International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

A Novel Cluster Based Video Object Segmentation for Key Frame Extraction

Jeyapandi Marimuthu[1], Vanniappan Balamurugan[2]

Submitted: 20/12/2023 **Revised**: 26/01/2024 **Accepted**: 06/02/2024

Abstract: Video key frames are the abstraction of content rich frames of a shot or a video that best reflects the nature of the whole video without redundancy. Object based key frame extraction techniques are capable of extracting key frames that are semantic. These techniques needs to extract the required objects or region through video object segmentation. The segmentation of objects is achieved by Fuzzy C-Means clustering as it distinguishes well across object boundaries. In this paper, Oppositional based Border Collie Optimization algorithm is proposed along with Gaussian Kernel FCM to optimize the centroids of clusters. The accuracy of the segmented objects are evaluated in terms of SSIM, BDE and VoI with the SBM-RGBD dataset. The resultant frames with segmented objects are compared with consecutive frames for change of pose of objects using key points features. When there is a considerable variations between two frames, one of the frames is selected as a key frame. The experimental results showed that the proposed BCOKFE technique improves the accuracy of the extracted key frames to 92% for the WEB data set.

Keywords: *Gaussian Kernel FCM, Clustering, Key Frame, Segmentation, oppositional-based BCO.*

1. Introduction

The usage of digital videos is inevitable due to its technological advancements and cost effectiveness. A video signal is a sequence of two dimensional images projected from a dynamic three dimensional scene onto the image plane of a video camera [1]. The sources of video data are documentaries, movies, animation videos, surveillance videos, sports and news telecasts, etc. The quality and size of the video depends upon the encoding techniques used. In order to accommodate the persistence of human vision, the video sequence is made up of a number of frames with a minimum frame rate of 25 frames per second.

The video can be analyzed by splitting the video into scenes, shots, and key frames. The extracted key frames can further be used for various applications such as video summarization, abnormal detection, and video indexing, etc. Video analysis becomes challenging when the contents are rich and voluminous. Hence, key frames alone are considered instead of whole video as it represents the semantic nature of the whole video without redundancy. This representation reduces the searching time significantly during video retrieval.

Key frame extraction techniques are classified based on histogram, clustering, object contents of a frame, and motion displacement vectors of pixels. Among the existing key frame extraction techniques, the cluster based key frames extraction techniques outperforms better due to its computational simplicity. Though many researchers have applied the cluster based key frame extraction techniques, the accuracy of the extracted key frames remains a constraint.

[1,2] Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli India, 627 012
[1]E-mail: athishmarimuthu@gmail.com [2]E-mail: bala_vm@msuniv.ac.in

The main challenges in cluster based key frame extraction techniques are as follows:

i) The centroids are often stuck in to local optimum

ii) Sensitivity to outliers

iii) Slow convergence rate in case of optimization

The centroid based clustering methods such as K-Means stuck into a local optimum due to sensitivity to outliers and random initialization of centroids. Hence, the representative object based clustering [2] such as fuzzy c-means is opted. In FCM the number of clusters are created and then data items are assigned iteratively to the representative cluster. The patterns are not distinguished well across cluster boundaries due to overlapping nature [3]. Such limitations are overcome by fuzzy clustering approaches. As far as the fuzzy c-means is concerned, the global optimal variables are optimized using any of the optimization technique. In this paper, the Border Collie Optimization technique [4] is adapted to find the global optimum with high convergence rate.

The proposed BCOKFE technique is tested against a WEB benchmark data set and the result reveals that the BCOKFE technique outperforms the existing cluster based key frame extraction techniques in terms of accuracy.

The major contributions of the proposed work are:

i) The initialization of cluster centroids are determined using the oppositional-based BCO algorithm along with Gaussian kernel FCM

ii) The variable stuck in local optimum are adjusted with nearby values in order to find global optimum

iii) Quick convergence is achieved

iv) Effective object segmentation from the selected clusters is done by the level-set algorithm

The input video is converted into sequences of frames and the frames are selected at the ratio of 10:1 for further processing. The foreground of each frame is extracted using blob detection for further segmentation of objects. While clustering, to avoid centroids get stuck in local optimum, the Border Collie optimization is carried out on with the fuzzy c-means. The segmented objects of each frame are compared with successive frames to ensure positional and transformational changes of objects. Whenever, there is a mismatch of objects between two frames, the frames are extracted as key frames.

The rest of the paper is structured as follows: Section 2 reviews the important research works which are relevant to the current work, section 3 deals with the proposed BCOKFE technique, section 4 describes the experimentation and results, and finally section 5, summarizes the BCOKFE technique based key frame extraction.

2. Related Work

The objective of the review is to present an overview of various key frame extraction techniques along with the critical analysis in order to identify the research gap. The research works that are related to key frame extraction techniques are classified as Histogram Based Key Frame Extraction (HBKE), Cluster Based Key Frame Extraction (CBKE), Motion Based Key Frame Extraction (MBKE), and Object Based Key Frame Extraction (OBKE).

The approach of HBKE [5][6][7] is based on the fact that the two frames will have little dissimilarity in their histograms with unchangeable objects and background. The reason behind this is that, for characterizing an image, color is considered to be the best feature. The standard color histogram based algorithm and its variants are widely used for detecting abrupt changes between frames. The limitation of this technique is that two different frames may also be considered as similar frames due to same histogram.

CBKE techniques [8][9] are used to group the frames into clusters and from each candidate clusters a key frame is chosen. Clustering is the powerful mechanism used in various fields such as pattern recognition and information retrieval, etc. The selection of key frames varies based on the number and size of the clusters which in turn varies based on the nature of the content. The similar visual frames are grouped into a cluster, where the visual information includes color, texture, and shape. This method yields better efficiency with improved computational simplicity. However it fails to preserve the temporal order.

In MBKE [10][11], a model of perceived motion energy to model patterns in video and a scheme to extract key frames are employed. Motion displacement vectors are an important feature of this technique. Motion is considered as the significant quality for the presentation of actions in videos and hence should be used in determining key frames. It has a high computational time.

The OBKE techniques segment video objects [12][13][14] using clustering or background subtraction techniques . When a color histogram is used for segmentation, it yields low efficiency with computational simplicity. The color features have limited semantics. As local features of objects have the capability to generate vigorous descriptors and to retain image semantics in a latent fashion, they are applied for key frame extraction.

The clustering techniques are classified by principle viz. hierarchical clustering, graph theoretical clustering, and objective function based clustering. By model it is further classified into deterministic, statistical, and fuzzy clustering. The fuzzy c-means is an objective function based clustering technique.

Cannon et. al [15] presented an approximated fcm technique in which lookup table approach was used for computing the Eucliean distance between the data item and cluster centers. It reduces the complexity of each iteration to one by sixth of time and the feature space is confined to a tuple of finite integers. Park et. al [16] explained a gradient based FCM technique in which the gradients of the objective functions with respect to membership grades are set to zero and only one data item is considered for every iteration. It has fast convergence over traditional FCM. Noordam et. al [17] briefed a geometrically guided FCM technique which includes the geometric neighborhood relationship i.e, membership values of neighboring pixels. The segmented image consists of more homogeneous regions and less spurious pixels. Ahmed et. al [18] described a modified FCM technique in which a regularization term that is inversely proportional to SNR is introduced in fuzzy objective function. The resultant image is free from salt and pepper noise and it is limited to single feature input. Fan et. al [19] presented a suppressed FCM technique which suppresses the less significant membership values. It is not sensitive to fuzzy factors and suffers with slow convergence rate. Chen et. al [20] explained a kernel based spatial FCM technique that uses kernel induced distance measure with spatial constraints.

The computational time increases exponentially due to curse of dimensionality. Kim et. al [21] briefed a FCM technique with novel initialization for color clustering in which color points that are closest to dominant colors are selected as centroids. The CIELAB color space model that was used in this technique well represents two color points correspond to human perception. Prior knowledge of number of clusters is essential to determine the dominant colors. Wang et. al [22] briefed a feature weighted FCM technique in which weighted Euclidean distance metric was used. It supports clusters of non spherical shape by feature space transformation. The time complexity increases due to weight learning. Chuang et. al [23] described a spatial FCM technique in which the spatial function is incorporated along with objective fnction. It resulted into more homogeneous segmented regions that are free from spurious blobs and noise. Dong et. al [24] presented a graph connectivity based FCM technique that is applied to data sets with both numerical and qualitative attributes. It support clusters of arbitrary shape such as spherical, linear, elongated and concave. Mukhopadhyay et. al [25] briefed a SVM based FCM technique in which fuzzy clustering solutions are enhanced using the classifier. The points with high membership values in each cluster are used to train the SVM classifier and in turn the classifier predicts the labels of remaining points. Kannan et. al [26] presented an effective FCM technique with novel initialization of cluster centers. It uses the kernel induced distance measure. Gao [27] briefed an ant colony optimization technique which improves the accuracy for multivariate data clustering. It suffers as it needs to tune multiple parameters. Kapoor et. al [28] explained a grey wolf optimization technique that is capable to show improved exploitation for unimodal problems and exploration for multimodal functions. It consumes less computational time. Verma et. al [29] discussed a particle swarm optimization technique that searches for both local and global optima. It avoids local minima trapping through candidate solution.

As these techniques fail to avoid getting stuck from local optima and also due to poor exploration and exploitation capabilities, there is degradation in the clustering accuracy. To improve the accuracy of clustering and segmentation, this paper proposes an oppositional based border collie optimization technique which applies eyeing mechanism to avoid getting stuck from local minima.

3. Border Collie Optimization Based Clustering for Video Key Frame Extraction

This section defines a problem of key frame extraction and proposes a novel BCOKFE technique for extracting the key frames of a given input color video. The technique avoids getting stuck from local minima by applying the concept of eyeing mechanism which increases the clustering accuracy leads to effective key frame extraction.

The major workflow of BCOKFE technique is illustrated in Figure 1. Consider an input color video sequence of N frames with $r \times s$ dimensions and let F be the set of frames sampled at regular interval T, i.e., $F = \{f_i\}$ where i is the frame sequence number. The value of i ranges from 1 to N / T and $F \subset F_N$. Now the problem is to identify those f_i's, which can be identified as key frames where the key frames are distinct frames that do not share common color features. The variations in similarity distance D between two consecutive frames of a frame sequence determine the key frames. Let F_k be the set of key frames where $F_k \subset F$.

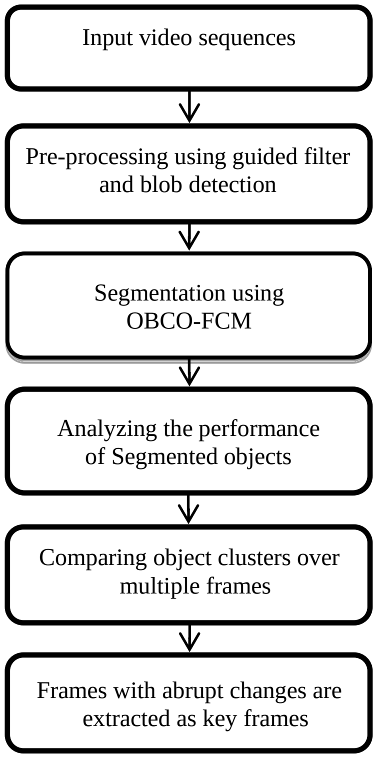

Fig 1. Block diagram of BCOKFE technique

Each frames of set F are resized to 256×256 and subjected to preprocessing using guided filter for noise removal. The resultant noise free set of frames are F_{nf}. The each frame f_i of set F_{nf} is further subjected to segmentation. For segmenting each frame or image, the number of clusters k is to be passed as an argument. As the cluster boundaries are not distinguishable due to overlapping nature and also to get optimized clusters, the BCO-FCM technique is applied.

While clustering, the population n of dimension d is initialized with the following control parameters fitness (f), velocity (v), acceleration (acc), position (pos) and time (t). The entire population is divided into two categories such as border collie dogs (k search agents) and number of sheep in the herd (the n-k cluster members). The first n population that give the best fitness values for the given objective function *obj* are selected as one lead dog and equal number of left and right dogs during every iteration. The dependent parameters *acc* and *pos* are updated for all dogs and sheep based on the independent parameters *t* and *v* for every iteration.

The membership matrix U of the fuzzy objective function is updated for every iteration of clustering based on the distance measure $\| d_i - v_k \|^2$ by considering the positions of each dog as centroids v_k where d_i is the distance of each data item *i* from v_k. The objective function is transformed using the conversion technique of kernel and by adapting the Gaussian function. Oppositional based approach is imposed on BCO-GKFCM to further optimize the initial centroids of the clusters. When the matrix values U remain unchanged with in threshold, the iterations comes to end and the data item with highest membership value to a particular cluster *Cj* is assigned to that cluster.

The resultant frames with segmented objects are compared with consecutive frames for change of objects in terms of position, translational, or rotational transformations using key points features. When there is a considerable variations between two successive frames f_i and f_j the frames, the frame f_i is selected as a key frame and the set of key frames denote $F_k \subset F$. The algorithm for OBCO-FCM is given in Algorithm 1.

Algorithm 1 OBCO-FCM

Input : Frame, Number of Clusters, Fuzziness Coefficient

Output : Optimized Clusters

```
function cluster_labels = obco_fcm(data, k, m)


// Initialize the OBCO parameters
swarm_size = 20;
max_iterations = 100;
// Initialize the cluster centers randomly
cluster_centers = rand(k, size(data, 2));
// Initialize the OBCO swarm
swarm = zeros(swarm_size, size(data, 2));
for every swarm i
   swarm(i, :) = rand(1, size(data, 2));
end
// Initialize the opposition-based swarm
opposition_swarm = zeros(swarm_size, size(data, 2));
for every opposition_swarm i
   opposition_swarm(i, :) = 1 - swarm(i, :);
end
// Initialize the best solution
best_solution = cluster_centers;
% Iterate until convergence.
While !converged
// Calculate the membership degrees of each data point to each cluster
   membership_degrees = zeros(size(data, 1), k);
   for i = 1:size(data, 1)
      for j = 1:k
         membership_degrees(i, j) = (norm(data(i, :) - swarm(j, :))^-(2 /
            (m - 1))) / sum((norm(data(i, :) - swarm(:, :))^-(2 / (m - 1))));
      end
   end
// Update the cluster centers
   for j = 1:k
      swarm(j, :) = sum(membership_degrees(:, j).^m .* data, 1) /
        sum(membership_degrees(:, j).^m, 1);
   end
// Update the opposition-based swarm
   opposition_swarm = 1 - swarm;
// Update the best solution
   if norm(best_solution - cluster_centers) > norm(swarm -
   cluster_centers) & norm(best_solution - cluster_centers) >
   norm(opposition_swarm - cluster_centers)
      best_solution = swarm;
   elseif norm(best_solution - cluster_centers) >
   norm(opposition_swarm - cluster_centers)
      best_solution = opposition_swarm;
   end
end
// Assign each data point to the cluster with the highest membership degree
cluster_labels = max(membership_degrees, [], 2);
```

end

3.2 Frame Conversion

The input frames at a ratio of 10:1 alone are considered for further processing, as the contents of adjacent frames are identical within the range of a video sequence. The lighting changes in the background of RGB videos degrades the performance of KFE algorithm as two similar comparable frames with varying lighting conditions are identified as two key frames. In order to avoid the impact of this luminance effect, the frames are converted into hue, saturation and value color model.

3.3 Blob Detection

The blob detection [30] process is carried out after the frame separation process to identify the particular region of interest for object segmentation. It finds the regions in the image with different properties like colour, brightness, and surrounding areas and represents the presence of objects in a frame. A blob is a group of connected pixels in an image that share some common property. In the BLOB identification process, pixels with intensity values that are greater than the threshold value are checked for with adjacent pixels[4]. Each blob is enclosed by rectangle box which is elastic in nature and can be stretched in both vertical and horizontal directions until the entire blob becomes enclosed. The process is then repeated for all the blobs in the image. As a result of blob detection, the size, shape and number of objects can be identified and further be subjected to object segmentation.

3.4 Gaussian Kernel-based Fuzzy C Means(GKFCM)

The unsupervised method for clustering is FCM. It is applied in several fields like astronomy, medical imaging and also in image segmentation. It groups the images into a cluster by combining related data in a specific feature space which belongs to different groups with the degree of membership. The cost function is minimized iteratively to achieve clustering. Based on the distance between every centre of the clusters and the point of data, this algorithm determines the membership. It is clear that all the data points summation should be equal to one.

The objective of the FCM algorithm is expressed in eq. (1) & (2).

$$N = \sum_{i=1}^{M} \sum_{j=1}^{C} \mu_{ij}^{m} \left\| x_i - C_j \right\|^2 \quad \ldots.. \quad (1)$$

$$u_{ij} = \frac{1}{\sum_{r=1}^{c} \left(\frac{d(C_j, x_i)}{d(C_j, x_r)} \right)^{\frac{2}{m-1}}} \quad \ldots.. \quad (2)$$

where M denotes the count of data points, the required cluster is denoted as C, the cluster centre of j is represented as C_j and $d^2(C_j, x_i) = (C_j - x_i)^2$ is the Euclidean distance between the data point C_j and x_i and the membership level of the group is denoted as μ_{ij}. The parameter m in the degree of membership controls the fuzziness of the segmentation output. When pixels near their cluster's centroid are given high membership values and those far from the centroid are given low membership values, the cost function minimises. The FCM algorithm's probability depends on how far a pixel is from each cluster centre in the feature domain. To overcome the drawbacks of the traditional FCM, the kernel information is involved in the FCM, which manages the few differences between the clusters. The KFCM maps the non-sequential data space in the high-dimensional feature space with the use of the kernel function. The objective of FCM is transformed by introducing the kernel method which is expressed in eq. (3).

$$N = \sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij}^{m} \left\| \varphi(c_i) - \varphi(x_j) \right\|^2 \quad \ldots.. \quad (3)$$

The objective function is converted by using the conversion technique of kernel. The mathematical expression is shown in eq. (4) & (5).

$$\left\| c_i - x_j \right\|^2 = \left(\varphi(c_i) - \varphi(x_j) \right)^T \left(\varphi(c_i) - \varphi(x_j) \right) \quad \ldots.. \quad (4)$$

$$= \kappa(x_j, x_j) + \kappa(c_i, c_i) - 2\kappa(c_i, c_j) \quad \ldots.. \quad (5)$$

Then the Gaussian function [31] is adapted to the kernel function, then the objective function is converted as eq. (6)

$$N = 2 \sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij}^{m} (1 - \kappa(c_i, x_j)) \quad \ldots.. \quad (6)$$

To minimize the objective function, the clustering centre c_i and the membership matrix μ_{ij} can be obtained as expressed in eq. (7) & (8).

$$\mu_{ij} = \frac{\left(1 - \kappa(c_i, x_j)\right)^{\frac{1}{(1-m)}}}{\sum_{k=1}^{c} (1 - \kappa(c_i, x_j))^{\frac{1}{(1-m)}}} \quad \ldots. \quad (7)$$

$$c_i = \frac{\sum_{k=1}^{n} \mu_{ij}^{m} (c_i, x_j) x_j}{\sum_{j=1}^{n} \mu_{ij}^{m} \kappa(c_i, x_j)} \quad \ldots.. \quad (8)$$

If the objective function of the kernel FCN induces the new metrics in the data space which is defined in eq. (9) & (10).

$$d(x, y) = \left\| \varphi(c_i) - \varphi(x_j) \right\| \quad \ldots.. \quad (9)$$

$$= \sqrt{2(1 - \kappa(x, y))} \quad \ldots.. \ (10)$$

where, $d(x, y)$ is the metric in the actual space, the Gaussian kernel function is taken as $\kappa(x, y)$.

It only measures the points and the current cluster centre. When the threshold condition reaches the maximum iteration the execution is stopped. Therefore the oppositional Based BCO algorithm is also applied to the

GKFCM which finds the initial centroids of the clusters which are explained below.

3.5 Oppositional-based Border Collie Optimization (OBCO)

The Border Collie Optimization algorithm is developed by imitating the sheep herding style of the Border Collie dogs which was proposed by [4]. This technique adopts both the front and side of the Border collie's unique herding style using three dogs, one is the leading dog and the others are the left and right dogs. The exploration and exploitation phase has an equal focus by partitioning the whole population into dogs and sheep. The dogs use three herding techniques: gathering, stalking and eyeing. The gathering is the process of bringing the sheep closer to the global best solution. Stalking is the process of moving around the sheep to guide them towards the global best solution. Eyeing is the process of intimidating the sheep to avoid local optima. The sheep follow a simple rule: move away from the nearest dog and towards the nearest sheep. The fitness value of sheep is not higher than dogs. The fitness value of three dogs is represented as F_f for the lead dog, F_l for the left dog, and F_r for the right dog, then the fitness of the sheep is F_s. The optimal solution of the distance between the points of the field to the farms travelled by the dogs and sheep is controlled by velocity, acceleration and time. The velocity of the three dogs is expressed in eq. (11), (12) and (13).

$$v_{fd}(t+1) = \sqrt{v_{fd}(t)^2 + 2 \times acc_{fd}(t) \times pos_{fd}(t)} \quad .... \quad (11)$$

$$v_{rd}(t+1) = \sqrt{v_{rd}(t)^2 + 2 \times acc_{rd}(t) \times pos_{rd}(t)} \quad .... \quad (12)$$

$$v_{ld}(t+1) = \sqrt{v_{ld}(t)^2 + 2 \times acc_{ld}(t) \times pos_{ld}(t)} \quad .... \quad (13)$$

From the above equations, $v_f(t+1)$, $v_r(t+1)$, and $v_l(t+1)$ are the velocity of three dogs at a time $(t+1)$, then the acceleration is denoted as acc_{fd}, acc_{rd} and acc_{ld} with time t, and the position of the dogs are represented as $pos_{fd}(t)$, $pos_{rd}(t)$ and $pos_{ld}(t)$ with time t. The value of D_g is positive when the sheep is nearer to the front dog and moves towards the dog in the front. The fitness value to choose the sheep is expressed in the Equation

The velocity of sheep is updated using gathering, stalking and eyeing techniques of dogs. In the Gathering technique, the value of D_g is positive when the sheep is nearer to the front dog and moves towards the dog in the front. The fitness value to choose the sheep is expressed in eq. (14).

$$D_g = (F_{fd} - F_s) - \left(\left(\frac{F_{ld}+F_{rd}}{2}\right) - F_s\right) \quad .... \quad (14)$$

Then the velocity of the sheep is updated using eq. (15).

$$v_{Gs}(t+1) = \sqrt{v_{fd}(t+1)^2 + 2acc_{fd}(t)pos_{Gs}(t)} .. \quad (15)$$

To keep the sheep on course when using the stalking technique, the sheep closer to the left and right dog must be stalked from the flanks. These sheep are those for which negative D_g values are discovered. The equations to update the velocity of the stalked sheep are presented in eq. (16), (17) & (18).

$$v_{rd} = \sqrt{(v_{ld}(t+1)\tan(\theta_1))^2 + 2acc_{rd}(t)pos_{rd}(t)} \quad .... \quad (16)$$

$$v_{ld} = \sqrt{(v_{ld}(t+1)\tan(\theta_2))^2 + 2acc_{ld}(t)pos_{ld}(t)} \quad .... \quad (17)$$

$$v_{Ss}(t+1) = \frac{v_{ld}+v_{rd}}{2} \quad .... \quad (18)$$

The stalked sheep' velocity v_{ss}be determined by both side dogs which guide the sheep from both sides, therefore the random traversing angles, θ_1 and θ_2 are taken. The range θ_1 is 1 to 89 and the range of θ_2 is 91 to 179 which are selected arbitrarily.

The sheep which are away from the path need eyeing by the dogs which have the least fitness value because the eyeing is used when the fitness of the individual does not increase after several iterations. Therefore, they experienced the retardation which is expressed in eq. (19) & (20).

$$v_{Es}(t+1) = \sqrt{v_{ld}(t+1)^2 - 2acc_{ld}(t) \times pos_{ld}(t)} \quad .... \quad (19)$$

$$v_{Es}(t+1) = \sqrt{v_{rd}(t+1)^2 - 2acc_{rd}(t) \times pos_{rd}(t)} \quad .... \quad (20)$$

The acceleration of the sheep and dogs are updated as $i \in \{f, ld, rd, Gs, Ss\ to\ Es\}$, which is expressed in eq. (21).

$$acc_i(t+1) = \frac{(v_i(t+1) - v_i(t))}{Time_i(t)} \quad .... \quad (21)$$

The update of time for each iteration is done by the eq. (22). Where the dimension of each iteration for the individual is denoted as d.

$$T_i(t+1) = Avg \sum_{i=1}^{d} \frac{(v_i(t+1) - v_i(t))}{acc_i(t+1)} \quad .... \quad (22)$$

To update the dog's population, the basic displacement equation is used which is expressed in the eq. (23).

$$pos_i(t+1) = v_i(t+1)T_i(t+1) + \frac{1}{2}acc_i(t+1)T_i(t+1)^2 \quad ....(23)$$

where pos_i denotes the position of the front, left and right dogs, v_i denotes the velocity of the front, left and right dogs, time of the front, left and right dos are represented as T_i and the acceleration of front, left and right dogs are represented as acc_i with time $t+1$.

Several heuristic optimization algorithms are enhanced their convergence speed and performance using the oppositional based-learning. The goal of this algorithm is to find the optimal solution to the problem by evaluating the opposite population and the estimated population from a similar generation to improve the convergence speed. The convergence speed is increased in case of the random close to the optimal solution. Otherwise, the convergence speed is low.

The BCO algorithm employs oppositional-based learning which utilizes the current best and worst solution of the information to produce a better candidate solution. Evaluate the fitness value of each dog using its position and generate opposite points and use them to create an opposite population. Sort both populations based on their fitness values and select the best solutions from them. Identify the best (a), second best (b) and third best (d) solutions among the selected solutions. Update the positions of the dog and evaluate the fitness value of each dog using its updated position. Then, generate opposite points using the jumping rate and create an opposite population from the current population and select the best solutions from both populations. Finally, the maximum iteration is reached to provide the best centroid of the cluster.

3.6 Level Set Algorithm

In the early days, the level-set algorithm succeeds in computer graphics and vision and also in image segmentation and the shape recovery problem. But there are some limitations while using it in the image segmentation process. If there are any fuzzy or discrete boundaries in the region then finding the result leads to difficulties and it may consume more time by solving the differential equation of each point of the image domain. Therefore, a variant of the level set algorithm [32] which increases the accuracy for identifying the fuzzy object borders by weighting the edge indicator function is proposed. To gain an accurate object counter and shorten the evaluation time, the GKFCM with the oppositional-based BCO algorithm was run simultaneously to the proper beginning contour of the evolution curve. In the proposed paper the appropriately segmented region of the object is extracted using the level set algorithm. It relies on the improved contour C that serves as the zero level set of the graph in the high dimensional function which is expressed as in eq. (23).

$$C_k = \{(x, y)|\varphi(x, y, k) = 0\} \quad .... \quad (23)$$

where C_k represents the artificial time marching parameter k of the counter C then the evaluation of the curve is expressed in eq. (24) with the curve evaluation speed V and the normal vector of inward units N.

$$\frac{\partial c}{\partial k} = VN \quad .... \quad (24)$$

The form of $\varphi(C_k, k)$ is expressed in the eq. (25).

$$\left(\frac{\partial \varphi}{\partial k}\right) + \nabla\varphi . \left(\frac{\partial c}{\partial k}\right) = 0 \quad .... \quad (25)$$

Then the related curve evaluation of LSA is expressed in eq. (26).

$$\frac{\partial \varphi}{\partial k} = V\|\nabla\varphi\| \quad .... \quad (26)$$

The primary curve is used to produce the initial Level set function. Hence, the object from the video sequence is segmented effectively. Then finally the key frames are extracted from the segmented videos that preserve the temporal and spatial coherence of the video content for the extraction and it provide a short and comprehensive representation of the video.

3.7 Key Frame Extraction

Let the input to extract the key frames is the set of sampled frames of a video sequence which is represented as $s = s_1, s_2, \dots \dots s_M$ of M resultant frames after having applied OBCO-FCM algorithm. Pose estimation is the task of estimating the pose of an object in a video frame. The pose of an object can be represented by a set of key points representing various regions of an object. Once the key points were detected, changes in the object's pose can be measured by calculating the difference between the object's pose in the current frame and the object's pose in the previous frame. The feature vector is calculated for every frame to denote its key points. The feature vector of the i^th frame s_i is represented as E_i. The feature vector is calculated for every frame. In the key frame selection process, three frames are selected viz. one present frame and two previous frames which are represented as s_i, s_{i-1}, and s_{i-2} respectively. Based on the change Ratio, Rx is calculated to measure the content of variation according to eq. (27).

$$R_x = \frac{max\{d(s_{i-2}, s_{i-1}), d(s_{i-2}, s_i)\}}{min\{d(s_{i-2}, s_{i-1}), d(s_{i-2}, s_i)\}} \quad .... \quad (27)$$

where d represents the Euclidean distance between the feature vectors. If the R_x value is greater than the predefined threshold then, s_i is selected as a key frame and this feature vector comparison is repeated over multiple frames under consideration in order to extract all possible key frames.

4. Experimental Results

A. Experimental Setup

The main focus of this section is to carry out the experiment on the BCOCKFE using the benchmark datasets such as SBM-RGBD dataset [33] and WEB benchmark dataset [34]. Around 2500 video sequences including natural sceneries, cartoons, TV shows, news and sports events, commercials, and home activities are comprised of in the data sets. The frame rate of videos ranges from 24 fps to 30 fps and the running time of videos varies from 5 secs to 20 secs. The algorithm was implemented in MATLAB version 2013a.

The input frames are selected at the sampling rate of 10 : 1 from the decomposed video. The frames are subjected to preprocessing using guided filter for noise removal. The foreground regions are extracted using blob detection in order to carry out segmentation.

Finding the best outcomes for meta heuristics is greatly influenced by the optimal tuning of exploration and exploitation. Three different herding techniques of the Border Collie dogs are used to obtain the effective feedbacks, which in turn help to find the optimal results. Negative feedback is attained by introducing the eyeing mechanism and the positive feedback is achieved by the gathering and stalking behavior of the Border Collie dogs respectively. The eyeing mechanism introduced in this algorithm serves as an important mechanism to avoid it from getting stuck into local optima. This enhances effective grouping and it is free from outlier and noises that lead to effective segmentation.

The accurate segmentation of foreground objects are ensured using various performance measures such as structural similarity index measure (SSIM), Border Displacement Error (BDE), Variation of information (VoI), and Computational time. SSIM is an index metric that is used to measure the similarity between two images. The BDE measures the average displacement error of one boundary pixels with respect to the closest boundary pixels that belong to other segmentation. The VoI characterizes the dissimilarity (distance) between the original and the segmented image. The various segmentation performance measures such as Computational time, SSIM, BDE, and VoI are depicted in table 1, 2, 3, and 4 respectively. The resultant frames with segmented objects are compared with consecutive frames for change of pose of objects using key points features. When there is a considerable variations between two frames, one of the frames is selected as a key frame. It is possible that the extracted key frames do not fully depict the nature of the video. The measurements true positive (TP), true negative (TN), false positive (FP), and false negative (FN) are generated to assess the accuracy rate of key frame extraction. The accuracy rate of key frame extraction is computed using the eq. (28).

$$Accuracy\ Rate = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \quad \ldots. \quad (28)$$

It shows that the BCOKFE complements the optimized clustering using oppositional based BCO by effective initialization of cluster centroids. The observations are listed as follows.

Table 1. Experimental Results of Segmented Images

Dataset / Methods	SSIM	BDE	VoI	Computational Time (in sec)
Frame 1	0.9678	10.3821	0.9984	31.26
Frame 2	0.9529	10.2908	0.9931	28.92
Frame 3	0.9771	10.1932	0.9826	22.34
Frame 4	0.9447	10.4567	0.9867	32.63
Frame 5	0.9586	10.5347	0.9875	48.58

Table 2. Comparison of SSIM based Accuracy of Segmented Image

Dataset / Methods	ACO	GWO	PSO	BCO
Frame 1	0.8593	0.8939	0.9197	0.9678
Frame 2	0.8917	0.9024	0.9138	0.9529
Frame 3	0.8402	0.9057	0.9369	0.9771
Frame 4	0.8383	0.8912	0.9064	0.9447
Frame 5	0.8274	0.8548	0.9218	0.9586

Table 3. Comparison of BDE based Accuracy of Segmented Image

Dataset / Methods	ACO	GWO	PSO	BCO
Frame 1	2.7859	1.8642	0.9197	0.9984
Frame 2	2.9268	2.1497	0.9138	0.9931
Frame 3	2.5117	1.7349	0.9369	0.9826
Frame 4	2.6352	1.9745	0.9064	0.9867
Frame 5	2.6771	2.3768	0.9218	0.9875

Table 4. Comparison of VoI based Accuracy of Segmented Images

Dataset / Methods	ACO	GWO	PSO	BCO
Frame 1	21.6785	14.6532	12.2761	10.3821
Frame 2	22.7313	16.6517	13.5124	10.2908
Frame 3	19.3497	13.3414	11.3168	10.1932
Frame 4	22.1764	15.6369	13.4706	10.4567
Frame 5	20.9702	17.7472	14.1427	10.5347

B. Result Analysis

The comparative analysis of the performance of the BCO based segmentation technique with the existing techniques such as ACO (Ant Colony Optimization) [27], GWO (Gray

Wolf Optimization) [28], PSO (Particle Swarm Optimization) [29] is given in Figure 2.

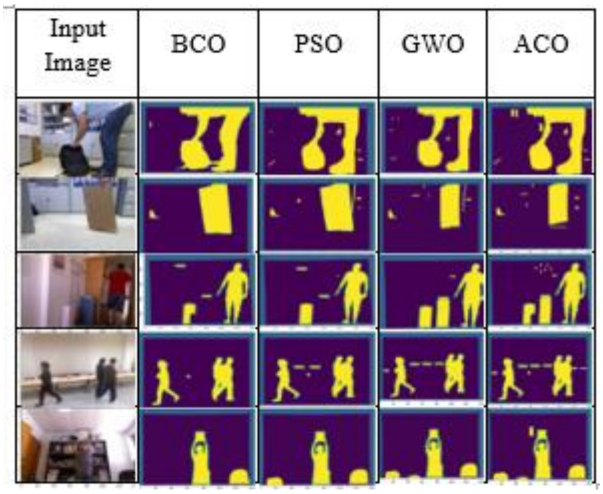

Fig 2. Segmented Images against ACO, GWO, and PSO

The comparative analysis of the performance of the BCOKFE key frame extraction technique with the existing techniques such as video representation based high density peaks search (VRHDPS) [35], Moving Object based Key Frame Extraction (MOKFE) [36], Row Echelon based Spectral Clustering (RESC) [37], and Adaptive Clustering based Key Frame Extraction (ACKFE) [38] is given in Figure 3.

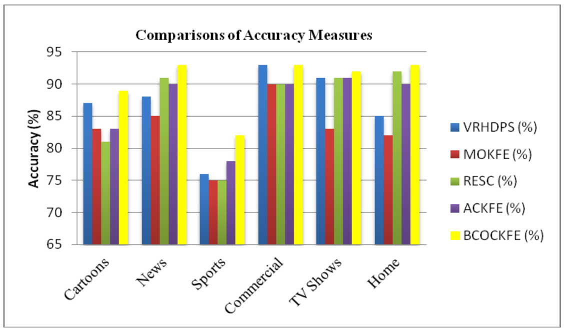

Fig. 3 Comparisons of Accuracy Measures

The proposed BCOCKFE technique performs well against cartoons, news, sports, commercial, tv-shows, and home videos with the accuracy of 89%, 93%, 82%, 93%, 92%, and 93% respectively. The performance measure of this technique outweighs other techniques all types of videos irrespective of redundant events or high content changes. This is achieved by preserving the clustering accuracy by means object based approach and further improves it by means of effective optimization techniques.

5. Conclusion

This BCOKFE technique is proposed to achieve optimal selection of cluster centroids to improve the clustering accuracy and hence to enhance the image segmentation that in turn leads to effective key frame extraction. It applies the eyeing mechanism to avoid getting stuck from local optima. It is capable of maintaining a balanced exploration and exploitation strategy in search space. In future, the complex multimodal fitness function may be introduced to evaluate clustering performance in order to have the search space with massive number of local optima. Further, the clustering and segmentation performance can be improved using deep learning methods.

References

[1] Yao Wang, Jorn Ostermann, and Ya-Qin Zhang, “Video Processing and Communications,“ Prentice Hall, 1st Edition, 2002, ISBN 0-13-017547-1.

[2] S. Ghosh, and S.K. Dubey, "Comparative Analysis of K-means and Fuzzy C-Means Algorithms," International Journal of Advanced Computer Science and Applications, Vol. 4, No. 4, 2013.

[3] J. Nayak, B. Naik, and H.S. Behera, "Fuzzy C-Means (FCM) Clustering Algorithm: A Decade Review from 2000 to 2014," Computational Intelligence in Data Mining, Vol. 2, pp. 133-149, 2015.

[4] T. Dutta, S. Bhattacharyya, S. Dey, and J. Platos, "Border Collie Optimization," IEEE Access, Vol. 8 pp. 109177-109197, 2020.

[5] C.C. Lo, and S.J. Wang, "Video Segmentation using A Histogram-Based Fuzzy C-Means Clustering Algorithm," Computer Standards & Interfaces, Vol. 23, No. 5, pp. 429-438, 2001.

[6] A.K. Mallick, and S. Maheshkar, "Video Retrieval based on Color Correlation Histogram Scheme of Clip Segmented Key Frames," 2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC), IEEE, 2016.

[7] Z. Zong, and Q. Gong, "Key Frame Extraction based on Dynamic Color Histogram and Fast Wavelet Histogram," 2017 IEEE International Conference on Information and Automation (ICIA), 2017.

[8] H. Gharbi, S. Bahroun, and M. Massaoudi, " Key Frames Extraction using Graph Modularity Clustering for Efficient Video Summarization," 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017.

[9] S. Zhao, "Application of A Clustering Algorithm in Sports Video Image Extraction and Processing," The Journal of Supercomputing, Vol. 75, No. 9, pp. 6070-6084, 2019.

[10] A.K. Mallick, and S. Mukhopadhyay, "Video Retrieval based on Motion Vector Key Frame Extraction and Spatial Pyramid Matching," 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN), IEEE, 2019.

[11] Z. Wang, and Y. Zhu, "Video Key Frame Monitoring Algorithm and Virtual Reality Display based on

Motion Vector," IEEE Access, Vol. 8, pp. 159027-159038, 2020.

[12] N. Mlik, W. Barhoumi, and E. Zagrouba, "Object-Based Event Detection for the Extraction of Video Key-Frames," International Conference on Multimedia Computing and Systems, Tangier, Morocco, 2012.

[13] Z. Wu, B. Liang, and G. Wu, "An Object-Behaviour based Key-Frame Extraction Method," International Journal of Simulation--Systems, Science & Technology, Vol. 17, No. 27 2016.

[14] M. Chen, X. Han, H. Zhang, and G. Lin, "Quality-Guided Key Frames Selection from Video Stream based on Object Detection," Journal of Visual Communication and Image Representation, Vol. 65, pp. 102678, 2019

[15] R.L. Cannon, J.V. Dave, and J.C. Bezdek, "Efficient Implementation of the Fuzzy C-Means clustering algorithms," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 2, pp. 248-255, 1986.

[16] D.C. Park, C. Dong, and I. Dagher, "Gradient based Fuzzy C-Means (GBFCM) Algorithm," Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94), Vol. 3, 1994.

[17] J.C. Noordam, W. H. A. M. Van Den Broek, and L.M.C Buydens, "Geometrically Guided Fuzzy C-Means Clustering for Multivariate Image Segmentation," Proceedings 15th International Conference on Pattern Recognition (ICPR-2000), Vol. 1, 2000.

[18] M.N. Ahmed, S.M. Yamany, N. Mohamed, "A Modified Fuzzy C-Means Algorithm for Bias Field Estimation and Segmentation of MRI Data," IEEE Transactions on Medical Imaging Vol. 21, No. 3, pp. 193-199, 2002.

[19] J.L. Fan, W.Z. Zhen, and W.X. Xie, "Suppressed Fuzzy C-Means Clustering Algorithm," Pattern Recognition Letters,Vol. 24, Issues 9-10, pp. 1607-1612, 2003.

[20] S.Chen, and D. Zhang, "Robust Image Segmentation using FCM with Spatial Constraints based on New Kernel-Induced Distance Measure," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) Vol. 34, No. 4, pp. 1907-1916, 2004.

[21] D.W. Kim, K.H. Lee, and D. Lee, "A Novel Initialization Scheme for the Fuzzy C-Means Algorithm for Color Clustering," Pattern Recognition Letters, Vol. 25, No. 2, pp. 227-237, 2004.

[22] X. Wang, Y. Wang, and L. Wang, "Improving Fuzzy C-Means Clustering based on Feature-Weight Learning," Pattern Recognition Letters, Vol. 25, No. 10, pp. 1123-1132, 2004.

[23] K.S. Chuang, H.L. Tzeng, S. Chen, and J. Wu, " Fuzzy C-Means Clustering with Spatial Information for Image Segmentation," Computerized Medical Imaging And Graphics, Vol. 30, No. 1, pp. 9-15, 2006.

[24] Y. Dong, Y. Zhuang, K. Chen, and X. Tai, "A Hierarchical Clustering Algorithm based on Fuzzy Graph Connectedness," Fuzzy Sets and Systems, Vol. 157, No. 13, pp. 1760-1774, 2006.

[25] A. Mukhopadhyay, and U. Maulik, "Towards Improving Fuzzy Clustering using Support Vector Machine: Application To Gene Expression Data," Pattern Recognition, Vol. 42, No. 11, pp. 2744-2763, 2009.

[26] S.R. Kannan, S. Ramathilagam, and A. Sathya, " Effective Fuzzy C-Means based Kernel Function in Segmenting Medical Images," Computers In Biology And Medicine, Vol. 40, No. 6, pp. 572-579, 2010.

[27] W. Gao, "Improved Ant Colony Clustering Algorithm and Its Performance Study," Computational Intelligence and Neuroscience, 2016.

[28] S. Kapoor, I. Zeya, C. Singhal, S.J. Nanda, "A Grey Wolf Optimizer based Automatic Clustering Algorithm for Satellite Image Segmentation," Procedia Computer Science, Vol. 115, pp. 415-422 2017.

[29] H. Verma, D. Verma, and P.K. Tiwari, "A Population based Hybrid FCM-PSO Algorithm for Clustering Analysis and Segmentation of Brain Image," Expert Systems with Applications, Vol. 167, pp. 114121, 2021.

[30] Patro and B. Narayana, "Design and implementation of novel image segmentation and BLOB detection algorithm for real-time video surveillance using DaVinci processor," International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE, Delhi, India, 2014.

[31] Yang, M. Shen, and H. S. Tsai, "A Gaussian kernel-based fuzzy c-means algorithm with a spatial bias correction," Pattern recognition letters, Vol. 29, Issue. 12, pp.1713-1725, 2008.

[32] Reddy G. R, Ramudu K, Zaheeruddin S, and Rao R. R, "Image segmentation using kernel fuzzy c-means clustering on level set method on noisy images," International Conference on Communications and Signal Processing, IEEE, Kerala, India, pp. 522-526, Feb, 2011.

[33] https://rgbd2017.na.icar.cnr.it/SBM-RGBDdataset.html

[34] WEB Dataset. [Online]. Available: http:/ /homepages. inf. ed. ac.uk/rbf/CVonline/ Imagedbase.html

[35] J. Wu, S. Zhong, J. Jiang, and Y. Yang, “ A Novel Clustering Method for Static Video Summarization,” Multimed Tools Appl., Vol. 76, pp. 9625–9641, 2017.

[36] Y. Luo, H. Zhou, Q. Tan, X. Chen, and M. Yun, " Key Frame Extraction of Surveillance Video Based on Moving Object Detection and Image Similarity," Pattern Recognition and Image Analysis, Vol. 28, No. 2, pp. 225-231, 2018.

[37] J. Marimuthu, V. Balamurugan, and S.R. Kumar, " Row Echelon based Spectral Clustering Framework for Key Frame Extraction," Fifth International Conference on Intelligent Computing and Control Systems (ICICCS), IEEE, Tamilnadu, India, 2021.

[38] G. Man, and X. Sun. " Interested Keyframe Extraction of Commodity Video Based on Adaptive Clustering Annotation," Applied Sciences, Vol. 12, No. 3, pp. 1502, 2022.