. Convergence Curve for 50 Iterations

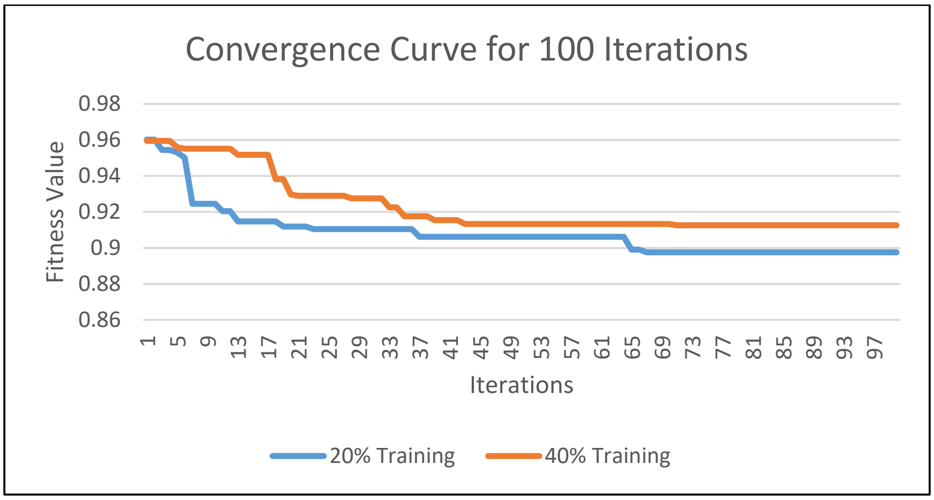


Figure 8. Convergence Curve for 100 Iterations

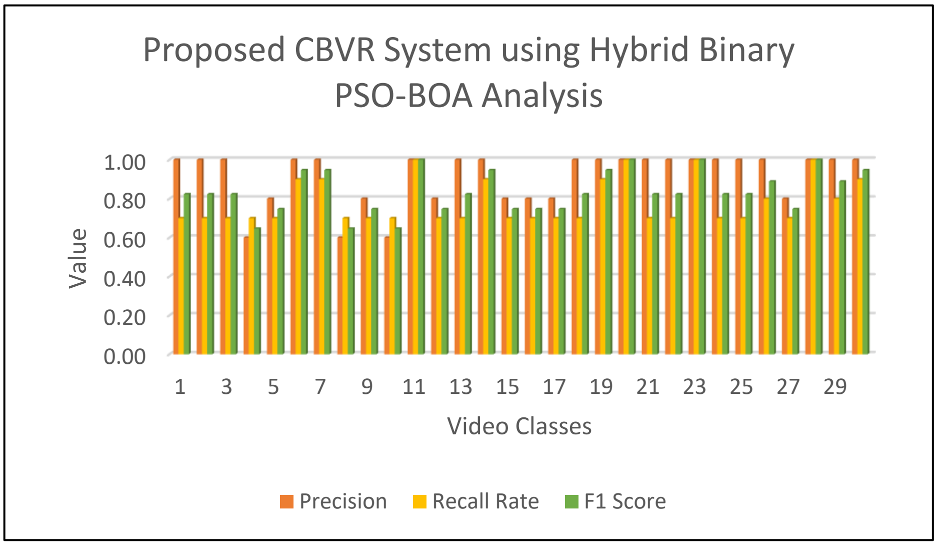


Figure 9. Result Analysis Bar Graph for Different Video Classes

Table 5. Comparison of the Proposed CBVR System with Other Systems

Method	Average Precision	Average Recall Rate	Average F1 Score	Compression Ratio
Proposed System	0.913	0.780	0.836	0.9858
Mallick et al. [3]	0.790	0.552	0.651	-
Thomas et al. [6]	0.71	0.71	0.71	-
Farhan et al. [7]	0.647	0.647	0.647	-
Maojin et al. [21]	-	-	-	0.9360
Rajeshwari et al. [1]	-	-	-	0.9854

5. Conclusion

A new optimized keyframe extraction approach for CBVR systems is proposed in this paper. While feature selection uses a hybrid PSO-BOA algorithm, this method adopts BBOA to diversify and reduce redundancy in keyframe selection. An iterative process like this identifies informative and representative keyframes, achieving a CR of 0.9858. Features extracted include CCV, GLRLM, and DTCWT-based magnitude and orientation to improve performance. To maximize precision and minimize the number of selected features, a multi-objective fitness function is designed. Experiments on the UCF101 dataset were conducted and proved that the PSO-BOA framework enhances retrieval performance.

References

[1] Rajeshwari, D., and C. Victoria Priscilla. "An Enhanced Spatio-Temporal Human Detected Keyframe Extraction." International journal of electrical and computer engineering systems 14, no. 9 (2023): 985-992.

[2] Zhong, Qi, Yuan Zhang, Jinguo Zhang, Kaixuan Shi, Yang Yu, and Chang Liu. "Key Frame Extraction Algorithm of Motion Video Based on Priori." IEEE Access 8 (2020): 174424-174436.

[3] Mallick, Ajay Kumar, and Sushanta Mukhopadhyay. "Video Retrieval Based on Motion Vector Key Frame Extraction and Spatial Pyramid Matching." In 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN), IEEE, 2019, 687-692.

[4] Hussain, Altaf, Mehtab Ahmad, Tariq Hussain, and Ijaz Ullah. "Efficient Content Based Video Retrieval System by Applying AlexNet on Key Frames." ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal 11, no. 2 (2022): 207.

[5] Kar, Tejaswini, Priyadarshi Kanungo, Sachi Nandan Mohanty, Sven Groppe, and Jinghua Groppe. "Video Shot-Boundary Detection: Issues, Challenges and Solutions." Artificial Intelligence Review 57, no. 4 (2024): 104.

[6] Thomas, Sinnu Susan, Sumana Gupta, and Venkatesh K. Subramanian. "Context Driven Optimized Perceptual Video Summarization and Retrieval." IEEE transactions on circuits and systems for video technology 29, no. 10 (2018): 3132-3145.

[7] Hamad, Sumaya, Ahmeed Suliman Farhan, and Doaa Yaseen Khudhur. "Content Based Video Retrieval Using Discrete Cosine Transform." Indones. J. Electr. Eng. Comput. Sci 21, no. 2 (2020): 839-845.

[8] Arora, Sankalap, and Satvir Singh. "Butterfly Optimization Algorithm: A Novel Approach for Global Optimization." Soft computing 23, no. 3 (2019): 715-734.

[9] Chen, Chao, Yuting Wan, Ailong Ma, Liangpei Zhang, and Yanfei Zhong. "A Decomposition-Based Multi objective Clonal Selection Algorithm for Hyperspectral Image Feature Selection." IEEE Transactions on Geoscience and Remote Sensing 60 (2022): 1-16.

[10] Keleş, Mümine Kaya, and Ümit Kiliç. "Binary Black Widow Optimization Approach for Feature Selection." IEEE Access 10 (2022): 95936-95948.

[11] Shami, Tareq M., Ayman A. El-Saleh, Mohammed Alswaitti, Qasem Al-Tashi, Mhd Amen Summakieh, and Seyedali Mirjalili. "Particle Swarm Optimization: A Comprehensive Survey." IEEE Access 10 (2022): 10031-10061.

[12] Su, Hongjun, Qian Du, Genshe Chen, and Peijun Du. "Optimized Hyperspectral Band Selection Using Particle Swarm Optimization." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7, no. 6 (2014): 2659-2670.

[13] Harjoko, Agus. "Dataset Feature Reduction Using Independent Component Analysis with Contrast Function of Particle Swarm Optimization on Hyperspectral Image Classification." In 2016 2nd International Conference on Science in Information Technology (ICSITech), IEEE, 2016, 285-290.

[14] Zhang, Mengjian, Daoyin Long, Tao Qin, and Jing Yang. "A Chaotic Hybrid Butterfly Optimization Algorithm with Particle Swarm Optimization for High-Dimensional Optimization Problems." Symmetry 12, no. 11 (2020): 1800.

[15] Ghaleb, Sanaa AA, Mumtazimah Mohamad, Waheed Ali HM Ghanem, Abdullah B. Nasser, Mohamed Ghetas, Akibu Mahmoud Abdullahi, Sami Abdulla Mohsen Saleh, Humaira Arshad, Abiodun Esther Omolara, and Oludare Isaac Abiodun. "Feature Selection by Multi Objective Optimization: Application to Spam Detection System by Neural Networks and Grasshopper Optimization Algorithm." IEEE Access 10 (2022): 98475-98489.

[16] Salmi, Meryem, and Bachir Boucheham. "Content Based Image Retrieval Based on Cell Color Coherence Vector (Cell-CCV)." In 2014 4th International Symposium ISKO-Maghreb: Concepts and Tools for knowledge Management (ISKO-Maghreb), IEEE, 2014, 1-5.

[17] Saadizadeh, Saman. "Breast Cancer Detection in Thermal Images Using GLRLM Algorithm." In 2021 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), IEEE, 2021, 1-6.

[18] Shrikant Chavate, Ravi Mishra. "An Efficient Approach for Shot Boundary Detection in Presence of Illumination Effects using Fusion of Transforms," International Journal of Engineering Trends and Technology, vol. 70, no. 4, (2022): 418-432, https://doi.org/10.14445/22315381/IJETT-V70I4P236.

[19] Selesnick, Ivan W., Richard G. Baraniuk, and Nick C. Kingsbury. "The Dual-Tree Complex Wavelet Transform." IEEE signal processing magazine 22, no. 6 (2005): 123-151.

[20] Soomro, Khurram, Amir Roshan Zamir, and Mubarak Shah. "Ucf101: A dataset of 101 Human Actions Classes from Videos in the Wild." arXiv preprint arXiv:1212.0402 (2012).

[21] Sun, Maojin. "An intelligent Retrieval Method for Audio and Video Content: Deep Learning Technology Based on Artificial Intelligence." IEEE Access (2024).