Improved Deer Hunting Optimization Algorithm for Video-based Salient Object Detection

Puttaswamy M R
Middle East College, Rusayl, Oman
puttaswamymr76@gmail.com

Abstract: Recent days, many researchers have concentrated on the development in salient object detection which is important in numerous computer vision applications. Nevertheless, the main confront is the competent SOD employing the still images. Hence, this work presents the SOD method employing the developed Improved-Deer Hunting Optimization (Improved-DHO) approach. The approach experiences 3 steps that include saliency mapping, keyframe extraction, contourlet mapping, subsequently, a combination of attained outputs by exploiting the optimal coefficients. Initially, extracted frames are given to saliency as well as contourlet mapping concurrently to decide each pixel quality. Subsequently, the outcomes attained from the contourlet mapping and saliency mapping are combined by exploiting the random coefficients in order to attain the last consequence which is used to recognize the salient objects. Moreover, the developed Improved-DHO is used to select the optimal coefficients in order to detect salient objects. The investigational analysis of developed Improved-DHO based on the performance measures exposes that the developed Improved-DHO obtained a maximum accuracy, specificity, and sensitivity.

Keywords: Image Steganography, Secret Sharing, Wavelet Transform, Haar Wavelet, UCID Database

Nomenclature

Abbreviations	Descriptions
SOD	Salient Object Detection
CNNs	Convolutional Neural Networks
DFB	Directional Filter Bank
SOD	Subspace Outlier Degree
GWO	Grey Wolf Optimization
ROI	Regions of Interest
NN	Neural Networks
STAN	Spatiotemporal Attention Neural Network
FTSD	Frequency Tuned Saliency Detection
SUN	Saliency Using Natural statistics
LSTM	Long Short Term Memory
CT	Contourlet Transform
FPR	False Positive Rates
FNR	False Negative Rates
KNN	K Nearest Neighbour

1. Introduction

The SOD intends to accurately and consistently recognize a large amount of visually distinctive areas in a video or image [1]. It has to turn out to be an extremely vigorous research issue in computer vision, since it is capable to sustain high-level visual tasks, like visual tracking, segmentation, photo cropping, and thumbnail creation. In the saliency detection community, it is classified into video and image salient object detection using the input of saliency models. Image salient object has been extensively considered, whilst video salient object detection draws less awareness in the past decades, because of the complexity of extracting healthy motion information in addition to the lack of large-scale explained datasets [3].

In videos, SOD aspires at positioning main foregrounds typically draws the human interest in every frame [8]. For every frame, the saliency map is an output; whereas the value of the pixel shows the equivalent pixel probability belong to a salient object. The superior saliency represents the superior value [4]. Generally, the human notices the SOD which is exploited in many applications like image

segmentation, autonomous driving, image change detection, autonomous facial expression recognition, and so forth whereas task is done.

From various viewpoints, saliency detection techniques can be generally categorized into 2 diverse techniques such as bottom-up as well as top-down techniques. The previous techniques are stimulus-driven that assume human interest based on the visual stimuli themselves without the image content knowledge. Conversely, the top-down concentration approaches are task-driven as well as typically necessitate unambiguous considerate of the scene context [11].

For video content a well-liked method demonstration for the frames extraction which integrates superior level semantic information similar to keyframes. For instance those from surgical operations, a basic requirement is to segment complete video to significant structural units that are well-thought-out to stand for an incessant spatiotemporal action for large video scale data [6]. After keyframes extraction from every shot, it is subsequently simple to set up the general video context as a compilation of illustrations happens as of individual shots. These illustrations might pass on each to the actual keyframes, otherwise to vigorous descriptors for the feature extraction from the keyframes. The previous shows the way to applications associated with the video browsing as well as summarization, whereas rather than the complete video the user might envisage a prechosen keyframes number. From keyframes, extraction descriptors are typically associated with image retrieval as well as video indexing, whereas the objective is for the keyframes retrieve like a query frame.

For salient object detection, several approaches are formulated taking into consideration still images with diverse saliency measures and prediction methods, such as graphical modeling, spatial prior, information theory, regional contrast, or deep learning [9]. The existing algorithms are formulated based on the deep learning algorithms encompass shown potential performance by generous significance to complete salient objects. Hence, several researchers initiated to construct efforts for using CNNs in SOD and have obtained better performance, as CNN's has the sturdy capability to eliminate high-level features abandoning the handcrafted features [10]. Nevertheless, various conventional SOD techniques require large scale CNNs which is trained for using pooling operations and strided convolution. A 2 stream CNN was formulated to learn information; however, these techniques are exploited to train the techniques on multi-frame that reason extreme calculational burden.

The main objective of the paper is to formulate a method in order to recognize salient objects by exploiting a recently formularized improved optimization approach called, Improved-DHO approach. The developed algorithm is consists of 3-steps, whereas the salient objects are recognized by exploiting the input video frames to effectual SOD. Initially, the frames are given into the saliency and contourlet mapping stage in order to extract the background as well as characteristics of the properties by means of the following frames. Subsequent to the recognition of objects, an improved optimization approach called Improved-DHO is used subsequent to experiences the contourlet and saliency mapping by choosing the optimal coefficients. After merging the 2 outputs produced by contourlet as well as saliency mapping, the position of the salient objects is decided.

2. Literature Review

In 2020, Mohammad Shokri et al [1], developed a new non-local fully CNN model in order to capture the global dependencies additional competently as well as examine employ of newly developed non-local NN in video SOD. The consequence of non-local operations was examined individually on dynamic as well as static saliency detection to employ both motions as well as appearance features. For video SOD, a new deep non-local fully CNN model was developed and examined on 2 renowned datasets FBMS as well as DAVIS.

In 2019, Yi Tang et al [2], developed a 2-stream based STAN for video salient object detection. The motion information was sufficiently extracted regarding the LSTM network and 3D convolutional operation from video sequences and optical flow-based prior. In addition, a conscientious element was modeled to incorporate the diverse kinds of spatiotemporal feature maps by learning the equivalent weights. In the meantime, to produce adequate pixel-wise interprets video frames, then produce many coarse labels that were fine used to train a vigorous saliency prediction network.

In 2019, PingZhang et al [3], worked on a new stereoscopic video saliency detection algorithm on the basis of the depth confidence optimization and spatiotemporal correlation. Initially, spatial saliency was attained using color contrast calculation as well as additional improving spatial correlation among neighbors. Next, temporal saliency was computed using propagating motion information among frames. In sequential frames, the exacting propagation and reverse sequential frames increase the temporal correlation. Moreover, depth confidence optimization was presented to combine the temporal and spatial saliency to produce the best saliency map adaptively.

In 2020, Qiong WANG et al [4], presented an outline of new advancements in this domain and evaluates the equivalent techniques modern, such as classification of the conventional techniques and their models; review of benchmark datasets as well as usually exploited estimate measures; investigational evaluation of the performances of the conventional technique; recommendations of few shows potential future works for uncertain issues.

In 2019, Yuming Fang et al [5], worked on a new spatiotemporal saliency detection model. Initially, the temporal, as well as spatial saliency maps, were computed using minimum-level visual features. Furthermore, these two saliency maps were merged of video sequences for spatiotemporal saliency prediction. The spatial saliency map was computed and 3 types of features were extracted namely texture, luminance, as well as color, whilst temporal saliency map was calculated from video sequences in order to extract the motion features calculated. To combine temporal and spatial saliency maps a new adaptive entropy-based uncertainty weighting algorithm was modeled to forecast the concluding spatiotemporal saliency map using the Gestalt theory. From spatial saliency, the Gestalt principle of similarity was exploited to calculate the spatial uncertainty.

In 2018, Constantinos Loukas et al [6], developed a new algorithm of the operational video from individual shots for keyframe extraction. Initially, the laparoscopic videos were segmented to video shots by exploiting an objectness technique that was trained to detain importantly alters in the endoscope area for analysis. Subsequently, into 3 saliency maps every frame for a shot was festering to design the predilection of human vision to areas with superior differentiation regarding motion, color, as well as texture.

3. Proposed Algorithm for Salient Object Detection

In this section, an effectual technique is exploited to perform the recognition of salient objects by exploiting the video frames set. The developed technique experiences 2 phases to detect the salient objects that include mapping as well as the coefficient selection stage. At first, the input video is used from that the extraction of the frames is done and it is subjected to the mapping stage. The saliency, as well as contourlet mapping, is done by exploiting the frames in the mapping phase. The frame exhibits the different each pixel quality comprised in the frame of the saliency mapping. The aspire of the saliency mapping is to compress else change the image model into a somewhat new thing that is highly comprehensible as well as easy to verify. For equal elevation, to connect the points an imaginary line is drawn as well as allows studying the objects' shape in contourlet mapping. Hence, contourlet mapping estimates the images as well as creates smooth regions divided using smooth boundaries to present faster execution. Subsequently, the coefficients of the output produced employing the contourlet as well as saliency mapping are combined by exploiting the novel designed optimization approach named developed Improved-DHO approach. Fig. 1 demonstrates the schematic model of the adopted Improved DHOA method.
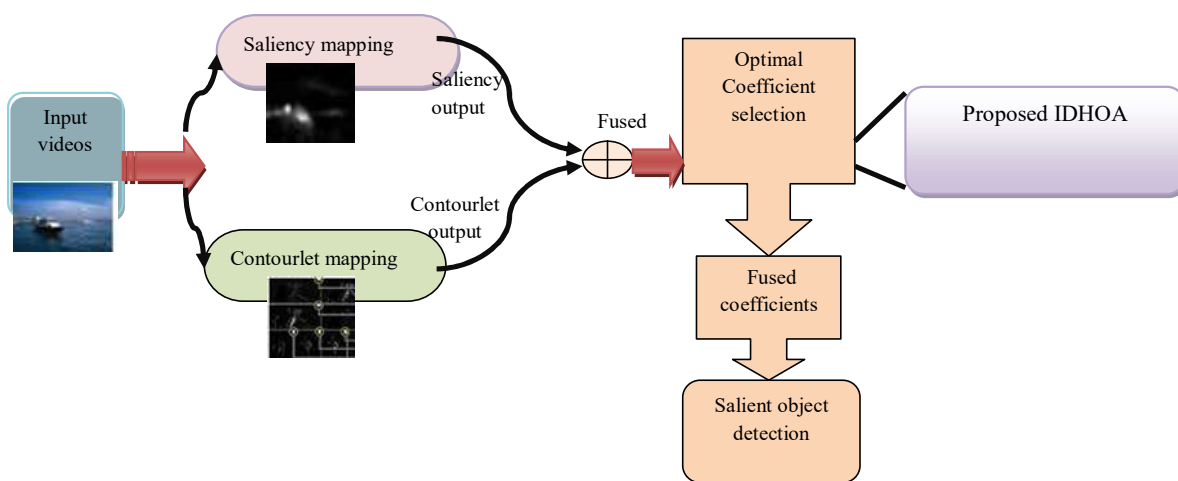

Fig. 1. *Architecture model of the proposed Improved DHOA method*

3.1 Keyframe Extraction by Exploiting the Input Videos

The process of SOD decides well-known objects by exploiting the image into the input video that extremely seizes notice of humans. The keyframe extraction is modified to increase system quality to

represent a video comprises of a high count of frames. Consider I as the input database consists of a number of videos that are stated in eq. (1).

$$I = \{V_q; \quad 1 \le q \le c\}$$
(1)

Subsequently, each video cause's d number of frames as well as it is stated as below:

$$V = \{D_g; \quad 1 \le g \le d\}$$
(2)

Among d number of frames, an important keyframe D_g is extracted in order to determine the salient objects, whereas $1 \le g \le d$.

3.2 Salient Objects Detection using Mapping Phase

The attained keyframe is fed to the mapping stage in that the saliency as well as contourlet mapping used for transformation. In saliency mapping, images exhibit the every pixel quality is exploited to simplify the demonstration of an image into a particular point which is highly noteworthy and easy for inspection. Likewise, the frame creates a multi-resolution directional fixed frame to evaluate the images for faster execution in contourlet mapping. In addition, the contourlet mapping creates an imaginary line that connects the equivalent elevation points as well as allows studying objects shape.

3.3 Salient Features for Estimating the Quality of the Pixel

In this section, the process of saliency mapping [7] is elaborated, in that the salient features are exploited to mapping. Moreover, a frame D_g saliency mapping is performed to show individual pixels quality. A saliency map aims to create uncomplicated image delineation that is highly important and easy to estimate. The ROI is mapped by exploiting the salient features as well as region intensity of frame which are extracted from the videos in saliency mapping.

The process of segmentation is performed to attain for each frame D_g, the super-pixels $A = \{A_1, A_2, ..., A_l, ... A_P\}$ whereas, P indicates the total number of superpixels as well as it is presumed as 100. Eq. (3) indicates the histogram features of the lth superpixel, whereas, p indicate the number of superpixels, whereas $1 \le u \le p$, k_{lv}, m indicate number of histogram bins, whereas $1 \le v \le m$, and k_{uv} indicates the superpixel histograms.

$$H_l = \sum_{u=1}^{p}\sum_{v=1}^{m}\frac{(k_{lv} - k_{uv})^2}{(k_{lv} + k_{uv})}$$
(3)

Eq. (4) indicates the global contrast color features of lth superpixel, whereas, $d(y_l, y_u)$ indicates Euclidean distance among the uth and l^{th} superpixel color values y_u and y_l.

$$I_l = \sum_{u=1}^{P} d(y_l, y_u)$$
(4)

Likewise, the local contrast color features of lth superpixel are stated in eq. (5), whereas, $W_{l,u}^r$ indicates the weight function as well as $r_l \in [0,1] \times [0,1]$ indicates the regularized location of superpixel l^{th} as well as it is stated in eq. (6).

$$L_l = \sum_{u=1}^{P} W_{l,u}^r d(y_l, y_u)$$
(5)

$$W_{l,u}^r = \frac{1}{X_l}\exp\left(\frac{-1}{2\sigma_r^2} \| r_l - r_u \|_2^2\right)$$
(6)

In eq. (6), σ indicates a constant which is represented as 0.25, and X_l indicates normalization factor, r_u indicates the normalized location of uth superpixel, and r_l indicates the normalized location of lth superpixel.

The global saliency is calculated presuming the salient region pixels which possess equal color distribution. From the salient regions background the saliency mapping aids to detach. The main aspire of the saliency mapping is to decide the amalgamation of coefficients of the color in a space. To enhance typical power, power-law alterations are used to the coefficient of every color after normalizing the coefficients [0, 1]. Moreover, 3 gamma values are exploited, that is 0.5, 1.0, and 2.0. Hence, the high-dimensional matrix to indicate colors of the image is stated in eq. (7).

$$R = \begin{bmatrix} S_1^{\gamma 1} & S_1^{\gamma 2} & S_1^{\gamma 3} & T_1^{\gamma 1} \\ S_2^{\gamma 1} & S_2^{\gamma 2} & S_2^{\gamma 3} & T_2^{\gamma 1} \\ \vdots & & & \\ S_P^{\gamma 1} & S_P^{\gamma 2} & S_P^{\gamma 3} & T_P^{\gamma 1} \end{bmatrix} \tag{7}$$

In eq. (7), T_l and S_l indicates the test image of lth superpixel of T as well as S, the color channel of the matrix R by $z = 11 \times 3$. The background as well as foreground and color coefficients are changed to detach the salient regions as well as background color and the issue of z_2 regularized least square issue and it is stated in eq. (8) in order to attain the saliency map, whereas, ρ indicates the weighting parameter for controlling β, $\beta \in F^z$ indicates the coefficient vector.

$$\min_{\beta} \| (Q - \tilde{R}\beta) \|_2^2 + \rho \| \beta \|_2^2 \tag{8}$$

In eq. (9), R indicates a $Y \times Z$ matrix, whereas Y indicates the number of color samples. Moreover, every $\tilde{R}$ associated with the color samples and it is stated in eq. (9), whereas, h_p and f_n indicate the lth foreground superpixel between complete subpixels as well as u^{th} background candidate superpixel between complete subpixels, p and n indicate the total number of regions.

$$\tilde{R} = \begin{bmatrix} S_{f_1}^{\gamma 1} & S_{f_1}^{\gamma 2} & S_{f_1}^{\gamma 3} & T_{f_1}^{\gamma 1} \\ S_{f_n}^{\gamma 1} & S_{f_n}^{\gamma 2} & S_{f_n}^{\gamma 3} & T_{f_n}^{\gamma 1} \\ S_{h_1}^{\gamma 1} & S_{h_1}^{\gamma 2} & S_{h_1}^{\gamma 3} & S_{h_1}^{\gamma 1} \\ S_{h_p}^{\gamma 1} & S_{h_p}^{\gamma 2} & S_{h_p}^{\gamma 3} & T_{h_p}^{\gamma 1} \end{bmatrix} \tag{9}$$

Because of the high color samples than the coefficient vector dimensions, l_2 regularized square problems are contemplated as a crucial issue which is to be minimized by presuming β as β^*. Hence, the saliency map for globally is stated in eq. (10).

$$A_I(M_l) = \sum_{u=1}^{z} R_{l,u}\beta_u \ , \ l = 1, 2, \cdots, P \tag{10}$$

The detection of the salient presents a better outcome using minimum FPR as well as has disadvantages that are simply pretentious with textures as well as hence, it comprises maximum FNR. A local saliency map is modeled based on the spatial and color distances regarding the neighboring pixel to surmount this disadvantage. For every superpixel, M_l, the K-nearest foreground superpixel $M_f = \{M_{f_1}, M_{f_2}, \dots M_{f_k}\}$ is decided as well as K-nearest background superpixel $M_h = \{M_{h_1}, M_{h_2}, \dots M_{h_k}\}$ is formulated by exploiting M_f and M_h as features. The Euclidean distance among the K nearby background and foreground of lth superpixel is stated in eq. (11).

$$d_{f_l} = \begin{bmatrix} \| r_l - r_{f_1} \|_2^2 \\ \| r_l - r_{f_2} \|_2^2 \\ \vdots \\ \| r_l - r_{f_K} \|_2^2 \end{bmatrix} \qquad d_{h_l} = \begin{bmatrix} \| r_l - r_{h_1} \|_2^2 \\ \| r_l - r_{h_2} \|_2^2 \\ \vdots \\ \| r_l - r_{h_K} \|_2^2 \end{bmatrix} \tag{11}$$

From l^{th} superpixel, the feature vector of color distance to KNN foreground $d_{Cf_l} \in F^{8K \times 1}$ as well as background $d_{Ch_l} \in F^{8K \times 1}$, it is stated in eq. (12).

$$d_{Cf_l} = \begin{bmatrix} d(y_l, y_{f_1}) \\ d(y_l, y_{f_2}) \\ d(y_l, y_{f_K}) \end{bmatrix} \qquad d_{C,h_l} = \begin{bmatrix} d(y_l, y_{h_1}) \\ d(y_l, y_{h_2}) \\ d(y_l, y_{h_K}) \end{bmatrix} \tag{12}$$

Once the local saliency map A_L, as well as global saliency map, A_I is produced, subsequently the amalgamation of local as well as global saliency map is performed to produce the last saliency map. Here, 2 algorithms are exploited to integrate saliency maps. The primary algorithm is to do the pixel-wise multiplication of 2 maps as well as stated in eq. (13).

$$A_E = \frac{1}{X}(p(A_I) \times p(A_L)) \tag{13}$$

In eq. (13), $p(.)$ indicates a pixel-wise integration function, X indicates a normalization function, A_L indicates the global saliency mapping outcome and A_I indicates the global saliency mapping outcome.

The aforesaid integration exhibits the restrains bright pixels as well as darker pixels. The second approach integrates the 2 maps exploiting summation and it is devised in eq. (14).

$$A_E^{'} = \frac{1}{X}\left(P(A_I) + P(A_L)\right) \tag{14}$$

The 2 maps are integrated to augment the analysis as well as superior weightage is stated to extremely salient regions. The values of the weight are set up and it is evaluated saliency map using the optimal weight as well as ground truth for the summing up is stated in eq. (15), whereas, B indicates the ground truth.

$$\min_{\substack{W_1 \geq 0, \ W_2 \geq 0 \\ W_3 \geq 0, W_4 \geq 0}} \| W_1 P(W_2 A_I) + W_3 P(W_4(A_L)) - B \|^2 \tag{15}$$

The eq. (15) is exploited to determine the best weights which are used to optimize non-negative least-squares to attain an optimal solution. Hence, the ultimate saliency map is stated in eq. (16).

$$A_f = \frac{1}{X}\left(W_1 P(W_2 A_I) + W_3 P(W_4 A_L)\right) \tag{16}$$

The performance is significantly enhanced after combining the 2 maps as well as the extremely salient regions are conserved as well as the FNR regions are eradicated.

3.4 Frame Representation using Contourlet Mapping

In numerous image directions, the CT is a transformation method modified to represent the even contours and presents directionality and anisotropy and it is shown in Fig 2. In an image, the algorithm modified a twice filter bank to detect the points discontinuities, where DFB is used to link discontinuities of point to a linear structure. The LP presents a means to attain the decomposition of the multi-scale. In every decomposition, a down-sampled input image is generated as well as the detailed image among maximum frequencies is produced. In addition, the DFB is a 2-D directional filter bank exploited to obtain perfect restoration. In two phases, the DFB used in contourlet transform is a process, creating twenty-one subbands using partitioning the wedge-shaped frequency. By exploiting DFB, contourlet transform is a sophisticated 2D wavelet transform. In different scales, contourlet development comprises diverse images that are slanting at diverse directions by possible feature ratios. In addition, it is important to record that contourlet which equivalents both directions as well as the position of image contours produces important coefficients. Fig. 2 exhibits the contourlet mapping of the input video frame. Concerning CT to input frame D_g produces 4 bands as well as it is stated in eq. (17).

$$CT(D_g) = \{LH, LL, HH, HL\} \tag{17}$$

Moreover, every band indicates the frequency and energy of contourlet coefficients. The contourlet bands attained by exploiting CT applications have diverse properties. The output attained after using CT on the frame is stated using, S_N.

3.5 Integration of Contourlet and Saliency Mapping

One outcome of contourlet mapping, as well as saliency mapping, is attained, subsequently, the ensuing outputs are amalgamated to produce effectual mapping. For the process of fusion, the 2 coefficients α and β are used. Hence, in the best way, the coefficients are chosen by exploiting the proposed technique. The output which is the integrated image improves object's quality in attendance in the frame. The principle of the method is to protect the object sharpness, saliency, as well as completeness.

$$U = \alpha A_f + \beta S_N \tag{18}$$

In eq. (18), S_N indicates the output attained after contourlet mapping and A_f indicates the output attained after saliency mapping, the two coefficients α, and β is used for the combination procedure.

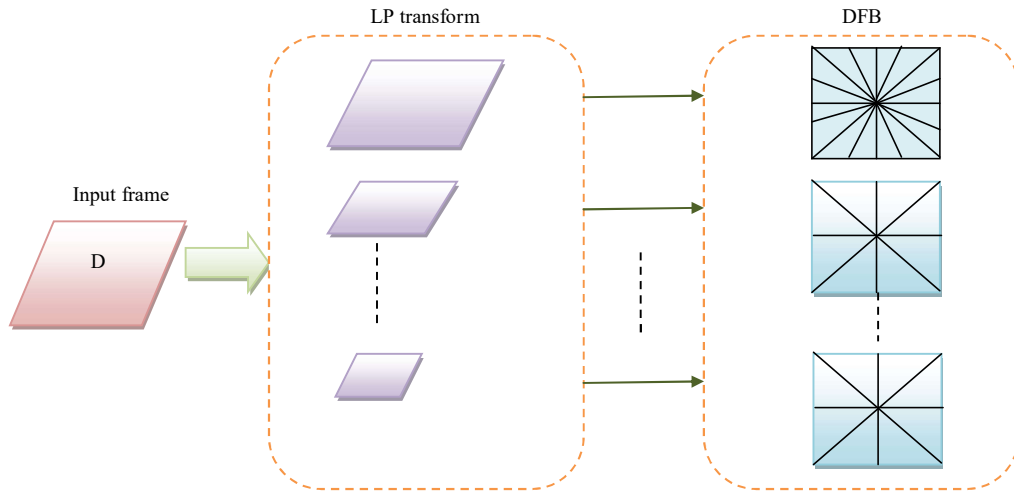**Fig. 2.** *Schematic diagram of the contourlet mapping*

4. Proposed Improved DHOA

A novel optimization algorithm based on the deer hunting process is presented in [12]. Because of the distinctive capability of deer, it can simply getaway by the predators from the hunting. The method initiates with a vector for the arbitrary population named hunters. This procedure is stated as eq. (19), Y indicates the total population and n indicates the number of hunters population.

$$Y = \{y_1, y_2, ..., y_n\} \tag{19}$$

The proposed method initiates with initializing the major parameters of the method like deer's position angle as well as wind angle. In this method, search space is presumed as a circle. Hence, wind angle is stated using the circle circumference.

$$\theta_i = 2\pi\delta \tag{20}$$

In eq. (20), c indicates the current iteration and δ indicates an arbitrary number in the range [0, 1]. Also, the location angle of the deer is devised in eq. (21), θ indicates the wind angle.

$$\omega_i = \pi + \theta \tag{21}$$

Subsequently, location propagation which includes successor location $\left(Z^s\right)$ which is the succeeding hunter location as well as the leader location $\left(Z^l\right)$ that denotes the first optimal location of the hunter is applied.

i) The leader position

After implementing and initializing the first iteration, by the updating, hunters try to attain the optimal location. Therefore, eq. (22) describes the encircling behavior.

$$Z_{i+1} = Z^l - X \times S_w \times \left[L \times Z^L - Z_i\right] \tag{22}$$

In eq. (22), Z_i indicates the current positions and S_w indicates an arbitrary value based on the wind speed within the interval 0 as well as 2, Z_{i+1} indicates the next locations, L as well as X indicates the coefficient vectors that are devised in eq. (23) and (24).

$$L = 2 \times \tau \tag{23}$$

$$X = 0.25 \times \log\left(I + \frac{1}{I_{max}}\right)\beta \tag{24}$$

In eq. (23), I_{max} indicates the utmost iteration, τ indicates an arbitrary value in the interval 0 and 1 and β indicates an arbitrary parameter in the interval -1 as well as 1. The updating of the location (Z, Y) indicates a hunter initial location that is updated based on the prey location. The location updating can be enduring formerly the optimal location $\left(Z^*, Y^*\right)$ is attained based on the L and X.

Hunters try to carry on into the direction of the leaders. Formerly leader possesses an ineffective movement, the hunters can stay put in their present location. The updation of the location is based on the eq. (23) merely while $S_w < 1$; namely, the hunters able to travel arbitrarily in diverse directions with no deliberation for location angle based on the Eq. (22) and Eq. (23).

ii) The location angle

In the rule of the update, the solution space is enhanced and taking into consideration the location angle. The angle assessment is most important for the assignment of the position for hunter such which prey is unacquainted with an attack which creates the hunting procedure triumphant. The angle of the visualization for the deer is formulated as eq. (25)

$$v_i = \frac{1}{8} \times \pi \times \delta \qquad (25)$$

Considering the variation among the wind angle as well as prey visual, a parameter can be devised which aids angle updating of the location, θ describes the wind angle.

$$p_i = \theta_i - v_i \qquad (26)$$

Eq. (27) is used for the updating of the position angle.

$$\omega_{i+1} = \omega_i + d_i \qquad (27)$$

Subsequently, based on the attained location angle, the novel location is attained using eq. (28).

$$Z_{i+1} = Z_1 - S_w \times |\cos(\omega_{i+1}) \times Z^1 - Z_i| \qquad (28)$$

Since the hunter is beyond the deer visual angle, so the deer cannot be observed.

iii) The successor location

This phrase aids the method to augment the exploration based on encircling behavior. By taking into consideration a first arbitrary search, the value of the vector L is not represented higher than 1. Hence, the successor location is assumed to update location rather than primary optimal solution consequence that creates a global search for an approach that is stated in eq. (29).

$$Z_{i+1} = Z^S - X \times S_w \times |L \times Z^1 - Z_i| \qquad (29)$$

In eq. (29), Z^S indicates the successor location of hunters from the current population. The method updates location of the hunters iteratively based on the optimal solution. This procedure creates an optimal solution while $|L| \geq 1$. The hunter can select arbitrarily while $|L| < 1$.

These procedure consequences in a switch named L which alters method among exploration and exploitation stages. An inadequacy of DHOA is its premature convergence. In subsequent, a novel method is used to minimize this disadvantage. The most important disadvantage of DHOA is that because of the abundance of arbitrary parameters, it occasionally consequences premature convergence.

To overcome this disadvantage, this paper uses a method to adapt the most important DHO parameters method to enhance its convergence speed. The solution parameters to enhance the DHO method are δ, S_w, τ and β.

Chaos theory indicates the arbitrary science procedure. The principle of chaos theory is to investigate the extremely dynamic sensitive systems that are pretentious by some minute deviations. Considering the aforesaid elucidation, the generation diversity in the DHO method can be extended using the chaos theory to enhance convergence speed [13] [14]. In the next, a characteristic outline of chaos theory is stated in eq. (30).

$$CT_{j+1}^k = f\left(CT_j^k\right) \quad k = 1,2,....n \qquad (30)$$

In eq. (30), n indicates the dimension of the map as well as $f\left(CT_j^k\right)$ indicates the chaotic generator model. In the IDHO algorithm, parameters δ, S_w, τ and β are modeled based on the sinusoidal chaotic map as sated as below:

$$\delta = v_1 \theta_k^2 \sin(\pi\theta_k) \qquad (31)$$

$$S_w = v_2 \theta_k^2 \sin(\pi\theta_k) \qquad (32)$$

$$\tau = v_3 \theta_k^2 \sin(\pi\theta_k) \qquad (33)$$

$$\beta = v_4 \theta_k^2 \sin(\pi\theta_k) \qquad (34)$$

In eq. (31), k indicates the iteration number and $\theta_0 \in [0,1]$, $v_j \in [0,4]$, $j = 1,2,3,4$.
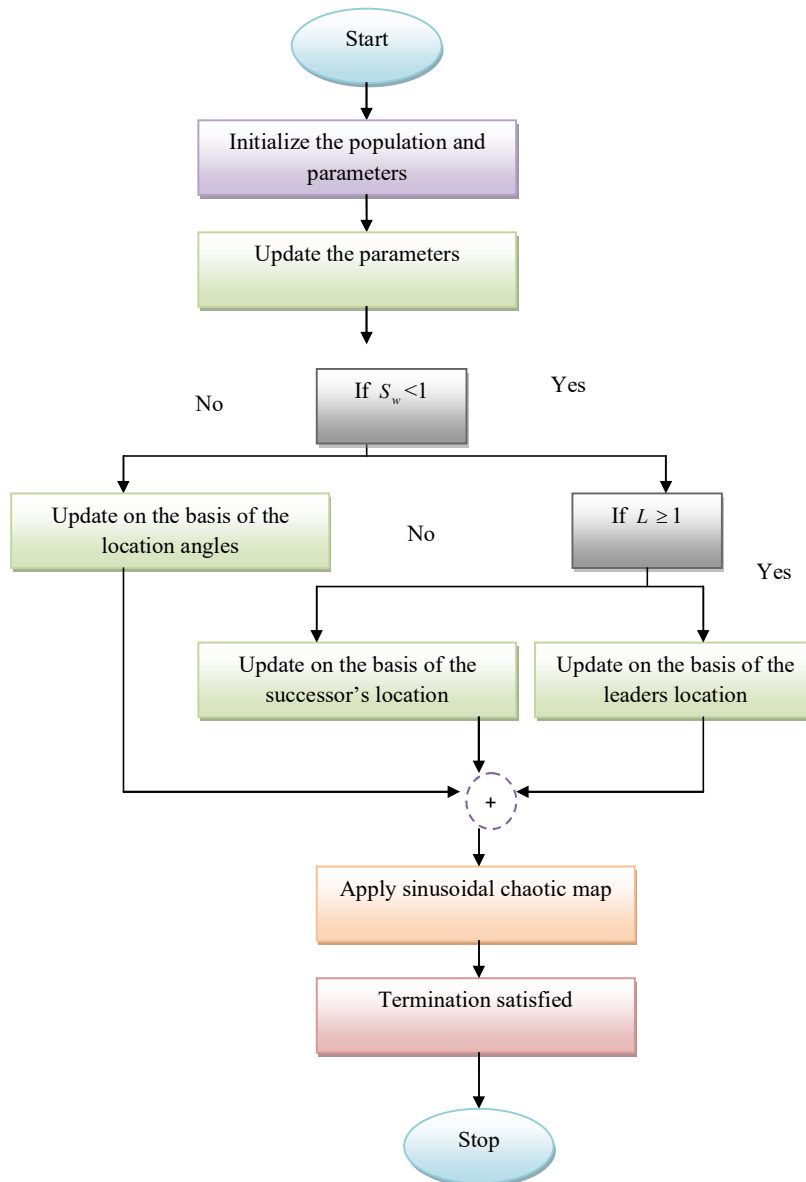
Fig 3 demonstrated the flow chart of the IDHOA.

Fig. 3. *Flow chart of the proposed IDHOA*

5. Results and Discussions

5.1 Experimental Procedure

In real-time, the simulation was performed from the DAVIS dataset [15] by gathering input annotated videos. In this dataset, every video sequences, either single case or multiple cases were interpreted. Moreover, 90 sequences were reinterpreted with diverse objects. Additionally, a small number of sample videos were measured as the input for executing salient object detection with querying any arbitrary search. Also, 3 videos were presented as an input, from that salient objects were recognized.

The evaluation was attained by performing the performances of the conventional methods, such as SUN [17], Spectral Residual technique, FTSD [16], GWO, and the proposed method by exploiting 3 metrics, namely accuracy, specificity and sensitivity.

5.2 Performance Analysis

Fig 4 demonstrated the performance analysis of conventional SUN, Spectral Residual method, FTSD, and GWO as well as a developed technique based on the accuracy, specificity, and sensitivity metrics. Moreover, the analysis is performed by using the utmost number of frames. The overall analysis exhibits that the developed technique validates superior to conventional algorithms.
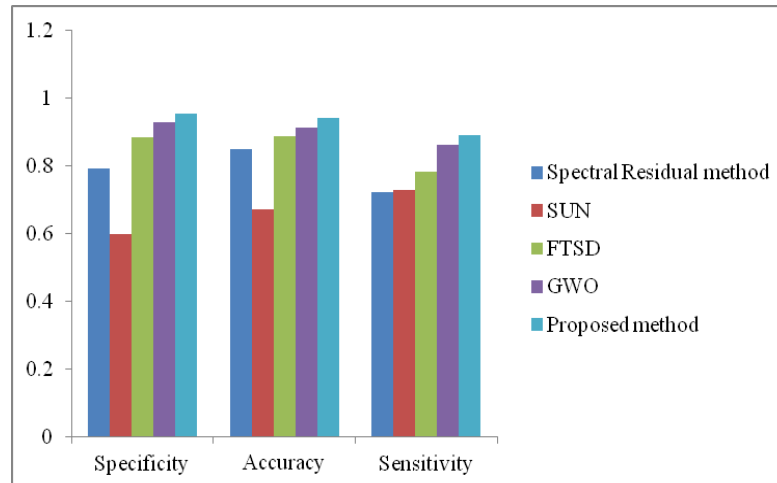
Fig. 4. *Analysis of the developed technique for various metrics*

6. Conclusion

On the basis of the saliency and contourlet mapping, recognition of salient objects was done that aspire to recognize the objects by exploiting the developed Improved-DHO approach from the image data. The contourlet and saliency mapping were exploited to demonstrate every pixel quality, while the contourlet mapping was exploited to attain the smooth contours for the presented image. The approach experiences three steps that include saliency mapping, keyframe extraction, contourlet mapping, and subsequently, the combination of outputs. At first, from large videos, the input video was given to the keyframe extraction that was in charge of extract the keyframes. Subsequently, the attained frames are given to contourlet and saliency mapping to detect the salient objects. At last, the outcomes produced from the contourlet and saliency mapping were combined to attain the outcome that was exploited for salient object detection. Moreover, the best coefficients were decided by exploiting the developed Improved-DHO approach. Also, the developed Improved-DHO approach chooses the best coefficients to detect salient objects. The investigational analysis of the developed Improved-DHO approach based on the performance measures, which exhibits that the developed Improved-DHO obtained maximum accuracy, specificity, and sensitivity.

References

[1] Mohammad Shokri, Ahad Harati, Kimya Taba,"Salient object detection in video using deep non-local neural networks", Journal of Visual Communication and Image Representation, Volume 68, April 2020.

[2] Yi Tang, Wenbin Zou, Yang Hua, Zhi Jin, Xia Li," Video salient object detection via spatiotemporal attention neural networks"Neurocomputing, Volume 37715, February 2020, Pages 27-37.

[3] Ping Zhang, Jingwen Liu, Xiaoyang Wang, Tian Pu, Zhengkui Guo,"Stereoscopic video saliency detection based on spatiotemporal correlation and depth confidence optimization", Neurocomputing, Volume 377, 15 February 2020, Pages 256-268.

[4] Qiong Wang, Lu Zhang, Yan Li, Kidiyo Kpalma,"Overview of deep-learning based methods for salient object detection in videos", Pattern Recognition, Volume 104, August 2020.

[5] Yuming Fang, Xiaoqiang Zhang, Feiniu Yuan, Nevrez Imamoglu, Haiwen Liu,"Video saliency detection by gestalt theory", Pattern RecognitionVolume 96December 2019.

[6] Constantinos Loukas, Christos Varytimidis, Konstantinos Rapantzikos, Meletios A. Kanakis,"Keyframe extraction from laparoscopic videos based on visual saliency detection", Computer Methods and Programs in Biomedicine, Volume 16, 5October 2018, Pages 13-23.

[7] Kim, J., Han, D., Tai, Y.W. and Kim, J., "Salient region detection via high-dimensional color transform and local spatial support," IEEE transactions on image processing, Volume.25, no.1, page no.9-23, 2016.

[8] Wang, Z., Ren, J., Zhang, D., Sun, M. and Jiang, J., "A deep-learning based feature hybrid framework for spatiotemporal saliency detection inside videos," Neurocomputing, Volume.287, page no.68-83, 2018.

[9] Chen, Y., Zou, W., Tang, Y., Li, X., Xu, C. and Komodakis, N., "SCOM: Spatiotemporal constrained optimization for salient object detection," IEEE Transactions on Image Processing, Volume.27, num.7, page no.3345-3357, 2018.

[10] Zhang, P., Liu, W., Lu, H. and Shen, C., "Salient Object Detection with Lossless Feature Reflection and Weighted Structural Loss," IEEE Transactions on Image Processing, 2019.

[11] Weng, S.K., Kuo, C.M. and Tu, S.K., "Video object tracking using adaptive Kalman filter," Journal of Visual Communication and Image Representation, Volume.17, number.6, page no.1190-1208, 2006.

[12] Brammya, G., Praveena, S., Ninu Preetha, N., Ramya, R., Rajakumar, B., Binu, D.,"Deer Hunting Optimization Algorithm: A New Nature-Inspired Meta-heuristic Paradigm". The Computer Journal, 2019.

[13] Rim, C., Piao, S., Li, G., Pak, U., "A niching chaos optimization algorithm for multimodal optimization", Soft Computing, vol. 22, no.2, page no 621-633, 2018.

[14] Yang, D., Li, G., Cheng, G., "On the efficiency of chaos optimization algorithms for global optimization", Chaos, Solitons & Fractals, vol. 34, no.4, page no 1366-1375, 2007..

[15] "Annotated Video Segmentation (DAVIS) dataset," from https://davischallenge.org/davis2017/code.html Accessed on March 2019.

[16] Achanta, R., Hemami, S., Estrada, F., & Susstrunk, S, "Frequency-tuned salient region detection," In proceedings of Conference on Computer Vision and Pattern Recognition, page no. 1597–1604, 2009.

[17] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell, "Sun: A bayesian framework for saliency using natural statistics," Journal of Vision, Volume.8, num.7, 2008.