Journal of Information Systems Engineering and Management
2025, 10(9s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

An Analytical Approach for Enhanced Video Stabilization Using a Bio-Inspired Neural Network

S. Afsal [1], Dr. J. Arul Linsely [2]
[1]*Research Scholar, Noorul Islam Centre for Higher Education, Kumaracoil,TamilNadu, India*
[2]*Associate Professor, Noorul Islam Centre for Higher Education, Kumaracoil,TamilNadu, India*
yuvaafsal222@gmail.com [1], arullinsely@gmail.com[2]
Corresponding Author: yuvaafsal222@gmail.com

ARTICLE INFO

Received: 29 Oct 2024

Revised: 20 Dec 2024

Accepted: 30 Dec 2024

ABSTRACT

The number of videos and photos shot over the past decade has dramatically increased owing to the rapid developments of hand-held digital cameras. However, viewers often see their videos difficult to watch, particularly because of severe camera shake and abrupt movements. As a result, this paper proposes a significant algorithm which results in creation of high-quality video stabilisation, thereby eliminating distracting jitters from amateur recordings and make them appear as though they were shot along smooth, purposeful camera tracks. Modern video stabilization algorithms typically possess an architecture that enables them to realize several benefits: detecting motion in video, constructing a motion model, refining the motion, and ultimately generating stabilized output frames. The video stabilisation includes a feature extraction mechanism in which Genetic Algorithm (GA) is employed due to its characteristics like invariant to scale, rotation, translations, illumination, and blur. In this paper, Whale Optimization Algorithm (WOA) tuned Convolutional Neural Network (CNN) technique is suggested for performing accurate estimation of inliers and outliers. The interframe motion is then estimated by fitting a set of matched point pairs into a linear transform model. Experimental analysis states that proposed system outperforms better than other state-of-art models under various measures (Rotation:94.2, Blue:95.1, Wrap:96.3 and Time cost:97.5).

Keywords: Video Stabilization, Deep Learning, WOA-CNN, Genetic Algorithm, Video Processing

INTRODUCTION

In today's world, it's common to use videos as a means of documenting everyday life. Handheld cameras are often utilized by individuals to capture significant occasions such as trips and social gatherings. These amateur recordings are affected by camera jitter because there are no professional stabilising tools present [1]. Such instabilities lower the quality of captured recordings and have an impact on subsequent operations, like video coding or surveillance, function [2]. The recent proliferation of inexpensive 360-degree cameras is expanding their use and, consequently, their accessibility to a large population. By utilizing a spherical video format, they offer a user-friendly introduction to Virtual Reality (VR) that allows individuals to enter and fully engage with a digital environment. VR is a component of mixed reality in a more general sense [3, 4]. With the expanding usage of video sequences in daily life, digital stabilisation technology with great flexibility, easy handling and cheap has emerged as a research hub [5, 6].

The majority of proposed solutions tackle the problem on a global scale by utilizing offline computations to predict and smooth camera paths. However, only a handful of online stabilization techniques are capable of performing a real-time, low-latency "capture, compute, and display" process for every incoming video frame [7].To enable 360° viewing of video, the framework has undergone a cube map translation. Omnidirectional videos use cube map [8] conversion in which spherical signals that come from cameras are with a full 360° view field. By monitoring the movements of the user’s head, the size of the sphere visible to the user is dynamically adjusted, which enhances their sense of immersion during recording. This interactive component has a significant influence on the user's overall experience, surpassing that of conventional videos [9, 10]. As a result, a feature extraction module has been incorporated into the video stabilization process, with multiple methods available for feature extraction. Figure 1 shows the conventional schema of video stabilization using neural network.

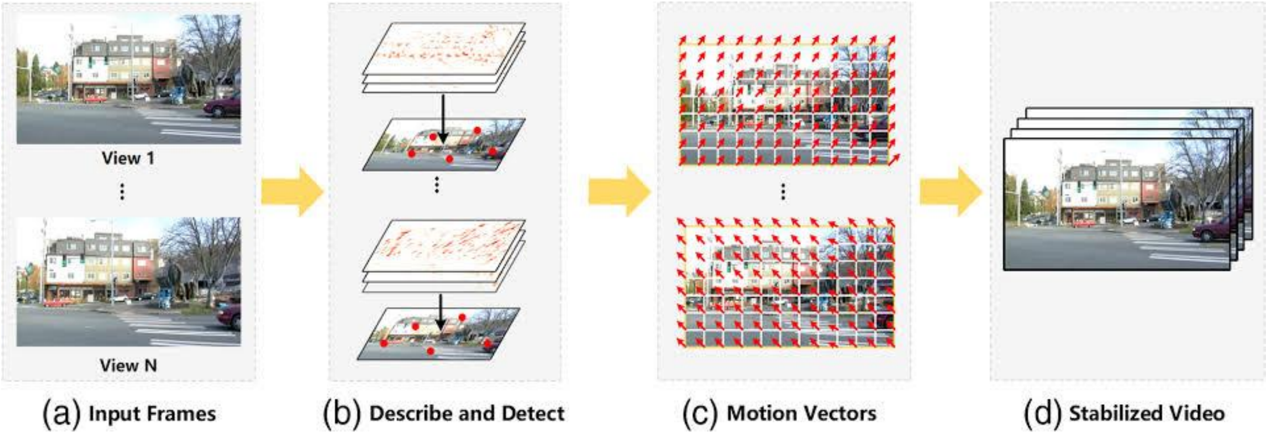

Figure 1. Overall schema of conventional video stabilization using neural network

The global motion estimation and motion correction module are the two main components of a digital video stabilisation system. The performance of motion correction during stabilisation phase is directly impacted by the quality of motion estimation, hence it is crucial to estimate global motion accurately [11]. The most recent techniques for global motion estimation frequently use scale invariant feature transform (SIFT) features [12], which are thought to be suitable for image scaling and rotation. Due to the enormous dimensionality of its feature vectors, SIFT is, however, not very effective [13]. Although SURF [14, 15] is another method that has been demonstrated to be effective and accurate, its applicability is limited due to assumption that foreground objects are more frequently found in image centre and are therefore less likely to be clipped in corners. To tackle such issues the proposed work utilizes Genetic Algorithm for accurate feature extraction process. To enable real-time processing, camera movement is determined using mesh flow or an affine transformation, or by utilizing homography. In this study, video stabilisation is concentrated based on neural network is introduced [16].

In recent times, the fields of computer vision and graphics have been transformed by Convolutional Neural Networks (CNNs) [17]. CNN-based techniques operate more precisely and effectively. For instance, CNNs are used to re-address a number of conventional video processing problems, including video stylization and video deblurring [18]. However, two key impediments to CNN-based stabilisation are found. First, there is a dearth of training data. To train a CNN model, pairs of stable and unsteady synchronised recordings with same recording route and contents are needed. While it is not essential for conventional approaches, it is for stabilising strategy based on learning. The second difficulty is in defining the problem correctly. Traditional stabilisation techniques calculate and smooth a camera path, which makes it difficult to apply them to a convolution neural solution [19, 20]. Therefore, it is necessary to define the problem slight differently with the adoption of optimization technique. In the way of providing solution WOA-CNN approach is established in the proposed work in which utilization WOA tunes the parameters of CNN and provides high resolution, accuracy and reliability.

In this work, to improve the viewing experience of captured video, a reliable and effective method for stabilising 360-degree camera motion is proposed. A novel genetic algorithm method for determining camera orientation from a set of features is established. Also, the proposed method accounts for estimation of feature uncertainty and suppression of outliers using a WOA-CNN.

KEY OBJECTIVES

The research work focus on bringing an effective neural network based video stabilization in which following are the objectives;

- Developing an effective CNN based video stabilization
- To overcome the instability, non-reliable and shaky vidoes , usage such quality and effective stages of neural network.
- With the ability to avoid local optima and get a global optimal solution, usage of WOA gives a boostage to classifier efficacy

- Experimental results states that proposed system outperfoms better on various measures

LITERATURE REVIEW

Xu et al. (2022) [21] approach involves multiple steps. First, we use a unique multi-homography estimation strategy and a motion refinement network to estimate the movement of keypoints and grids. Then, we use temporal association to obtain trajectories based on the grids. Finally, we utilize a new network that predicts dynamic smoothing kernels to smooth the trajectories, which can adjust to different dynamic patterns. By utilizing the consistency of keypoints and grid vertices in both space and time, we have developed a training approach that does not require supervision. Our experimental results on publicly available benchmarks demonstrate that DUT surpasses current leading techniques in terms of both quality and quantity.

Zhu et al. (2022) [22] investigates a method for video stabilization is suggested, involving a motion field that is sparse and spatially smooth, with motion vectors present solely at the mesh vertexes. The authors use neighboring correspondences and a median filter to assign a unique motion vector to each vertex. The video stabilization is carried out on the vertex profiles, which are collections of motion vectors obtained from the same vertex location over time. The proposed technique is evaluated both quantitatively and qualitatively, and the results demonstrate that it is comparable to latest methods and exhibits greater stability in challenging situations.

Chen et al. (2022) [23] article proposes an affordable technique for stabilizing videos that have fast and significant shaking frames. The primary approach of the suggested method is to identify the best match of feature points to produce an optimal transformation matrix, ensuring real-time and superior video stabilization for such frames. Additionally, the method utilizes image pre-processing by downsampling the image and setting the ROI area to enhance the processing speed without compromising the feature point detection.

Li et al. (2022) [24] robust video stabilization method with high real-time performance for mobile devices is proposed. We first employed a fast feature extraction method based on block-wised gradient. Then, the motion information of adjacent frames can be used to get the global camera path through the feature point-pairs. Finally, we smooth it and calculate the corresponding compensation matrix to obtain the stable video. However, large foreground object may lead to wrong motion estimation. Thus, we proposed a fast foreground segmentation method based on GMM for jitter scenes to eliminate feature points on foreground objects and get a more accurate camera trajectory.

Mudassar et al. (2022) [25] Suggest a reliable action detection process that can effectively handle camera motion disturbance and validate its effectiveness through experimentation. In particular, our approach involves aligning features of actors across different frames and integrating global scene features with localized actor-specific features. To achieve feature alignment, we adopt a unique formulation of the STSN i.e " Spatio-temporal Sampling Network" that employs a pyramid structure to predict and refine multi-scale offsets. Additionally, we introduce a new input-dependent weighted averaging technique to combine local and global features.

In their 2022 study, Singh et al. [26] introduce a method for real-time video stabilization that uses approximate computing principles to identify and correct camera movement. Their approach involves determining the direction and magnitude of camera motion, distinguishing intended movement from unintentional shaking, estimating the magnitude and direction of unwanted shake, and applying corrections. The authors describe the technical aspects of their method, including its design and implementation, as well as its limitations. They also propose enhancements to overcome these limitations by leveraging the principles of approximate computing.

Dervişoğlu et al., (2021) [27] examine novel learning-based techniques for video stabilization interpolation that do not rely on annotated data.The authors discuss the shortcomings and potential improvements of these methods, which are based on supervised learning approaches that require recording both stable and unstable versions of the same video. The paper highlights the importance of developing learning-based methods that can work without labeled data.

Yang et al. (2021) [28] present a novel video stabilization approach that combines a multi-region grey projection method and a long-short term memory (LSTM) encoder-decoder network. The proposed method calculates the camera's motion by using the grey projection of four areas in each frame, eliminates camera jitter and main movement direction, and then applies the LSTM network to predict and stabilize the video. The authors evaluate the performance of the proposed algorithm on videos with jitter and demonstrate its ability to achieve real-time video stabilization while improving object tracking accuracy and trajectory analysis.

Luchetti et al. (2021) [29] propose an offline stabilization process for 360-degree videos using a unique technique that obtains rotations with a PSO i.e " Particle Swarm Optimization algorithm". The authors incorporate an uncertainty estimation method among features and use a modified " Chauvenet criterion " to suppress outlier features. Additionally, they use a time-weighted color filter to handle various types of camera movement such as translational jitter, rolling shutter wobble, parallax, and lens deformation. The method is validated using virtual and real 360-degree video data, and a user study is conducted to evaluate the simulator sickness before and after the stabilization process.

In their 2023 paper, Afsal and Arul Linsely [30] introduce a video stabilization method that utilizes feature extraction modules to track gestures in the video, fit a motion model, smooth the motion, and generate stabilized output frames. The authors adopt SURF techniques for feature detection and matching and propose a hybridized approach that combines RANSAC and MSAC to perform accurate estimation of inliers and outliers. They use matched point pairs to estimate the interframe motion by fitting them into an affine transformation model.

Table 1. Comparison analysis of authors and it's methodology

Authors	Methodology	Remarks
Xu et al. (2022)	DUT	Computationally light Worked based on supervised than unsupervised
Zhu et al. (2022)	VVS	Performed on vortex profiles Comparitively less efficacy
Chen et al. (2022)	ROI frame extraction	Performed on larger and fast shaking frames No additional stages of analysing certain features
Li et al. (2022)	UECO	Optimization is less Didn't worked on real time analysis Computational complexity is high
Mudassar et al. (2022)	StSN	The results show a 4.1% improvement in frame mean average precision (mAP) and a 17% improvement in video mAP. However, it is worth noting that these improvements were achieved on a small dataset, which may limit the generalizability of the findings.
Singh et al. (2022)	Approximate computing	Discusses the technical details of using approximate computing for video stabilization, including its design, implementation, and limitations
Dervişoğlu et al. (2021)	Learning-based interpolation	This article examines recent learning-based interpolation techniques used in video stabilization and discusses their limitations and possible enhancements.

Yang et al. (2021)	Multi-region grey projection method and LSTM encoder-decoder network	The proposed approach suggests a video stabilization algorithm that utilizes a combination of a multi-region grey projection method and an LSTM network to predict camera jitter and stabilize the video accordingly.
Luchetti et al. (2021)	Particle Swarm Optimization algorithm and modified Chauvenet criterion	Develops a method for offline stabilization of 360◦ video using a particle swarm optimization algorithm and modified Chauvenet criterion to suppress outliers, and validates it through a user study
Afsal and Arul Linsely (2023)	Hybridized RANSAC and MSAC approaches	Proposes a method for feature detection and matching using SURF and hybridized RANSAC and MSAC approaches for reliable estimation of interframe motion in video stabilization

PROPOSED SYSTEM

When a platform is in motion, cameras mounted on it experience unwanted shaking and blurring, which can result in unstable video footage. To overcome this problem, digital video stabilization is used to remove undesired image movements and produce a stabilized video sequence that only contains smooth global motions. Figure 1 describes the block diagram of a best real-time video stabilisation technique utilizing GA and WOA-CNN.

The proposed technique for video stabilisation is built on GA feature extraction which replaces the SURF and is easier to compute than SURF technique. GA consists of two main steps: identifying interest points and describing those points. In our methodology, the features between subsequent frames are first computed, and then the camera motion is estimated by fitting resulted features to a simplified linear motion model using WOA-CNN method. Using a regularization technique, the computed camera movements are then temporally smoothed to lessen the motion vibrations. Finally, stabilisation is achieved in the video by transforming each frame depending on original and smoothed motions. The effectiveness and stability performance of the suggested approach are demonstrated by experimental data.

A) VIDEO STABILIZATION

The 360° cameras, which record a whole viewing sphere are considered to be advantageous. While stabilizing a video through random frame warping, it is not necessary to worry about frame clipping, which is a common occurrence when using a narrow field of view in recordings. Whatever the case, there must be a short term fix for this severe issue. Consider a scenario where a videographer is moving forward with a 360 degree camera while keeping a straight sight line. At a crossroad, the cameraman takes a 90 ° right turn. A perspective of the camera varies unless stabilization has been achieved by distorting frame to the reference. Although, this is challenging as well as degrades the quality experience, a viewer may shift the viewpoint by turning it back to the front. Thus, a key component of resolving this problem is video stabilization technology.

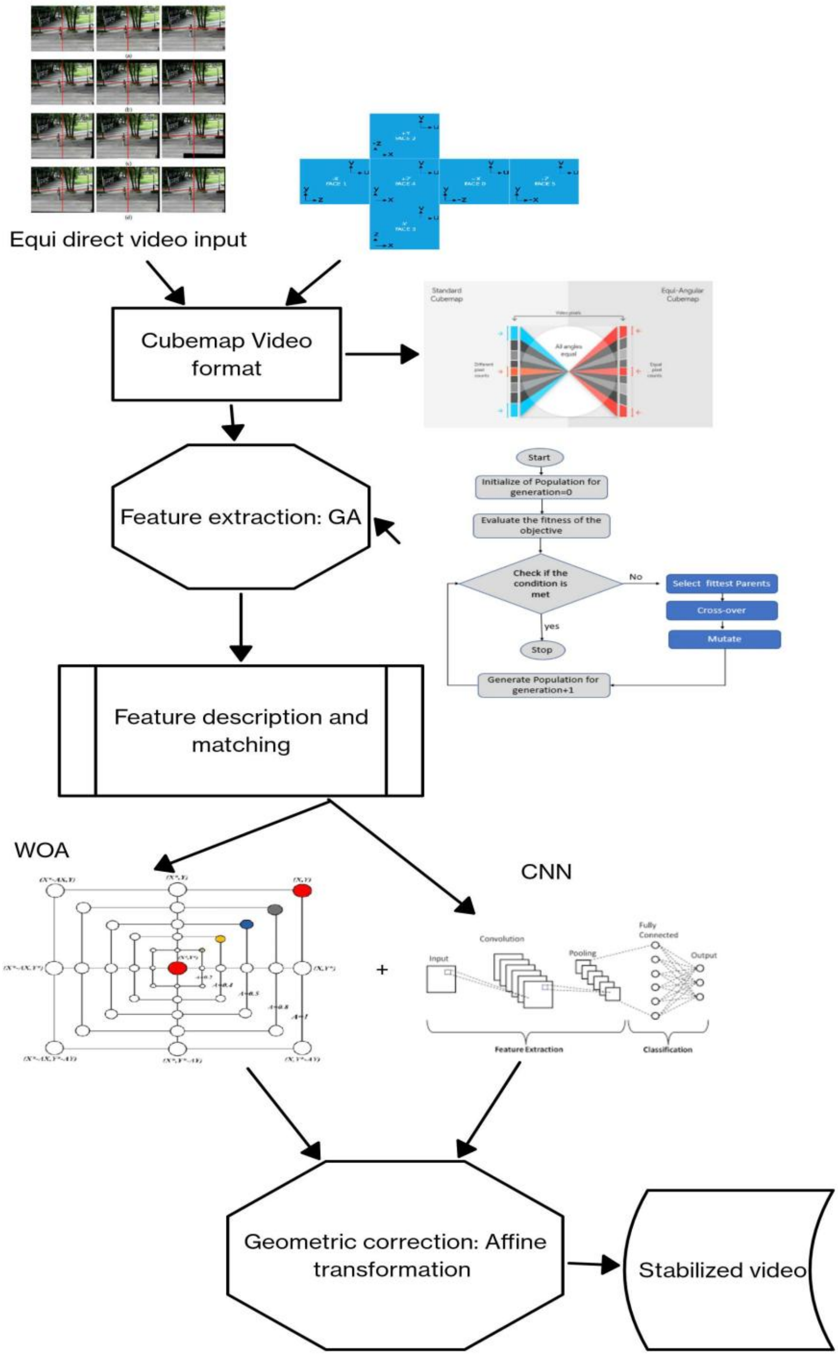

Figure 2: Block Diagram of Proposed System

B) CUBEMAP REPRESENTATION

In the environment mapping technique called "cube mapping," the six faces of a cube are employed as the map shape. The world is then unfolded into six parts of a single texture or mapped onto the sides of a cube and saved as six square textures.The scene is initially shown six times from each viewpoint, with each view specified by a 90-degree view frustum that represents one cube face. This creates the cube map.

Videos recorded in 360 ° are often depicted in an equirect format. While being an excellent format for viewing, computer vision systems consider it difficult for processing. As a result of subsequent processing it has been transformed into cube map format. The six sides of a unit cube are projected with a viewing sphere for generating the image. A unit focal length pinhole camera has been positioned in the center of the cube and every face is represented by an image captured by such camera. Face images, which follow the standard unipolar geometry have been utilized most frequently in computer vision methods for determining camera pose (Figure 2).

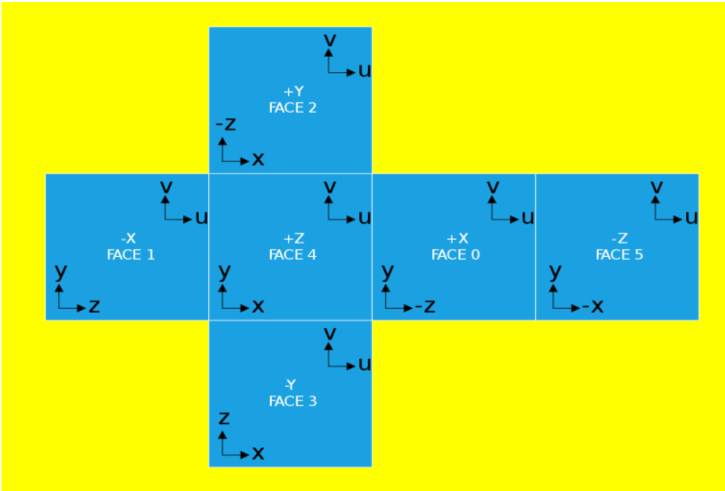

Figure 3. Cube map representation from equidirect inputs

C) FEATURE EXTRACTION USING GENETIC ALGORITHM

Holland first introduced the principles of GAs, which imitate specific natural processes of evolution. Considering GAs, the characteristics that describe an individual are frequently binary coded as well as combined for creating a string. In the solution space, the string packaging generated from various bit combinations has been regarded to as a point. The basic functions of a GA (Figure 3) are described as follows.

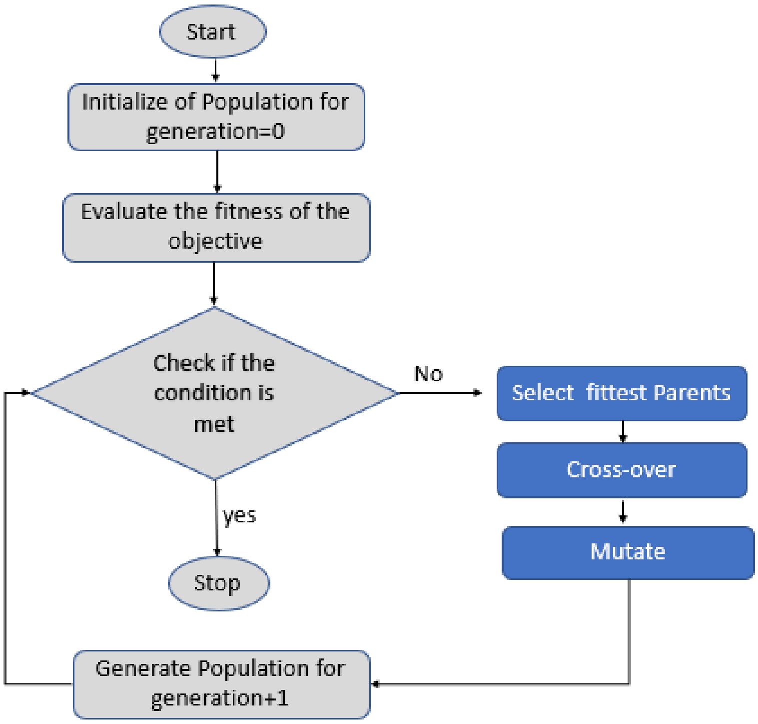

Figure 3(a). Genetic algorithm schema for video stabilization

Initialization: The genetic algorithm (GA) operates on a population of strings instead of a single one, and these strings evolve by generating new members that replace their predecessors. In the context of the given application, each string encodes a set of PSS parameters.

Objective function: The "fitness" of each structure is evaluated to determine its performance, and it is a figure of merit that must be minimized and is always non-negative.

Genetic operations: The GA transforms the prior population into a novel as well enhanced one by evaluating each person's fitness function. The following are the operations, which are most frequently used.

- **Reproduction:** An old string has been placed into a "mating pool" in accordance with its fitness in this process. The subsequent generation produces additional copies of strings, which are more highly fitted.
- **Crossover:** By exchanging genetic data between two parent chromosomes, crossover enables the combination of advantageous genes from both parents in the offspring.
- **Mutation:** Through the modification of certain chromosomes in accordance with a probabilistic law, it is a method, which can produce novel genetic material in the population.

This has been addressed by implementing dynamic weights to the representative points. It appears to be sense that actual representative points, as opposed to forged ones, will frequently become visible in images captured from the training set.

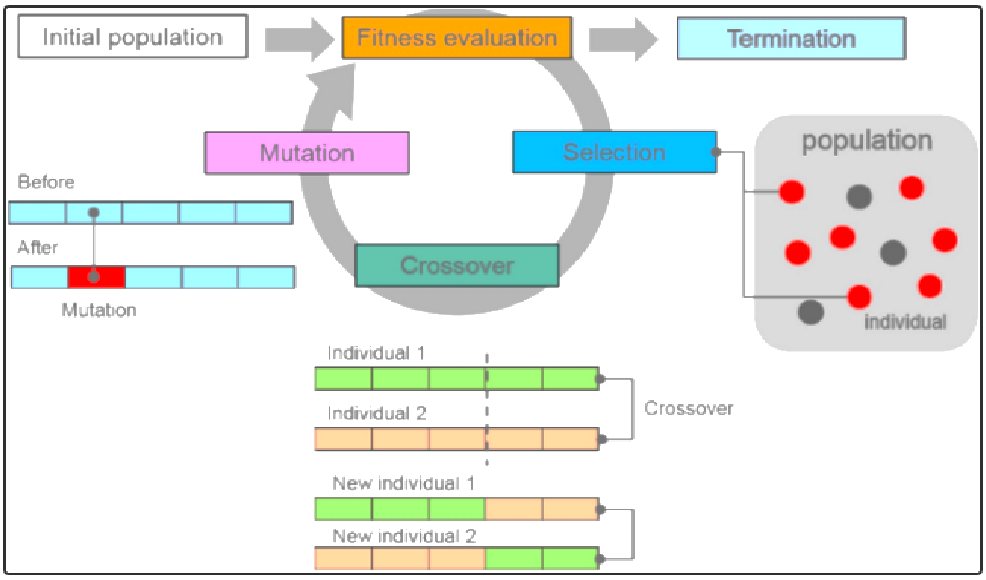

Figure 3(b). Basic steps in Genetic algorithm schema for video stabilization

For extracting the frames, the input video file has to be initially processed. Each pair of succeeding frame comparisons requires a considerable amount of time. The execution time has been reduced by selecting a first frame, skipping the next and repeating this process for each and every frame in the video file. As a result, this technique is utilized for processing half of the video files alone. Consequently, matching and removing the features require less time. Table 1 provides a clear explanation that F1, F2, F3, and F4 are similar but separate from F5, F6, F7, and F8, which suggests that the following shot frames are different from the initial shot frame. Moreover, F5 has been chosen as the frame for the sudden change, leading to the rejection of F2, F4, F6, and F8. Only F1, F3, F5, and F7 are retained for further processing. As a result, F1 has been contrasted with F3, which is subsequently contrasted with F5 and so on.

Table 1. Frame selection

Shot 1				**Shot 2**			
$F1$	$F2$	$F3$	$F4$	$F5$	$F6$	$F7$	$F8$
$F1$		$F3$		$F5$		$F7$	
$Ft(1)$		$Ft(2)$		$Ft(3)$		$Ft(4)$	

The extracted frames are resized to N×N and then transformed into grayscale images. The GA feature descriptor is then utilized to estimate the feature vector (Ft). Initially, key points are identified and subsequently characterized into a feature vector during the second stage. The descriptor features matrix (Ft) in this instance has a size of (p64), whereas (p) is a sum of key points identified in every frame and the length of feature vector for each key point is 64. A notion the GA features remain scaled invariable that indicates changing the picture size has no effect on them, which is considered as a significant aspect.

D) FEATURE DETECTION AND MATCHING

After extracting the features, feature matching has been performed, which compares the feature matrices $Ft(i)$ as well as $Ft(i+1)$. Based on the matching of feature matrices, the distance function is implemented for identifying commonalities among matrices. An expression for Hamming distance has been utilized to compare the corresponding feature vectors of feature matrices $Ft(i+1)$ and $Ft(i)$. Adjacent feature vectors are considered comparable if their distance from each other is less than a specified threshold value. Moreover, the feature vectors are not taken into account when the distance is greater than the threshold value. Few of matching features (MF) between both 2 feature matrices are represented as a vector $\mathrm{p} \times 1$ in the output. Considering that just 50% of video frames are often used for extracting the features. Therefore, the exact index for an abrupt change is determined by multiplying a sudden transition index by two.

E) WOA-CNN APPROACH

The extracted features are stabilized with the utilization of WOA-CNN technique as shown in Figure 2. WOA has been implemented for minimizing redundant as well as unwanted features during feature optimization. Two steps are involved in the optimization process. The prey is circling in the first stage, which also involves updating the spiral location. A random search for prey has been carried out in the second phase. The proposed process is mathematically modelled as follows.

a) Encircling prey

Whales track down their prey's position and encircle it. The position of the prey is unidentified in the search area. According to WOA, the leading component is an ideal prey. By shifting their locations, the existing search agents attempt becoming an optimum possible agents. The following behaviours of search agents are described as,

$$\vec{Y}(u+1) = \vec{Y}^*(u) - \vec{B}.\vec{E}, \quad (1)$$

$$\vec{E} = \left|\vec{C}.\vec{Y}^*(u) - \vec{Y}(u)\right|, \quad (2)$$

Whereby a whale's optimal position following iteration is defined as $\vec{Y} * (u)$. The whale's current location is indicated as $\overrightarrow{Y}(u+1)$ as well as a distance vector $\vec{E}$ represents the separation between both the whale and its prey. An absolute value has been denoted using | |.The equations for coefficient vectors $\vec{B}$ and $\vec{C}$ are as follows:

$$\vec{B} = 2.\vec{b}.\vec{s} + \vec{b} \quad (3)$$

$$\vec{C} = 2.\vec{s} \quad (4)$$

For applying to shrink,$\vec{b}$’ value has been minimized and B's fluctuating range is decreased to b. The value of $\vec{B}$changes from $(-b, b)$ as well as the number of iterations for b has been decreased from 2 to 0. By choosing a random value of $\vec{B}$ among (-1, 1), the position of the finest agent and also the beginning location of an agent are determined.

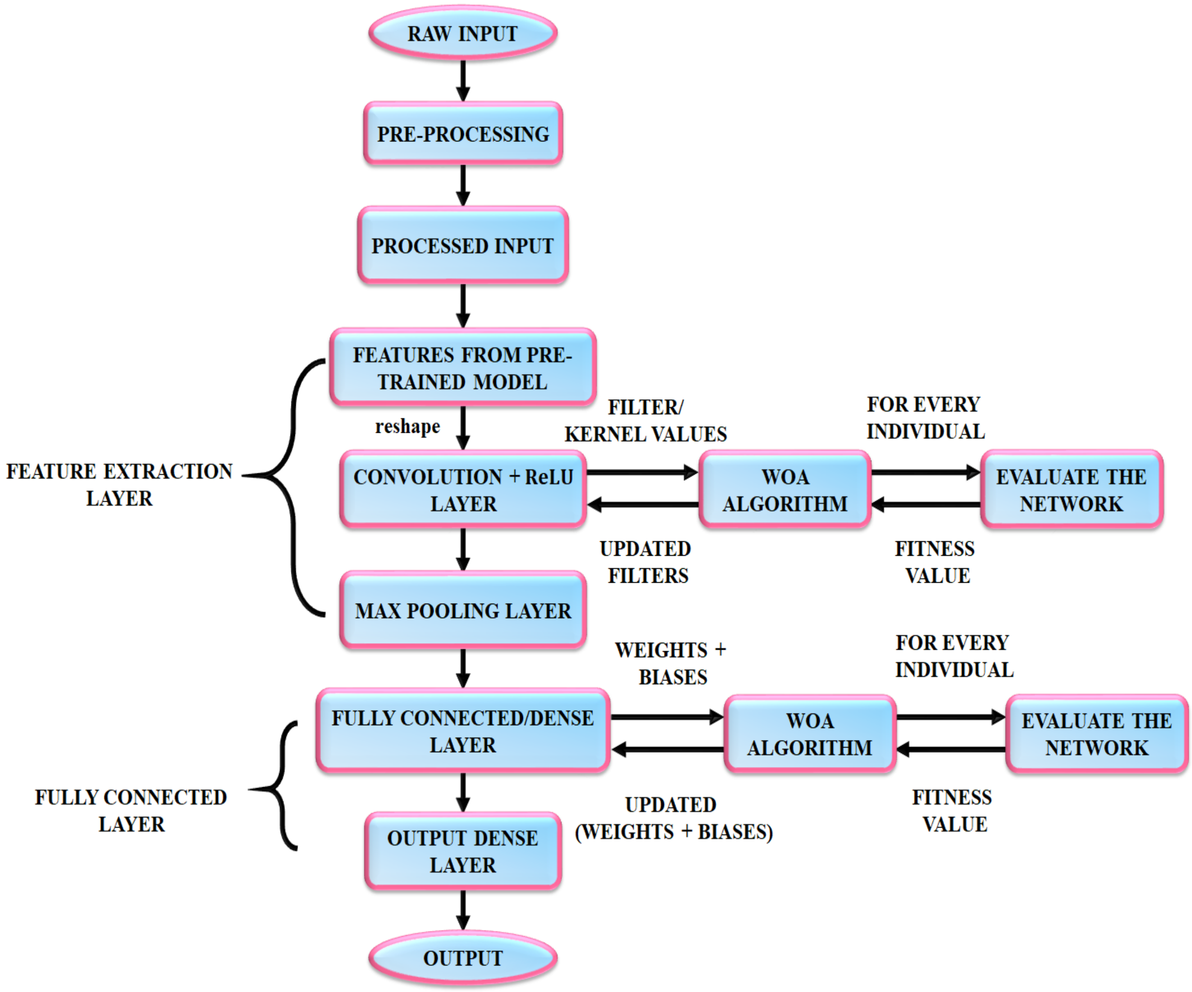

Figure 4: WOA-CNN Architecture

b) Spiral position updating

The gap between both the whale's position (Y, Z) along with the location of prey (Y^*, Z^*) has been used for determining the helix formation for whale prey tracking. The forward motion approaching prey has been defined as,

$$\vec{Y}(u+1) = e^{bk}.COS(2\pi k).\vec{E}^* + \vec{E}^*(u) \quad (5)$$

$$\vec{E}^* = \left|\vec{Y}^*(u) - \vec{Y}(u)\right| \quad (6)$$

Constant b determines the logarithmic spiral's form and the range of possible random numbers has been depicted as [-1, 1]. Whales migrate in a spiral motion, which allows them to change locations while performing reductions. Therefore, spiral as well as shrinking encircling have a 50% probability of being selected.

$$\vec{Y}(u+1) = \begin{cases} \vec{Y}^* - \vec{B}.\vec{E} & if\ p < 0.5, \\ e^{bk}.COS(2\pi k).\vec{E}^* + \vec{Y}^*(u), & if\ p \geq 0.5, \end{cases} \quad (7)$$

Here, 0 to 1 indicates a range of the random number p.

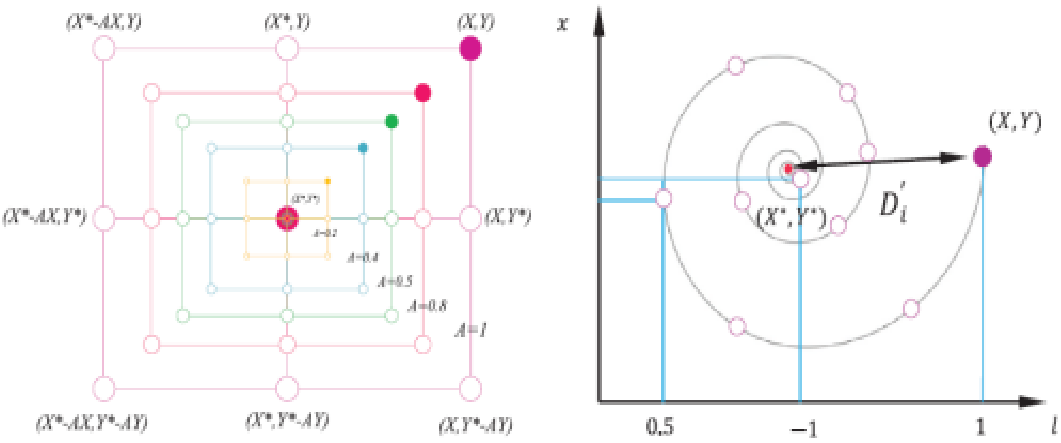

Figure 4(a). Bubble-net hunting attack model consists of two stages: shrinking encircling mechanism (left) and spiral updating position (right), chosen in semi-random order [8]

c) Prey search

An exploration phase that is reliant on variation of the vector $\vec{B}$ is also referred as the prey searching process. In terms of finding its prey based on the location, the whale undertakes a random search. The search agent leaves the whale which is constantly searching owing to the whale's location. WOA makes use of the vector B, which has random values either less than 1 or higher than 1. The search agent is randomly selected for beginning the exploration process. Through the reduction of local optimization problem, random selection converts WOA into a global search agent. Parameters of the global search are:

$$\vec{Y}(u+1) = \overrightarrow{Y_{rand}} - \vec{B}.\vec{E}, \quad (8)$$

$$\vec{E} = \left|\vec{C}.\overrightarrow{Y_{rand}} - \vec{Y}\right|, \quad (9)$$

Here, a random whale chosen from the population selected has been represented as $\overrightarrow{Y_{rand}}$. The WOA assumes the optimum approach of the function using minimum or maximum value as well as initializes the algorithm via distributing random results among the whale population.

The fitness function has been employed for examining the features chosen for the given phase, while it is discovered that a small number among them are nevertheless redundant as well as influence the final classification's accuracy. Each iteration's difference is calculated using the ensemble subspace discriminant (ESD) classifier as a fitness function. Consequently, a new phase with the term extra features approval (EFA) has been established. The standard error of the mean (SEM) provides the foundation for the presented phase. The threshold function is applied to the SEM value to achieve the ultimate recognition.Formulation of the suggested function is expressed as follows,

$$SEM(y_i) = \frac{\sigma}{N}, \sigma = \sqrt{\sigma^2}, \sigma^2 = E[(y_i - \mu)^2] \quad (10)$$

$$Selection = F(i) = \begin{cases} F_i^s & for\ y_i \geq SEM(y_i) \\ Remove, & Elsewhere \end{cases} \quad (11)$$

For final identification, an ensemble subspace classifier receives the features which were finally selected (F_i^s). The whale optimization algorithm utilized in the proposed work is tabulated as follows.

Algorithm: Whale Optimization Algorithm
Initial population Y_i where ($i = 1, 2, 3,...$) Fitness calculation for each solution Y^* = best search agent **While** ($u < Max\_iteration$) **For every solution** Updated b, B, C, M and p **If 1** ($p < 0.5$) **If 2** ($/B/ < 1$) Update the current search agent location with respect to equation 1. **Else if 2**($/B/ > 1$) Random search agent selection (Y_{rand}) Current search agent location changes according to equation 7. **End if 2** **Else if 1**($p \geq 0.5$) Current search agent location changes according to equation 5. **End if 1** **End** Inspect movement of search agent if search agent goes away from search location and changes it. Change Y^* if the better solution is presented $u = u + 1$ **End While** Return Y^* Refine using Eq. (10)-(11) Output: $F_i^s \leftarrow$ Best Feature Vector

By the whale's hunting behaviour, it ultimately satisfies the most stringent CNN criteria. The predictive algorithm is qualified as a consequence of locating the best fitness function or optimal solution. The prediction model generated for best fitness structure is highly suited to predict unknown data because the target function is to increase the accuracy of training data. The most intriguing aspect of suggested technique is that the CNN structure extracts an image's characteristics locally, allowing the network to learn particular patterns inside image and recognise them anywhere in image. The process will be continued until image has been scanned. Thus, effective stabilization is achieved.

The CNN architecture is optimized using the WOA-CNN technique, which stabilizes the extracted features by minimizing redundant and unwanted features. The optimization process involves two steps: encircling prey and spiral position updating. The prey search involves variation of the coefficient vector, and a random search is undertaken by the whale to find its prey. The fitness function is used to examine the chosen features, and a new phase called extra features approval (EFA) has been established to improve the final classification's accuracy. The architecture is mathematically modeled using equations, and the CNN architecture is shown in Figure 5.

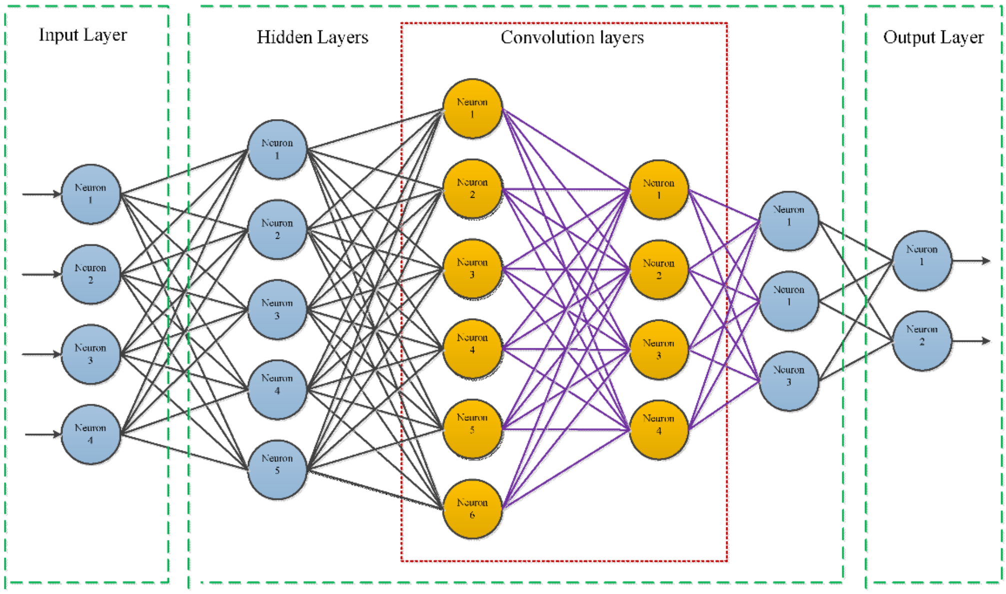

Figure 5. Details of the convolution neural network architectures network with convolution layers.

In the WOA approach, the CNN architecture is designed to perform four main tasks: the input layer pre-processes the raw data, the convolutional layers extract features from the data, the pooling layers down sample the features, and the fully connected layers generate a prediction for the output.

The input layer receives the raw data and pre-processes it to ensure that it is in a suitable format for the subsequent layers. This involves normalizing the data or applying other pre-processing steps. The convolutional layers use filters to extract features from the input data. The filters perform a convolution operation by sliding over the data and computing a dot product at each location. The result is a feature map that represents the presence or absence of a particular feature at each location in the input. The convolutional layers learn to extract relevant features from the input data, such as edges, corners, or textures. The pooling layers down sample the feature maps produced by the convolutional layers. This reduces the dimensionality of the feature maps and makes the subsequent layers more efficient. The pooling layers may also help to prevent overfitting by reducing the no. of parameters in the model. The fully connected layers connect each neuron in the previous layer to each neuron in the current layer. The purpose of this layer is to combine the features learned by the previous layers and generate a prediction for the output. The fully connected layers learn to map the high-level features extracted by the convolutional layers to the output classes or regression values.

By optimizing the weights and biases in each layer using a WOA algorithm, the CNN can learn to accurately classify or predict new data. Overall, the WOA approach to CNN architecture involves a series of layers that work together to pre-process, extract, down sample, and generate a prediction for the input data.

RESULTS AND DISCUSSION

This study makes extensive use of Genetic Algorithm feature extraction approach for video stabilisation in order to follow the precise motion and gesturing of the objects in videos with greatest possible clarity. Additionally, improved CNN technique with the utilisation of WOA approach provides highly reliable inlier and outlier estimation. The entire project is assessed using MATLAB Simulink, and the results are clearly discussed in this section in order to best validate the proposed methodology.

Figure 6: Video Frame in Equirect Format

The image of input video frame in equirect format, which effectively depicts both the 180° vertical perspective and the 360° horizontal view of input frame is shown in Figure 6. Since it comprises of illumination data from all directions, it is very likely that using this equirect format, the data is viewed as points on sphere.

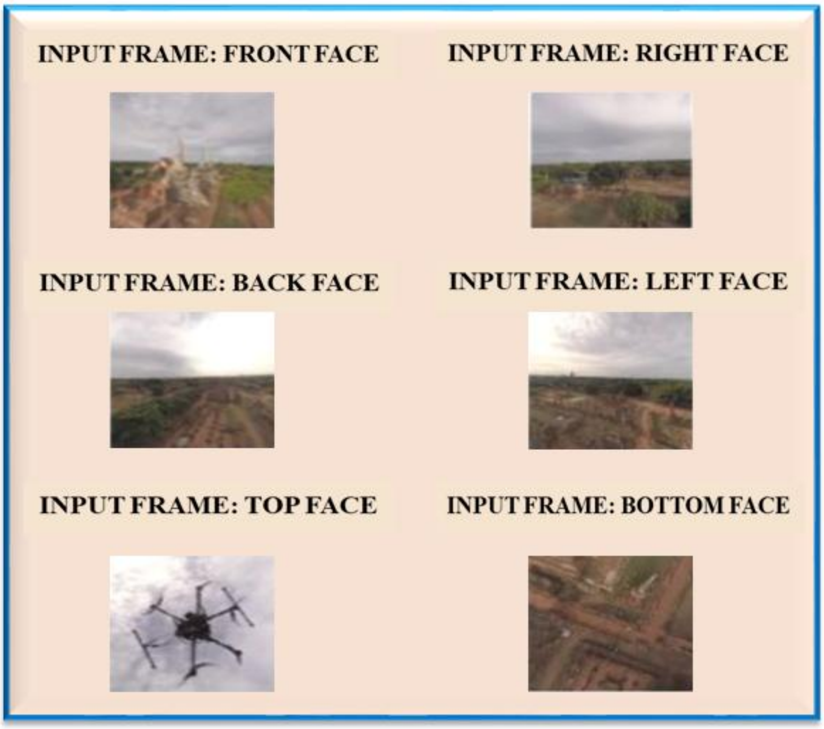

Figure 7: Representation of Cube maps

Figure 7 shows how cube maps for input video frames can create steady specular highlights in a relatively simple and effective way. A cube map texture which stores several specular highlights, is accessed by interpolating across surface's reflection vector to provide coordinates.

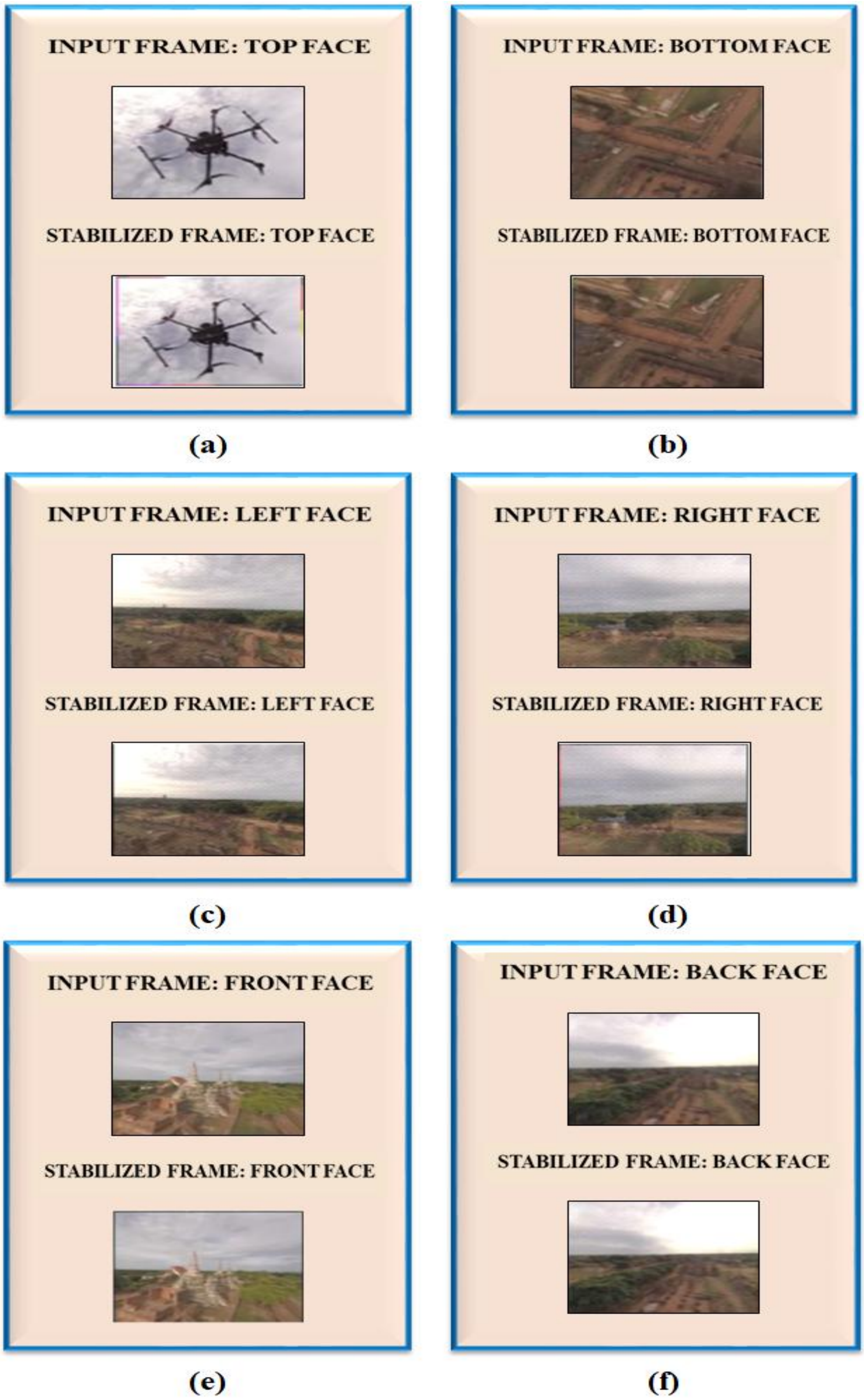

Figure 8: Stabilized Frame (a) Top Face, (b) Bottom Face (c) Left Face (d) Right Face (e) Front Face and (f) Back Face

In this work, video stabilisation is successfully accomplished with the use of GA-based feature extraction to remove unwanted camera disturbance from a video sequence. Figure 8 (a-f) shows the input frame and stabilised frame for each of six faces separately. On validating that the introduced approach using WOA-CNN algorithm effectively stabilises video frame because the inliers and outliers are accurately and significantly analysed without any complications. For the purpose of validating the introduced technique, the performance of WOA-CNN technique with GA during the feature extraction process is compared with the SIFT and SURF based extraction strategy. The overall computation time in accordance with detection time, matching time along with total time is depicted in Table 2.

Table 2: Computation Time Comparison

Image		1and 2	2 and 3	3 and 4	4 and 5	5 and 6	6 and 7	7 and 8
SIFT	Detection Time	0.142	0.144	0.157	0.168	0.173	0.164	0.156
	Matching Time	0.006	0.004	0.005	0.004	0.005	0.004	0.005
	Total Time (s)	0.148	0.148	0.162	0.172	0.178	0.168	0.160
SURF	Detection Time	0.021	0.020	0.033	0.035	0.021	0.018	0.022
	Matching Time	0.005	0.005	0.007	0.004	0.007	0.004	0.004
	Total Time (s)	0.026	0.024	0.039	0.0039	0.026	0.022	0.025
GA	Detection Time	0.020	0.019	0.032	0.022	0.021	0.017	0.021
	Matching Time	0.004	0.002	0.003	0.002	0.001	0.002	0.002
	Total Time (s)	0.024	0.021	0.035	0.024	0.022	0.019	0.023

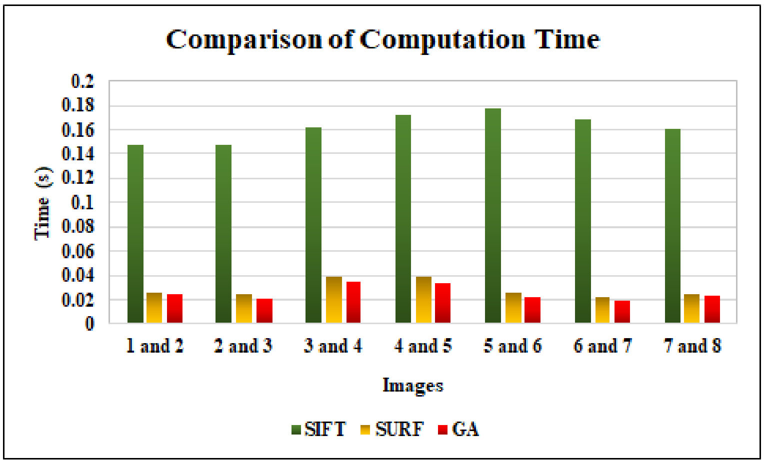

Figure 9: Computation Time Comparison

The analysis of computation time for approaches like SIFT, SURF and GA are compared in Figure 9. From the figure, it is noticed that the proposed GA results in faster computation time in comparison with others.

Table 3: Comparison of Matching Pairs

Image	SIFT	SURF	GA
1 and 2	37	40	42
2 and 3	40	43	45
3 and 4	39	43	44
4 and 5	40	40	41
5 and 6	38	39	42
6 and 7	40	43	46
7 and 8	42	38	43

The matching pair comparison for approaches like SIFT, SURF and GA is specified in Table 3. From the table, the proposed Genetic Algorithm shows improved matching pair results. The corresponding Graphical representation is illustrated in Figure 10.

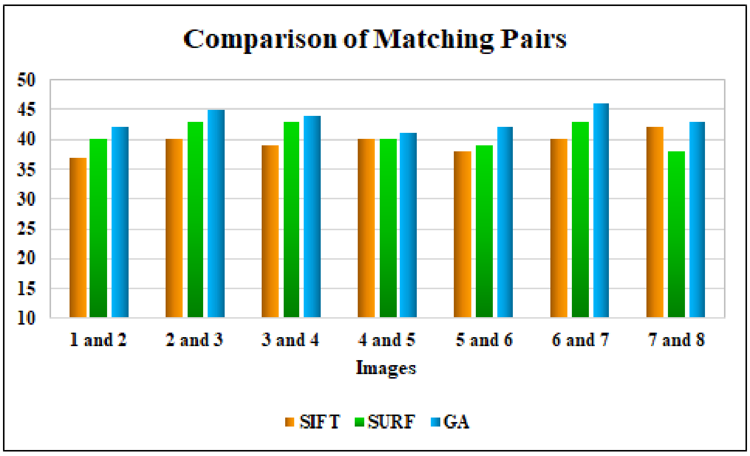

Figure 10: Graphical Representation of Matching Pair Comparison

The Accuracy obtained in terms of Rotation, Scale, Blur, Illumination, Wrap and Time cost is analysed and tabulated in Table 4, for approaches like SIFT, SURF and GA.

Table 4: Accuracy Specification

Approach	Accuracy (%)					
	Rotation	Scale	Blur	Illumination	Warp	Time Cost
SIFT	91.2	91.8	93.2	94.4	94.6	94.9
SURF	93.4	94.1	94.9	95.2	95.8	96.8
GA	94.2	94.8	95.1	95.4	96.3	97.5

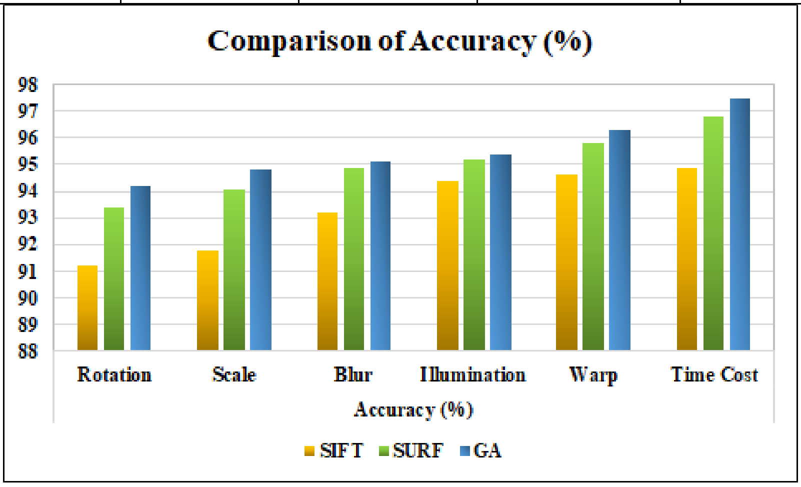

Figure 11: Comparison of Accuracy (%)

The graphical representation of Accuracy comparison is illustrated in Figure 11. It is observed that the proposed approach shows improved accuracy in all terms while compared to approaches like SIFT and SURF.

CONCLUSION

This research suggests an innovative real-time high-quality video stabilisation technique for stabilising videos, which is more difficult than usual videos acquired by handheld camera systems due to huge and more foreground objects. The present work provides the comprehensive analysis for validating excellence of video processing through stabilisation utilising GA-based feature extraction since it has many meritorious effects in improving the video quality throughout a larger range. Utilizing a genetic algorithm to extract features entails effectively enhancing the dependability of video frame, which enhances the system's total performance capacity.

Meanwhile, the WOA-CNN algorithm is impressively used to estimate Inliers and Outliers in order to acquire the interruption-free recognition of gestures and movement of objects in video without any difficulties. Future scope will be integrating advance deep learning models to improve the efficacy and analysing with other various measures and also use of transfer learning to improve the understanding knowledge of model to improve the training and testing efficacy.

REFERENCES

[1] Z. Wang, L. Zhang and H. Huang, "High-Quality Real-Time Video Stabilization Using Trajectory Smoothing and Mesh-Based Warping," in IEEE Access, vol. 6, pp. 25157-25166, 2018.

[2] M. Okade, G. Patel and P. K. Biswas, "Robust Learning-Based Camera Motion Characterization Scheme With Applications to Video Stabilization," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 3, pp. 453-466, March 2016.

[3] C. Tang, O. Wang, F. Liu and P. Tan, "Joint stabilization and direction of 360° videos", ACM Transactions on Graphics (TOG), Vol.38, no.2, pp.1-13, 2019.

[4] J. Kopf, "360 video stabilization", ACM Transactions on Graphics (TOG), Vol. 35, no.6, pp.1-9, 2016.

[5] M. Zhao and Q. Ling, "A Robust Traffic Video Stabilization Method Assisted by Foreground Feature Trajectories," in IEEE Access, vol. 7, pp. 42921-42933, 2019.

[6] H. Huang, X. -X. Wei and L. Zhang, "Encoding Shaky Videos by Integrating Efficient Video Stabilization," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 29, no. 5, pp. 1503-1514, May 2019.

[7] M. Zhao, S. Deng and Q. Ling, "A Fast Traffic Video Stabilization Method Based on Trajectory Derivatives," in IEEE Access, vol. 7, pp. 13422-13432, 2019.

[8] H. Jiang, G. Jiang, M. Yu, Y. Zhang, Y. Yang et al., "Cube map-based perception-driven blind quality assessment for 360-degree images", IEEE Transactions on Image Processing, vol. 30, pp. 2364–2377, 2021.

[9] X. Chen, A. T. Z. Kasgari and W. Saad, "Deep learning for content-based personalized viewport prediction of 360-degree VR videos", IEEE Networking Letters, vol. 2, no. 2, pp. 81 – 84, 2020.

[10] G. de A. Roberto, N. Birkbeck, I. Janatra, B. Adsumilli and P. Frossard, Multi-feature 360 video quality estimation", IEEE Open Journal of Circuits and Systems, vol. 2, pp. 338–349, 2021.

[11] A. Walha, A. Wali and A. M. Alimi, "Video stabilization with moving object detecting and tracking for aerial video surveillance", Multimedia Tools and Applications, Vol. 74, no. 17, pp.6745-6767, 2015.

[12] H. Pedrini,"Combination of local feature detection methods for digital video stabilization. Signal", Image and Video Processing, Vol. 12, no. 8, pp.1513-1521, 2018.

[13] D. Mistry and A. Banerjee, "Comparison of feature detection and matching approaches: SIFT and SURF", GRD Journals-Global Research and Development Journal for Engineering, vol. 2, no. 4, pp. 7–13, 2017.

[14] W. Zhang, X. Li, J. Yu, M. Kumar, and Y. Mao, "Remote sensing image mosaic technology based on SURF algorithm in agriculture", EURASIP Journal on Image and Video Processing, Vol. 2018, no.1, pp.1-9, 2018.

[15] S. Zhang et al., "CASIA-SURF: A Large-Scale Multi-Modal Benchmark for Face Anti-Spoofing," in IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 2, no. 2, pp. 182-193, April 2020.

[16] M. Wang et al., "Deep Online Video Stabilization With Multi-Grid Warping Transformation Learning," in IEEE Transactions on Image Processing, vol. 28, no. 5, pp. 2283-2292, May 2019.

[17] B. Kim, Y. S. Hariyani, Y. H. Cho and C. Park, "Automated white blood cell counting in nailfold capillary using deep learning segmentation and video stabilization", Sensors, Vol. 20, no. 24, p.7101, 2020.

[18] S. Dervisoglu, M. Sarıguland L. Karacan, "Interpolation-Based Smart Video Stabilization" Akıllı Sistemler ve Uygulamaları Dergisi, Vol. 4, no. 2, pp.153-156, 2021.

[19] M. Zhao and Q. Ling, "PWStableNet: Learning Pixel-Wise Warping Maps for Video Stabilization," in IEEE Transactions on Image Processing, vol. 29, pp. 3582-3595, 2020.

[20] K. M. Lee, and C. H. Lin, "Video stabilization algorithm of shaking image using deep learning", The Journal of the Institute of Internet, Broadcasting and Communication, Vol. 19, no. 1, pp.145-152, 2019.

[21] Xu, Y., Zhang, J., Maybank, S. J., & Tao, D. (2022). DUT: learning video stabilization by simply watching unstable videos. IEEE Transactions on Image Processing, 31, 4306-4320.

[22] Zhu, A., Lu, H., & Cao, Z. (2022). Video stabilization based on image registration. Journal of Electronic Imaging, 31(2), 023007-023007.

[23] Chen, C. H., Lin, C. E., Chen, T. Y., Wang, D. J., Liao, C. F., & Wen, C. K. (2022, July). The Cost-Effective Video Stabilization Method for Wearable Camera. In 2022 IEEE International Conference on Consumer Electronics-Taiwan (pp. 327-328). IEEE.

[24] Liu, S., Ji, Z., He, Y., Lu, J., Lan, G., Cong, J., ... & Gu, B. (2022). Deep-Learning Image Stabilization for Adaptive Optics Ophthalmoscopy. Information, 13(11), 531.

[25] Mudassar, B. A., Ko, S., Li, M., Saha, P., & Mukhopadhyay, S. (2022). RADNet: A Deep Neural Network Model for Robust Perception in Moving Autonomous Systems. arXiv preprint arXiv:2205.00364.

[26] Singh, Ritesh, Kalgesh Singh, and Amir Ansari. "Applying approximate computing principles for real-time rolling shake hyper-stabilization during video recording." Design Automation Conference. 2022.

[27] Dervişoğlu, Semiha, Mehmet Sarıgül, and Levent Karacan. "Interpolation-based smart video stabilization." Akıllı Sistemler ve Uygulamaları Dergisi 4.2 (2021): 153-156.

[28] Yang, Chenyi, Yuqing He, and Danfeng Zhang. "LSTM based video stabilization for object tracking." AOPC 2021: Optical Sensing and Imaging Technology. Vol. 12065. SPIE, 2021.

[29] Luchetti, A., et al. "Stabilization of spherical videos based on feature uncertainty." The Visual Computer (2022): 1-14.

[30] Afsal, S., and Arul Linsely. "Optimal Process of Video Stabilization Using Hybrid RANSAC-MSAC Algorithm." International Journal of Intelligent Systems and Applications in Engineering 11.2 (2023): 564-571.

AUTHORS PROFILE

Afsal S., is currently working as Assistant Professor at ACE College of Engineering, Trivandrum, Kerala, India. and pursing PhD in Image Processing from Noorul Islam Centre for Higher Education, Thuckalay, Tamilnadu,India. He completed his M.Tech in Signal Processing from Govt. Engineering College, BartonHill, Trivandrum, Kerala, India .

Dr. Arul Linsely J, is currently working as Associate Professor of EEE department in Noorul Islam Centre for Higher Education, Thuckalay, Tamilnadu,India. He completed his PhD from Noorul Islam Centre for Higher Education, Thuckalay, Tamilnadu,India.