

GMM-based Imbalanced Fractional Whale Particle Filter for Multiple Object Tracking in Surveillance Videos

Avinash Ratre

Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, 110042, India
E-mail: avinashratre@dtu.ac.in
ORCID iD: <https://orcid.org/0000-0001-9803-2665>

Received: 19 December 2024; Revised: 11 February 2025; Accepted: 04 March 2025; Published: 08 April 2025

Abstract: The imbalanced surveillance video dataset consists of majority and minority classes as normal and anomalous instances in the nonlinear and non-Gaussian framework. The normal and anomalous instances cause majority and minority samples or particles associated with high and low probable regions when considering the standard particle filter. The minority particles tend to be at high risk of being suppressed by the majority particles, as the proposal probability density function (pdf) encourages the highly probable regions of the input data space to remain a biased distribution. The standard particle filter-based tracker afflicts with sample degeneration and sample impoverishment due to the biased proposal pdf ignoring the minority particles. The difficulty in designing the correct proposal pdf prevents particle filter-based tracking in the imbalanced video data. The existing methods do not discuss the imbalanced nature of particle filter-based tracking. To alleviate this problem and tracking challenges, this paper proposes a novel fractional whale particle filter (FWPF) that fuses the fractional calculus-based whale optimization algorithm (FWOA) and the standard particle filter under weighted sum rule fusion. Integrating the FWPF with an iterative Gaussian mixture model (GMM) with unbiased sample variance and sample mean allows the proposal pdf to be adaptive to the imbalanced video data. The adaptive proposal pdf leads the FWPF to a minimum variance unbiased estimator for effectively detecting and tracking multiple objects in the imbalanced video data. The fractional calculus up to the first four terms makes the FWOA a local and global search operator with inherent memory property. The fractional calculus in the FWOA oversamples minority particles to be diversified with multiple imputations to eliminate data distortion with low bias and low variance. The proposed FWPF presents a novel imbalance evaluation metric, tracking distance correlation for the imbalanced tracking over UCSD surveillance video data and shows greater efficacy in mitigating the effects of the imbalanced nature of video data compared to other existing methods. The proposed method also outshines the existing methods regarding precision and accuracy in tracking multiple objects. The consistent tracking distance correlation near zero values provides efficient imbalance reduction through bias-variance correction compared to the existing methods.

Index Terms: Multiple Object Tracking, Particle Filter, Fractional Whale Optimization, Gaussian Mixture Model, Imbalanced Surveillance Video Data.

1. Introduction

Visual tracking is a critical task in vision-based applications in which the actions of humans are identified, and the human-computer interfaces are developed. Stochastic point tracking methods employ a state-space approach to model the linear or non-linear dynamics of the objects using the Kalman filter or particle filter. The Kalman filter fails when vision tracking problems are non-linear, non-Gaussian, and multi-modal. The PF solves such vision problems, which rely on the recursive construction of the state space's posterior probability density function (pdf). A set of random particles is produced to estimate the posterior pdf of the states of the tracked objects. In estimating the states of objects, these sets of particles are updated and propagated based on sequentially arriving observations. The particle filter better estimates hidden state variables than the other non-linear filters [1, 2]. However, due to the resampling procedure, the particle filter suffers from the problem of sample degeneration and impoverishment propagating in state space [3]. The resampling technique involves replacing the old set of particles with related weights with the new set of particles with related weights in more probable regions of state space when sample degeneration increases above a threshold. The resulting sample impoverishment pertains to the frequent selection of particles with high weights and the dying out of other particles with negligible weights when particles propagate after each time step [4, 5].

The imbalanced surveillance video dataset consists of normal and anomalous instances. The normal and anomalous instances cause majority and minority samples associated with high and low probable regions. In the standard particle filter, the minority particles tend to be at high risk of being suppressed by the majority particles, as the biased proposal or importance pdf encourages only majority particles in the highly probable regions of the input data space. However, resampling tries to correct the bias to a small extent. The difficulty in designing the correct proposal or importance pdf prevents standard particle filter-based tracking in the imbalanced surveillance video data. Further, the problem of sample degeneration and sample impoverishment occurs due to the choice of a biased proposal or importance pdf ignoring the minority particles [6, 7].

Traditional tracking methods for imbalanced data suffer from the problem of occlusion and identity switching between objects. This problem worsens when the cost of tracking false negatives and false positives is the same. The resampling methods apply three actions for the imbalanced data: data, algorithmic, and hybrid-level. In this paper, multiple object tracking is based on hybrid-level actions on the majority and minority particles. This paper discusses converting the biased proposal or importance pdf into the unbiased proposal or importance pdf using hybrid-level action for the particle filter.

The authors [8] provided a theoretical justification for an evolutionary particle filter to prevent the problem of sample impoverishment without the application of visual tracking of the objects. The authors [9] proposed a whale optimization algorithm (WOA), which is a highly efficient swarm-based metaheuristic algorithm stimulated by the hunting behaviour of humpback whales (H-whales) and determine the best solution. The WOA avoids local optima and gets a global optimal solution. The WOA solves different constrained or unconstrained optimization problems for practical applications without reformulation in the algorithm. This algorithm is highly efficient compared to other metaheuristic algorithms and is found to help handle imbalanced data.

In keeping with the advantages of metaheuristic algorithms and fractional calculus in dealing with the problem of sample degeneration and impoverishment, this paper proposes a new fractional whale particle filter (FWPF) which fuses a novel fractional whale optimization algorithm (FWOA) [9] with the standard particle filter for multiple object tracking (MOT) in the imbalanced surveillance video data with tracking challenges. Multiple objects from the initial video frame are extracted through human intervention, and the particles and related weights are generated. Then, the particles are updated using the FWPF method. This operation is repeated until the best solution is determined. In this paper, contributions are as follows:

- Development of a complementary fractional whale particle filter (FWPF) that fuses the fractional calculus-based whale optimization algorithm and the standard particle filter under weighted sum rule fusion. Integrating the FWPF with an iterative Gaussian mixture model (GMM) with unbiased sample variance and sample mean allows the proposal or importance probability density function (pdf) to be adaptive to deal with the imbalanced nature of surveillance video data. The adaptive importance or proposal pdf leads the FWPF to a minimum variance unbiased estimator (MVUE) for effectively detecting and tracking multiple objects in the imbalanced surveillance video data. The novelty of the work is to convert the fixed biased proposal pdf into an unbiased adaptive proposal pdf to deal with imbalanced data.
- A novel imbalanced evaluation metric, tracking distance correlation (TDC), is employed to evaluate the bias and variance correction (imbalance reduction) caused by an unbiased proposal pdf in the imbalanced FWPF-based tracking over the biased proposal pdf in the standard particle filter-based tracking. Apart from TDC, other performance measures, such as multiple objects tracking precision and accuracy (MOTP and MOTA), the sample mean and variance are used for comparative evaluation with other existing methods in the presented work.

This paper is structured as follows: The second and third sections relate to the related works and the challenges, respectively. The fourth section recounts the mathematical formulations for the standard PF. The fifth section introduces the proposed FWPF for MOT in imbalanced surveillance video data. The sixth section details the results and discussion. Finally, the seventh section winds up the paper as the conclusion.

2. Related Works

In particle filter-based visual tracking, the object's states are unknown and can only be deduced from visual feature observations. Visual tracking has two main difficulties: the state hypothesis evaluation of the observed image and the search for the state space. Metaheuristic algorithm-based particle filter tries to solve such problems in visual tracking. The two main visual tracking procedures are bottom-up and top-down procedures. The former typically favors rebuilding the states of the objects by examining the objects in the video sequences. At the same time, the latter is used to build and examine a state hypothesis set based on object models[10]. The MOT has many tracking challenges, such as occlusion, confusion between objects, scale changes of objects tracked, tracking resumption, and variability in the number of objects. There are two steps in MOT: In the first step, the number and position of targets are inferred from the current frame at every iteration, and this step requires either a generative or discriminative model, whereas the second step refers to the

modelling of detection errors and the movement of the target [11]. The Markov-Chain-Monte-Carlo (MCMC) framework approximates the posterior pdf adaptively and estimates the best state of objects using the Markov chain. This method encounters the problems of sample degeneration and impoverishment while tracking multiple objects [12].

The authors [13] proposed an evolutionary particle filter using modified grey-wolf metaheuristics to resolve sample impoverishment. The method enhances exploitation and exploration properties in tracking. However, the technique had limited exploration of dynamic target object behaviors. In [14], authors proposed a chaotic local search model for irregular motion with fewer particles. A chaotic sampling procedure enhances particle distribution in high-likelihood areas. The method effectively handles occlusion and fast-motion challenges by introducing a histogram of dynamic information for motion representation. The method exhibits superior performance in comparison to leading-edge approaches in tracking accuracy. The method has limited exploration of chaotic optimization in higher dimensions. The authors [15] proposed the combination of quantum particle swarm and adaptive genetic metaheuristics to focus on interference factors affecting tracking efficiency and accuracy. The method effectively handles occlusion and fast motion. In [16], the authors proposed a Mean shift algorithm for particle prediction and modified particle swarm metaheuristics for detecting occlusions in dynamic environments. The anti-occlusion module in this method improves tracking performance during occlusions.

The authors [17] proposed the covariance matrix-based learning with biogeography-based metaheuristics for real-world object tracking utilizing suitability index variables and color histograms. In [18], the authors proposed mayfly position-weighting strategies and dynamic control for multi-target tracking. Additionally, chaotic optimization strategies enhanced velocity parameters in mayflies. This method required addressing new and faded targets. The authors [19] proposed an improved firefly algorithm for optimization for multitarget tracking addressing the particle filter's weight degradation issue. The method fell into premature convergence quickly. The authors [20] utilized convolutional features and superpixel segmentation to address challenges in target association during complex scenarios of intelligent transportation. The method lacks extensive evaluation of diverse datasets. In [21], the authors proposed an iterative particle filtering based on ant colony metaheuristics and locality-sensitive histogram models to address the target appearance under illumination variations. The method discussed an iterative unscented Kalman filter for proposal distribution. However, the technique does not explore real-time tracking applications.

3. Challenges

The most difficult problem in visual tracking is the occlusion of objects, in which the portions of occluded objects are only seen, and the tracking becomes very difficult after the occurrence of object occlusion. The entry and exit of multiple objects in video sequences also pose problems for tracking systems. Tracking 3D visual persons has also been a challenging task in computer vision. Multiple object confusion leads to mismatches in tracking. Another difficulty is handling variable-size objects and their inter-occlusion in video sequences, affecting the tracking process in video surveillance and monitoring systems. The other challenges include noisy video data, changes in scene illumination, multi-camera tracking, etc. The current literature does not address the imbalanced data issues of surveillance video data but suggests different techniques for multiple object tracking (MOT).

4. Particle Filter Fundamentals

The main aim of the PF method is the estimation of a posterior pdf $p(x_n | y_{1:n})$ from a time-varying state-space model consisting of motion model specified by a state transition pdf $p(x_n | x_{n-1})$ and observation model specified by the observation likelihood pdf $p(y_n | x_n)$ where, x_n is the state vector of an object at discrete time-instant n , $y_{1:n}$ is the set of observations i.e. video frames from discrete time-instant 1 to the current discrete time-instant n , and y_n denotes the video frame at discrete time-instant n [1, 2, 3]. In the PF method, a posterior pdf or filtering pdf is computed according to prediction step and filtering or update step under the two-step recursion when a new video frame y_n is considered [4, 5]:

$$p(x_n | y_{1:n-1}) = \int p(x_n | x_{n-1}) p(x_{n-1} | y_{1:n-1}) dx_{n-1} \quad (1)$$

$$p(x_n | y_{1:n}) = \frac{p(y_n | x_n) p(x_n | y_{1:n-1})}{\int p(y_n | x_n) p(x_n | y_{1:n-1}) dx_n} \quad (2)$$

But, the recursive computation of the above posterior pdf is not possible analytically. Instead, it is numerically approximated by a set of M random particles with related weights w at discrete time-instant n , $\{x_n^m, w_n^m\}_{m=1}^M$ and it is given as follows:

$$p(x_n | y_{1:n}) \propto \sum_{m=1}^M w_n^m \delta(x_n - x_n^m) \quad (3)$$

Where $\delta(\cdot)$ denotes the Kronecker delta function. The particles are random samples of the state vector, and the weights have non-negative values that sum up to unity, i.e. $\sum_{m=1}^M w_n^m = 1$ and are interpreted as probabilities of the particles. In practice, we often cannot draw samples directly from a posterior pdf because it may not be a standard pdf. Instead, particles are easily drawn from an importance or proposal pdf $\pi(x_n | x_{n-1}^m, y_{1:n})$ which is a biased pdf approximating the important regions of input variables. In (3), RHS is characterised as follows:

$$x_n^m \propto \pi(x_n | x_{n-1}^m, y_n) \quad (4)$$

$$w_n^j = \frac{\tilde{w}_n^m}{\sum_{m=1}^M \tilde{w}_n^m} \quad (5)$$

$$\tilde{w}_n^m = w_{n-1}^m \frac{p(y_n | x_n^m) p(x_n^m | x_{n-1}^m)}{\pi(x_n^m | x_{n-1}^m, y_n)} \quad (6)$$

If the importance pdf is made equal to the state transition pdf, i.e. $\pi(x_n^m | x_{n-1}^m, y_n) = p(x_n^m | x_{n-1}^m)$ which, then, simplifies to the above equation as $\tilde{w}_n^m = w_{n-1}^m p(y_n | x_n^m)$. The initialization of PF requires the generation of a set of random particles, $\{x_0^m, w_0^m\}_{m=1}^M$ from the prior pdf $p(x_0)$. Equations (4) to (6) show the recursive computing of the set of random particles with related weights for the representation of $p(x_n | y_{1:n})$ from the previous posterior pdf $p(x_{n-1} | y_{1:n-1})$ corresponding to a new observation y_n . The number of efficacious particles is expressed as

$$M_n^{eff} = \frac{1}{\sum_{m=1}^M (\tilde{w}_n^m)^2} \quad (7)$$

In the PF method, all the particles are assigned the exact weights, and at each discrete time-instant, the particles are propagated and assigned weights. As time evolves, degeneration of a set of particles occurs, and only a few particles are left with meaningful weights. A well-biased importance pdf and a resampling procedure can reduce the sample degeneration to a small extent. Still, it presents the supplementary challenges of sample impoverishment, resulting in a deficiency of sample diversity in the state space. Recursive update calculations of ineffective particles seriously degrade the performance of particle filters. The resampling procedure eliminates the small weighted random particles and replicates the large weighted random particles. This resampling procedure poses the problem of sample impoverishment due to the over-replication of large weighted random particles and the reduction in the number of meaningful particles. Finally, it becomes difficult for the new set of particles to keep track of the true states of targets. To overcome this problem, this paper proposes FWPF for MOT in surveillance video data.

5. Proposed FWPF for Multi-object Tracking

Fig. 1 represents the proposed FWPF-based hybrid model for MOT in surveillance videos. The first step is object extraction, in which the objects are extracted from the video through manual intervention by considering the features of the objects. The next step is particle generation, in which the particles are generated randomly for the extracted object in each video frame. Then, particle updation is performed to select the position of the multiple objects in the video frame. The particle updation uses a hybrid model of standard particle filter and FWOA methods. The standard particle filter updates the particles by calculating the weight of each particle in the extracted object and discarding the particles with small weights. In this way, the particles are resampled, and the optimal position of the object is determined. WOA is a swarm-based metaheuristic that assumes the present solution is the best prey [9]. Then, the search agents update their position to the high likelihood region of the best search agent to find the optimal solution.

5.1. Object Extraction from the Input Video

Let us assume that the proposed MOT method's input is a video consisting of n frames. The input video taken as input to the proposed method is represented as $V = \{V_i; 1 \leq i \leq n\}$ where V represents the video and n represents the number of frames in the video. The primary task in the visual tracking is to discover the moving objects from the initial frame by extracting the foreground objects. The extracted object is then given for the tracking process, which involves the identification of the object's trajectories. So, the MOT method requires an object extraction mechanism either from the initial frame of the video or when the object appears initially in the video. In this MOT method, the object is extracted manually from the video sequence at the initial frame. The detected objects from the video through manual intervention are denoted as $O = \{O_j; 1 \leq j \leq k\}$ where O represents the multiple objects set and k represents the number of objects presented in the input video V .

The preprocessing technique is based on weakly supervised object detection (WSOD). An adaptive background subtraction approach is employed to initialize the detection of moving objects over the frames. After moving objects are detected in the initial frames with little human intervention, they are weakly labelled as objects or non-objects by applying a distance-based classification method, and the same is continued in successive frames within the small spatiotemporal neighbourhood. With known positions and weak labelling for objects or non-objects, real-world video datasets are always unlabelled and imbalanced because they contain anomalies somewhere in the frames, and their exact locations are unknown. These weakly labelled objects are refined and localized by the weighted sum rule fusion of the complementary trackers for multiple objects over the frames. The WSOD-based complementary trackers (FWPF and standard PF) are suitable for unlabelled and imbalanced video datasets where limited data availability is a major concern.

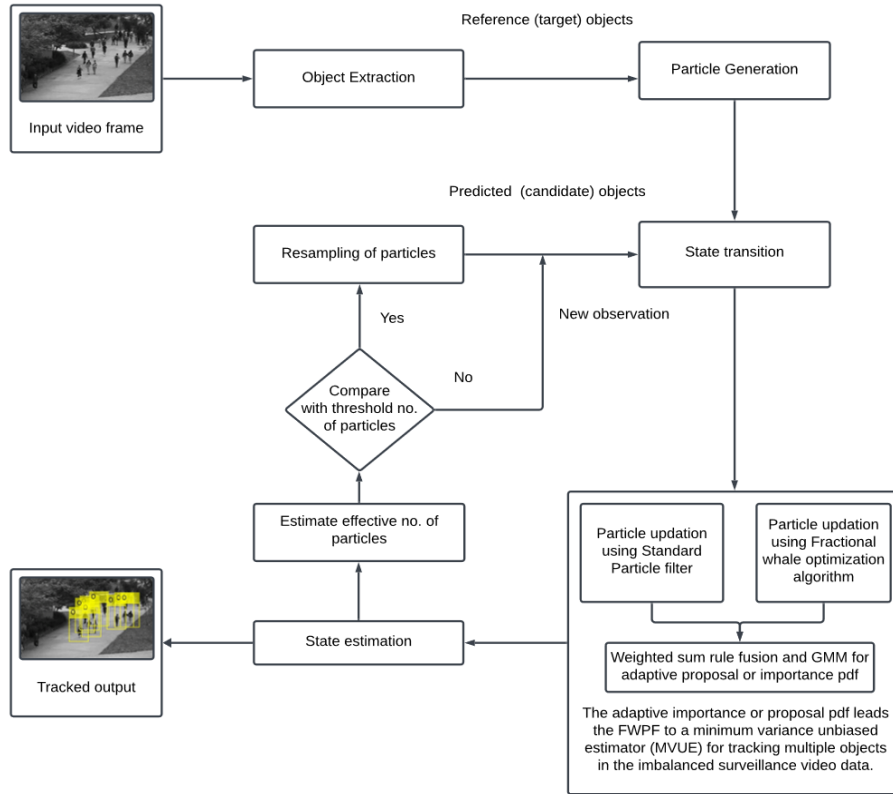


Fig.1. A proposed fractional whale particle filter method for MOT in the surveillance video

5.2. Multiple Object Tracking by Fractional Whale Particle Filter

This section describes the particle updating process in which the particles are updated using the standard particle filter and fractional calculus-based WOA method. Given the problem of sample degeneration and impoverishment in standard particle filters, the weighted sum rule fusion of FWOA and standard particle filter allows the biased proposal or importance pdf to be adaptive using the iterative Gaussian mixture model (GMM) with unbiased sample variance and sample mean.

A. Generation of Initial Particles

The particles are generated for every object after detecting the position of every object in the initial frame. Here, every object O_j contains v number of particles. The particles corresponding to every object are generated randomly to

evaluate the system's initial states. Here, the particles are used to describe the spatial and temporal attributes of moving objects in the video frames, and the particles are sampled based on the importance of the pdf.

B. Updating of Particles by Particle Filter for Visual Tracking

Once the particles are located in the initial frame, the movement (updating) of the particle is to be found to obtain the objects' location in the next frame. In a standard particle filter, the weight of each particle is determined based on the spatial and pixel correspondences. The spatial correspondence is computed based on the location of the objects using the Euclidean distance, and the pixel correspondence is computed based on the object's intensity using the Bhattacharyya distance. Once the weights of every particle are computed, they are multiplied by the previous position of the particle to find the location in the next frame. The new position of the object O_j at time T is determined as,

$$P_{O_j}^T(F) = \sum_{u=1}^v M_u^T C_u^T \quad (8)$$

Where v is the number of particles, is the centre of the particle u at time T and M_u^T is the weight of the particle u at time T . Here, the determination of the weight of each particle depends on the spatial pdf, $p(B_{O_j}^T | C_u^T)$ and the colour pdf, $p(D_{O_j}^T | C_u^T)$. The weight of the particle u at time T is determined as follows:

$$M_u^T = \frac{p(B_{O_j}^T | C_u^T) p(D_{O_j}^T | C_u^T)}{\sum_{e=1}^v p(B_{O_j}^T | C_e^T) p(D_{O_j}^T | C_e^T)} \quad (9)$$

The spatial pdf, $p(B_{O_j}^T | C_u^T)$ is determined as follows:

$$p(B_{O_j}^T | C_u^T) = \frac{1 - k(y(C_u^T), y(P_{O_j}^{T-1}))}{H} \quad (10)$$

Where, $y(C_u^T)$ is the centre of C_u^T and $y(P_{O_j}^{T-1})$ is the centre of $P_{O_j}^{T-1}$. $k(\cdot)$ represents the Euclidean distance between $y(C_u^T)$ and $y(P_{O_j}^{T-1})$. The denominator H in the above equation is computed based on the mean distance of the object region S located in the previous video frame with the height q and width p .

$$H = \frac{\sqrt{S_p (P_{O_j}^{T-1})^2 + S_q (P_{O_j}^{T-1})^2}}{2} \quad (11)$$

Then, the color pdf, $p(D_{O_j}^T | C_u^T)$ is calculated as follows,

$$p(D_{O_j}^T | C_u^T) = 1 - W(g(P_{O_j}^{T-1}), g(C_u^T)) \quad (12)$$

Where, $W(\cdot)$ represents the Bhattacharyya distance obtained from (13). $g(P_{O_j}^{T-1})$ and $g(C_u^T)$ denote the number of pixels in the bounding box surrounding the position of the object O_j at time $(T-1)$, i.e. $P_{O_j}^{T-1}$ and C_u^T , respectively.

Once the weights are computed, the particles which have lower weights are eliminated to improve the tracking accuracy. The remaining particles are the new set of particles for the next iteration. When a new set of particles is sampled, failure in the tracking gets repaired in the next few frames. In this way, the particles are resampled, and the object's new position is determined iteratively. The estimated result is applied to minimize the search area and the computation time during tracking.

C. Updating of Particles by Fractional-whale Optimization

The original whale optimization algorithm (WOA) copies the social behavior of the H-whales, which identifies the prey's location and encircles them. Since the placement of the optimal solution is unfamiliar at the beginning, the WOA

assumes the present solution as the target solution, beginning with a set of random solutions. Each iteration updates the search agent's position by either a randomly selected agent or the optimal solution. The process is repeated till it reaches the best solution. However, the WOA has a limited global search and a slow convergence rate [9].

The imbalanced surveillance video dataset introduces the drawbacks of overlapping objects during multiple object tracking, which causes problems with between-object imbalance and within-object imbalance with some missing or incomplete data. Such object imbalances occur when minority particles associated with low probable regions are not well separated from majority particles associated with high probable regions in specific areas of the input data space due to incomplete data. To counter this problem, the position of the particles is updated by the Fractional-whale optimization algorithm (FWOA), which aims to search the nearby position of the particles in the next frame to find the optimal position of the objects. This algorithm integrates the fractional calculus up to the first four terms, also known as adaptive multiple variable polynomials (AMFP), with the position update process of the whale optimization algorithm [22, 23].

The FWOA is a local and global search operator with inherent memory property. The FWOA adjusts the distribution of between-object and within-object generated particles using multiple imputation-based oversampling of minority particles in the targeted specific area of the input data space. The AMFP generates multiple different imputations for each missing data. If the missing or incomplete data are multivariate, they can be imputed from a joint multivariate normal or log-linear model. The AMFP imputes each majority particle only once for its missing data, whereas it oversamples each minority particle by replacing it with multiple imputations. Thus, all incomplete data are imputed with the oversampling of minority particles. One significant advantage of AMFP is that the oversampled minority particles are diversified with multiple imputations to eliminate data distortion, and all the observed data are kept unmodified. The FWOA approach results in low-bias and low-variance with greater efficacy for tracking precision and tracking distance in the imbalanced surveillance video dataset [24, 25].

i. Solution Encoding: Every solution is represented by a two-dimensional vector that indicates the position of the particles in the next frame.

ii. Fitness Calculation: The fitness is used to evaluate whether the generated solution is better. The fitness of every solution is calculated using the Bhattacharyya distance, which is given as follows:

$$Fitness = \sqrt{1 - \frac{\sum_j \sqrt{l(j)m(j)}}{\sqrt{\sum_j l(j) \sum_j m(j)}}} \quad (13)$$

Where l and m represent the position of search agents in the present video frame and previous video frame, respectively.

iii. Algorithm: The algorithm for FWOA is explained below:

a. Initialization: The search agents are initially populated randomly and initiate their position randomly by surrounding the object's position. This process kicks up the search agents to move toward the likelihood region of the objects placed in the previous frame.

b. Fractional encircling prey: H-whales identify the prey's location and encircle them to find the position in the next iteration. In the shrinking encircling phase, the new position of the search agents is computed as follows:

$$\bar{G}(r+1) = (1 - \bar{H} \cdot \bar{F}) \bar{G}^*(r) - \bar{H} \cdot \left(\alpha \bar{G}(r) + \frac{1}{2} \alpha \bar{G}(r-1) + \frac{1}{6} \alpha (1-\alpha) \bar{G}(r-2) + \frac{1}{24} \alpha (1-\alpha)(2-\alpha) \bar{G}(r-3) \right) \quad (14)$$

Where $\bar{G}^*$ and $\bar{G}$ are the position vectors corresponding to the best solution and the search agent with r as the current iteration. $\bar{H}$ and $\bar{F}$ are both coefficient vectors. The encircling prey is modified with fractional calculus up to the four terms with fractional order of α to be in the range of $0 < \alpha < 1$. The fractional calculus is based on the generalized Grunwald-Letnikov Equation for the discrete fractional derivative with fractional order of α [24, 25].

Then, the updation of the coefficient vectors $\bar{H}$ and $\bar{F}$ are computed as follows:

$$\bar{H} = 2\bar{h} \cdot \bar{s} - \bar{h} \quad (15)$$

$$\bar{F} = 2 \cdot \bar{s} \quad (16)$$

Where, $\vec{h}$ linearly decreases in the range $[2 \rightarrow 0]$ over the iterations, and $\vec{s}$ is a random vector in the range $[0,1]$.

c. Fractional exploitation phase: In the spiral updating position phase, the new position of the search agents is computed based on the following equation:

$$\vec{G}(r+1) = (1 + e^{df} \cdot \cos(2\pi f)) \vec{G}^*(r) - e^{df} \cdot \cos(2\pi f) \cdot \left(\begin{aligned} &\alpha \vec{G}(r) + \frac{1}{2} \alpha \vec{G}(r-1) + \frac{1}{6} \alpha (1-\alpha) \vec{G}(r-2) + \\ &\frac{1}{24} \alpha (1-\alpha) (2-\alpha) \vec{G}(r-3) \end{aligned} \right) \quad (17)$$

Where d is a constant denoting the state of the logarithmic helix in the range, f is a random number $[-1,1]$ and ' $\cdot$ ' is the element-wise multiplication. In the shrinking circling mechanism, the position of H-whales is updated as per (13) and (17) when the random probability of occurrence is less than or greater than $p = 0.5$.

d. Fractional exploration phase: The search agents search for prey (perfect solution) randomly, and the position of the search agent is updated based on the position of other search agents. This can be represented as:

$$\vec{G}(r+1) = (1 - \vec{H} \cdot \vec{F}) \vec{G}_m - \vec{H} \cdot \left(\alpha \vec{G}(r) + \frac{1}{2} \alpha \vec{G}(r-1) + \frac{1}{6} \alpha (1-\alpha) \vec{G}(r-2) + \frac{1}{24} \alpha (1-\alpha) (2-\alpha) \vec{G}(r-3) \right) \quad (18)$$

Where, $\vec{G}_m$ is specified as a random position vector.

e. Termination: The FWOA is performed between either a circular or spiral movement, considering the value of p . Then, the process is repeated until it reaches the maximum number of iterations the user gives. After reaching the maximum iteration, the best prey location is assigned to $P_{O_j}^T(FW)$, which is the position of the particle in the next frame.

D. Weighted Sum Rule Fusion for Updating the New Position

The new position obtained through the PF and FWOA is combined with the weighted sum rule fusion to track multiple objects in the video sequence. The fusion provides the mean position of the PF and FWOA as the final tracking paths of the multi-objects. The following equation determines the new position of the object:

$$\begin{aligned} P_{O_j}^T &= x_n = \frac{P_{O_j}^T(PF) + P_{O_j}^T(FW)}{2}, \\ Z &= \sum_{i=1}^{N=2} \alpha_i X_i \text{ such that } \sum_{i=1}^{N=2} \alpha_i = 1, \text{ and} \\ \sigma_Z^2 &= \sum_{i=1}^{N=2} \alpha_i^2 \sigma_{X_i}^2 = \alpha_1^2 \sigma_{X_1}^2 + \alpha_2^2 \sigma_{X_2}^2 \end{aligned} \quad (19)$$

Where $P_{O_j}^T(PF)$ is the new position of the object obtained by the PF method and $P_{O_j}^T(FW)$ is the new position of the object obtained by the FWOA. Since $P_{O_j}^T(PF)$ and $P_{O_j}^T(FW)$ are originated from the same video frame (data) distribution; they are not independent but are uncorrelated due to the FWOA-based tracker being complementary to the PF-based tracker. When treated as uncorrelated random variables, these complementary trackers produce a minimum variance, undergoing weighted sum rule fusion with equal weighting coefficients. If these complementary trackers achieve minimum variance for multiple objects, it will result in higher precision. The other score-level fusion techniques, like the maximum rule, minimum rule, product rule, median rule, etc., are not applied because these trackers are not independent [26, 27].

The weighted sum rule fusion is based on optimization by complementarity. One tracking model's incorrect detection and localization are complemented by another. If the trackers complement each other, they cancel out each other's random errors and help output correct or unbiased detection. These complementary trackers are uncorrelated but not independent because they originate from the same video frame (data) distributions. These trackers cause two complementary events which are mutually exclusive to each other. These trackers undergo the weighted sum rule fusion as per (19), where the terms X_i and α_i are uncorrelated random variables (two complementary trackers) and real weighting constants. The term Z is another random variable formed through weighted sum rule fusion. Further, we obtain the expression of the variance of Z in terms of the variances of X_1 and X_2 as given in (19). In statistics, precision is the reciprocal of the variance. In order to verify the approximated values of the precision, we assume $\sigma_{X_1}^2 = \sigma_{X_2}^2 = \sigma^2$ and consider different

values of α_1 and α_2 . With $\alpha_1 = \alpha_2 = 0.5$, it is found that the averaging of two complementary trackers $\left(Z = \frac{X_1 + X_2}{2} \right)$ results in higher precision $\left(MOTP \propto \frac{1}{\sigma_z^2} \right)$.

E. Adaptive Proposal pdf for the Imbalanced Data

The imbalanced surveillance video dataset consists of majority and minority classes as normal and anomalous instances. The normal and anomalous instances cause majority and minority samples associated with high and low probable regions. The minority particles tend to be at high risk of being suppressed by the majority particles, as the proposal or importance pdf encourages the highly probable regions of the input data space to remain a “biased distribution” in the standard particle filter. The proposal or importance sampling depends entirely on the choice of biased distribution, which results in a biased estimation of the discrete random measure. However, true and normalized importance weights are introduced to correct the bias to a small extent. The difficulty in designing the correct proposal or importance pdf prevents the use of particle filtering-based tracking in the imbalanced surveillance video dataset. Further, the standard particle filter tracker suffers from the problem of sample degeneration and sample impoverishment due to the biased proposal or importance pdf ignoring the minority particles [28, 29].

To alleviate this problem, The FWOA is fused with the standard particle filter under the weighted sum rule fusion to represent the proposal or importance pdf as a mixture of Gaussian distributions to deal with highly non-Gaussian true posterior pdf in the imbalanced data. Thus, FWOA adapts each component's weights and parameters in the GMM to result in the adaptive proposal or importance pdf for the FWPF, which balances exploring new areas of the state space with exploiting promising regions [30, 31].

i. Further to (19), a set of M random particles with associated weights u at discrete time-instant n , $\{x_n^m, u_n^m\}_{m=1}^M$ is considered to denote x_n^1 and x_n^M as best (with a small selection pressure coefficient, α) and worst (with a large selection pressure coefficient, α) initial solution, whereas the associated weight u_n^m is defined to be Gaussian distributed with constraint as under:

$$u_n^m \propto \frac{1}{\sqrt{2\pi\alpha M}} \exp \left\{ -\frac{1}{2} \left(\frac{m-1}{\alpha M} \right)^2 \right\} \quad \text{such that} \quad \sum_{m=1}^M u_n^m = 1 \quad (20)$$

ii. The GMM, as the importance or proposal pdf, is generated for every p^{th} observations with Gaussian probability density function, characterized by the mean (μ) and variance (σ^2) as follows:

$$G_n^p \{x_n^m(p)\} = \sum_{m=1}^M u_n^m \mathcal{N} \left(x_n^m(p); \mu_n^m(p), (\sigma_n^m(p))^2 \right) \quad (21)$$

$$\mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \right\} \quad (22)$$

Where $x_n(p)$ denotes the p^{th} element of the vector x_n . The unbiased sample mean and sample variance for the observations drawn from the Gaussian pdf are defined as follows:

$$\mu_n^m(p) = \bar{x}_n^m(p) = \frac{1}{M} \sum_{m=1}^M x_n^m(p) \quad (23)$$

$$\{\sigma_n^m(p)\}^2 = \frac{1}{(M-1)} \sum_{m=1}^M \{x_n^m(p) - \bar{x}_n^m(p)\}^2 \quad (24)$$

iii. Thus, the GMM generates N new samples that represent the progeny of the primary population. The most optimal function is assessed for all the newly generated samples. The M_n^{eff} fittest solution among the M current samples and N new samples are formed to find the best solution. The adaptive importance or proposal pdf, $G_n(x_n^m)$ now modifies (6) with (21) and (22) as follows:

$$G_n(x_n^m) = \prod_{p=1}^Q G_n^p(x_n^m(p)) = G_n^1(x_n^m(1)) \times G_n^2(x_n^m(2)) \times \cdots \times G_n^Q(x_n^m(Q)) \quad (25)$$

$$\tilde{w}_n^m = \frac{\frac{1}{M} \sum_{l=1}^M p(y_n | x_n^m) p(x_n^m | x_{n-1}^l)}{G_n(x_n^m)} \quad (26)$$

Once the weights are modified according to adaptive importance or proposal pdf, the particles $\{x_n^m, \tilde{w}_n^m\}_{m=1}^M$ are completely updated/resampled based on the modified posterior pdf from (3) till the completion of the epoch n as follows:

$$p(x_n | y_{1:n}) \propto \sum_{m=1}^M \tilde{w}_n^m \delta(x_n - x_n^m) \quad (27)$$

6. Results and Discussion

The proposed approach is simulated using the "MATLAB2019b" software tool on a system with an "Intel" core processor and 2 GB RAM, as well as the "Windows-11" operating system. The imbalanced UCSD video data – 1 and 2 [32] are obtained using a stationary camera mounted on the building, observing the pedestrian walkways where the density varies from sparse to crowded. Some occlusions and scale changes cause anomalies like cyclists, carts, and skateboarders to move at nonlinear speed. Each video has 100 frames with an image resolution of 158 × 238 pixels. The reason behind considering this video data is that the camera's field of view leads to a heightened degree of perspective distortion and pronounced variations in the scale of moving objects. Furthermore, anomalous events encompass motion anomalies elicited by cyclists, carts, and skateboarders, as well as contextual anomalies. The simulation results are discussed in this section.

6.1. Evaluation Metrics

Multiple objects tracking accuracy (MOTA) can reveal how class imbalance affects the overall tracking performance. For instance, a disproportionately high false negative rate for a minority class could indicate that the tracker consistently struggles to detect or maintain tracks of these minority or anomalous objects. In statistics, precision is often defined as the inverse of variance. Intuitively, variance measures how spread out a set of values is from its mean, while precision indicates how close or concentrated those values are to the mean. As a minimum variance unbiased estimator, the proposed method considers the Multiple objects tracking precision (MOTP), which can assess how accurately the tracker localizes those objects, regardless of their frequency in the video. Tracking distance correlation (TDC) can assess object tracks' temporal consistency and reliability. This is particularly important when dealing with imbalanced video data, where the performance of minority or anomalous objects might be compromised.

A. Multiple Object Tracking Accuracy (MOTA)

This metric incorporates three types of errors: false negatives, false positives, and identity mismatches. For the most accurate tracking, a high MOTA value is desired. The total count of ground truth objects is GT_t at frame (time) t . The terms FN_t , FP_t , and $IDMM_t$ are the number of false negatives, false positives, and identity mismatches at frame (time) t , respectively [33, 34].

$$MOTA = 1 - \frac{\sum_t (FN_t + FP_t + IDMM_t)}{\sum_t GT_t} \quad (28)$$

B. Multiple Object Tracking Precision (MOTP)

The metric refers to the average difference between all true positives and their ground truth hypothesis. A high MOTP value is preferred for the most precise tracking. The average dissimilarity among all true positives and their corresponding ground-truth hypothesis is the Multiple Object Tracking Precision [33, 34].

$$MOTP = \frac{\sum_{t,i} d_{t,i}}{\sum_t c_t} \quad (29)$$

Where the terms $d_{t,i}$ and c_t are the bounding box overlap of the target i with its assigned ground truth object and the count of matches at frame (time) t , respectively. It is used to represent the ability of the tracker to evaluate the correct position of the object positions in a video frame. MOTP and MOTA values vary between 0 (no tracking) and 1 (fully tracked). Higher values of both MOTP and MOTA are desirable for accurate and precise tracking.

C. Tracking Distance Correlation (TDC)

This metric measures both linear and nonlinear association between the original tracking path (X) and algorithmic tracked path (Y) in the imbalanced particle filter-based tracking. TDC value varies between 0 (when two paths are the same) and 1 (when two paths are independent). This metric depends on distance covariance, and its minimum value shows the bias and variance correction from tracking based on a biased proposal pdf to tracking based on an unbiased proposal pdf as required for FWPF. The tracking distance correlation is represented as follows [35]:

$$TDC(X, Y) = 1 - \frac{TDCov(X, Y)}{\sqrt{TDVar(X)TDVar(Y)}} \quad (30)$$

Where $TDCov(\cdot)$ and $TDVar(\cdot)$ represent the paths's distance covariance and distance variance.

6.2. Experimental Results

The performance of the proposed method is evaluated using two videos obtained from the UCSD dataset [32], each containing 100 frames. The UCSD video datasets are considered limited, unlabelled, real-world imbalanced surveillance video datasets. Figs. 2 (a) and (c) represent the 7th, 21st, and 35th frames of video-1 (Bicycle) and video-2 (Cart). The experimentation is performed to track the different numbers of multiple objects (from one to eight) from video-1 and video-2, as shown in Figs. 2 (b) and (d).

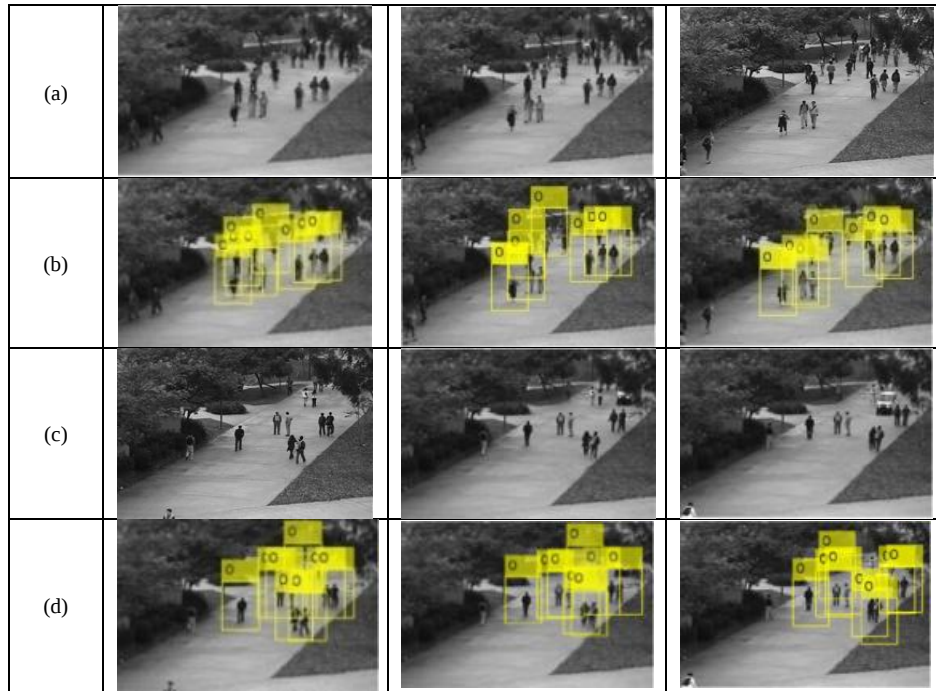


Fig.2. Proposed FWPF-based tracking over the original frames 7, 21, and 35 of the (a, b) UCSD video data – 1 [32] and (c, d) UCSD video data – 2 [32]

6.3. Performance Analysis

This section presents the comparative performance evaluation of the proposed FWPF with the existing methods like standard PF [1], Modified grey wolf particle filter (MGWPF) [13], Adaptive chaotic sampling particle filter (ACPF) [14], and Quantum particle swarm genetic particle filter (QPGPF) [15] when applying on the 100 frames extracted from UCSD video data – 1 and 2 [32].

A. Evaluation based on MOTP

Table 1 compares the average MOTP values of the proposed FWPF method with the existing methods [1, 13, 14, 15]

for the different numbers of individual tracked objects (from one to eight) in UCSD video data-1 and 2. Figs. 3 (a) and (c) depict the MOTP analysis curve for UCSD video data -1 and 2, respectively. The proposed method attains an average MOTP of 92.71% for UCSD video data -1, outperforming the MOTP values of 56.85%, 84.58%, 76.80%, and 70.63% achieved by PF, MGWPF, ACPF, and QPGPF, respectively. Similarly, the proposed method attains an average MOTP of 91.43% for UCSD video data -2, outperforming the MOTP values of 71.55%, 85.93%, 81.75%, and 78.82% achieved by PF, MGWPF, ACPF, and QPGPF, respectively.

B. Evaluation based on MOTA

Table 1 compares the average MOTA values of the proposed FWPF method with the existing methods [1, 13, 14, 15] for the different numbers of individual tracked objects (from one to eight) in UCSD video data-1 and 2. Figs. 3 (b) and (d) depict the MOTA analysis curve for UCSD video data -1 and 2, respectively. The proposed method attains an average MOTA of 90.73% for UCSD video data -1, outperforming the MOTA values of 53.03%, 81.76%, 73.27%, and 67.31% achieved by PF, MGWPF, ACPF, and QPGPF, respectively. Similarly, the proposed method attains an average MOTA of 89.58% for UCSD video data -2, outperforming the MOTA values of 65.93%, 82.67%, 76.86%, and 75.46% achieved by PF, MGWPF, ACPF, and QPGPF, respectively.

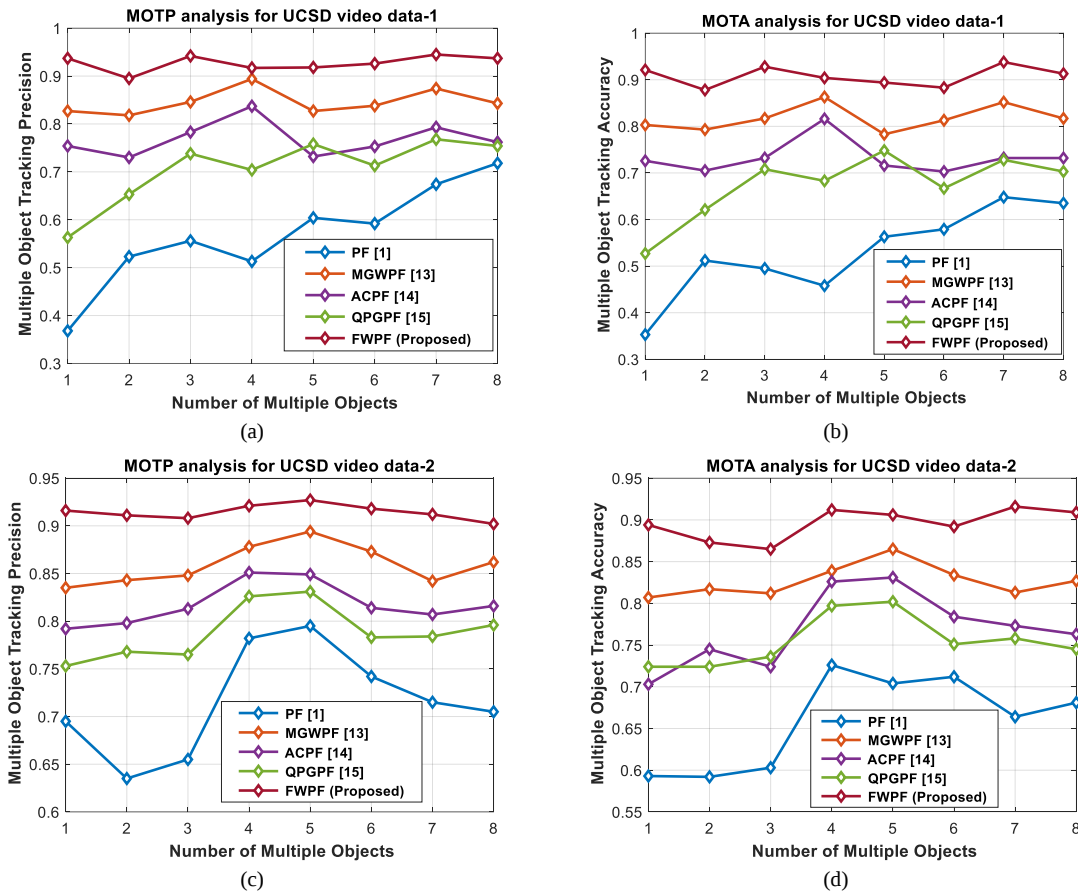


Fig.3. (a, b) MOTP and MOTA analysis curve and (c, d) MOTP and MOTA analysis curve for the UCSD video data -1 and 2 [32]

C. Evaluation based on TDC

Table 2 compares the tracking distance correlation (TDC) values of the proposed FWPF method with the existing methods [1, 13, 14, 15] for the different numbers of individual tracked objects (from one to eight) in UCSD video data-1 and 2. Figs. 4 (a) and (b) depict the TDC analysis curve for imbalanced tracking over UCSD video data -1 and 2, respectively. The proposed method attains TDC value of 0.0038 (for tracking of eight multiple objects) for UCSD video data -1, outperforming the TDC values of 0.0388, 0.0166, 0.0214, and 0.0318 achieved by PF, MGWPF, ACPF, and QPGPF, respectively. Similarly, the proposed method attains TDC value of 0.0035 (for tracking of eight multiple objects) for UCSD video data -2, outperforming the TDC values of 0.0550, 0.0128, 0.0175, and 0.0319 achieved by PF, MGWPF, ACPF, and QPGPF, respectively.

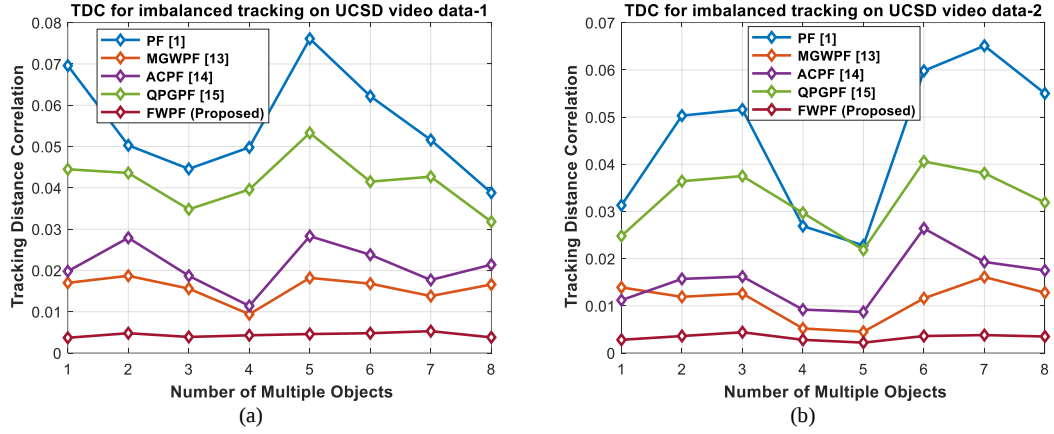


Fig.4. (a, b) Tracking distance correlation for imbalanced tracking on the UCSD video data – 1 and 2 [32]

Table 1. MOTP and MOTA comparison of methods in UCSD video data-1 and 2

Video	Methods	Average MOTP	Average MOTA
UCSD video data - 1 [32]	PF [1]	56.85	53.03
	MGWPF [13]	84.58	81.76
	ACPF [14]	76.8	73.27
	QPGPF [15]	70.63	67.31
	FWPF (Proposed)	92.71	90.73
UCSD Video data - 2 [32]	PF [1]	71.55	65.93
	MGWPF [13]	85.93	82.67
	ACPF [14]	81.75	76.86
	QPGPF [15]	78.82	75.46
	FWPF (Proposed)	91.43	89.58

Table 2. Tracking distance correlation for imbalanced tracking of (1 to 8) multiple objects over UCSD video data – 1 and 2

Video data	Methods	Tracking distance correlation for imbalanced tracking of (1 to 8) multiple objects over video data							
		1	2	3	4	5	6	7	8
UCSD-1 [32]	PF [1]	0.0696	0.0503	0.0446	0.0498	0.0761	0.0622	0.0516	0.0388
	MGWPF [13]	0.0170	0.0187	0.0156	0.0094	0.0182	0.0168	0.0138	0.0166
	ACPF [14]	0.0198	0.0279	0.0187	0.0114	0.0283	0.0238	0.0177	0.0214
	QPGPF [15]	0.0445	0.0436	0.0348	0.0396	0.0533	0.0415	0.0427	0.0318
	FWPF (Proposed)	0.0037	0.0048	0.0039	0.0043	0.0046	0.0048	0.0053	0.0038
UCSD-2 [32]	PF [1]	0.0313	0.0503	0.0516	0.0269	0.0228	0.0598	0.0651	0.0550
	MGWPF [13]	0.0139	0.0119	0.0126	0.0052	0.0045	0.0116	0.0161	0.0128
	ACPF [14]	0.0112	0.0157	0.0162	0.0092	0.0087	0.0264	0.0193	0.0175
	QPGPF [15]	0.0248	0.0364	0.0375	0.0297	0.02185	0.0406	0.0381	0.0319
	FWPF (Proposed)	0.0028	0.0036	0.0044	0.0028	0.0022	0.0036	0.0038	0.0035

D. Evaluation based on Sample Mean and Sample Variance

Tables 3 and 4 compare the sample mean and sample variance of the proposed FWPF method with the existing methods [1, 13, 14, 15] when applied over the imbalanced UCSD video data-1 and 2. The proposed method attains minimum sample mean and variance of 0.0044 (close to zero) and 3.25714×10^{-7} for UCSD video data - 1, as compared to sample mean (sample variance) values of 0.0554 (1.62203×10^{-4}), 0.0157 (8.8856×10^{-6}), 0.0211 (3.1324×10^{-5}), and 0.0415 (4.25221×10^{-5}) achieved by PF, MGWPF, ACPF, and QPGPF, respectively. Similarly, the proposed method attains minimum sample mean and variance of 0.0033 (close to zero) and 4.82679×10^{-7} for UCSD video data - 2, as compared to sample mean (sample variance) values of 0.0454 (2.57158×10^{-4}), 0.0110 (1.68079×10^{-5}), 0.0155 (3.45568×10^{-5}), and 0.0326 (4.56215×10^{-5}) achieved by PF, MGWPF, ACPF, and QPGPF, respectively.

Table 3. Sample mean and sample variance for imbalanced tracking of multiple objects (multiple observations) over UCSD video data -1

Methods	Sample mean	Sample variance
PF [1]	0.0554	1.62203×10^{-4}
MGWPF [13]	0.0157	8.88560×10^{-6}
ACPF [14]	0.0211	3.13240×10^{-5}
QPGPF [15]	0.0415	4.25221×10^{-5}
FWPF (Proposed)	0.0044	3.25714×10^{-7}

Table 4. Sample mean and sample variance for imbalanced tracking of multiple objects (multiple observations) over UCSD video data -2

Methods	Sample mean	Sample variance
PF [1]	0.0454	2.57158×10^{-4}
MGWPF [13]	0.0110	1.68079×10^{-5}
ACPF [14]	0.0155	3.45568×10^{-5}
QPGPF [15]	0.0326	4.56215×10^{-5}
FWPF (Proposed)	0.0033	4.82679×10^{-7}

E. Evaluation based on Bias-variance Correction for Imbalance Reduction

Tables 5 and 6 compare the bias-variance correction for imbalance reduction by the proposed FWPF method to the existing methods [1, 13, 14, 15] when applied over the imbalanced UCSD video data-1 and 2. The proposed method attains the correction of bias (variance) by -92.06% (-99.79%), -72.12% (-96.33%), -79.21 (-98.96%), and -89.40 (-99.23%) for imbalance reduction over UCSD video data - 1 when compared to PF, MGWPF, ACPF, and QPGPF, respectively. Similarly, the proposed method attains the correction of bias (variance) by -92.64% (-99.81%), -69.90% (-97.12%), -78.53 (-98.60%), and -89.77 (-98.94%) for imbalance reduction over UCSD video data - 2 when compared to PF, MGWPF, ACPF, and QPGPF, respectively.

Table 5. Bias-variance correction (imbalance reduction shown with negative sign) by the proposed method (FWPF as minimum variance unbiased estimator) in comparison with other methods for imbalanced tracking over UCSD video data -1

Methods	Bias correction (%)	Variance correction (%)
PF [1]	-92.06	-99.79
MGWPF [13]	-72.12	-96.33
ACPF [14]	-79.21	-98.96
QPGPF [15]	-89.40	-99.23

Table 6. Bias-variance correction (imbalance reduction shown with negative sign) by the proposed method (FWPF as minimum variance unbiased estimator) in comparison with other methods for imbalanced tracking over UCSD video data -2

Methods	Bias correction (%)	Variance correction (%)
PF [1]	-92.64	-99.81
MGWPF [13]	-69.90	-97.12
ACPF [14]	-78.53	-98.60
QPGPF [15]	-89.77	-98.94

F. Evaluation based on Computational Time and Limitations of Methods

Table 7 compare the computational time of the proposed FWPF method as compared to the existing methods [1, 13, 14, 15] when applied over the imbalanced UCSD video data-1 and 2. The proposed method processes the 100 frames each of UCSD video data-1 and 2 in 55.993 seconds and 25.775 seconds, respectively. The computational time is the lowest compared to the existing methods. The proposed method is suitable for limited (short-duration) video data obtained from fixed surveillance cameras under weakly supervised object detection settings. The limitations of other competing methods are as follows. In PF-based method, grid-based methods require a predefined state space, limiting uneven partitioning for high-probability density regions without prior knowledge. It highlights that the variance of importance weights increases over time, making degeneracy unavoidable. The computational cost of grid-based methods increases dramatically with state space dimensionality [1]. The MGWPF-based method may face challenges in motion blur and fast motion scenarios, leading to sample degeneracy. The convergence of the MGWO-PF algorithm can vary across different datasets, affecting performance consistency [13]. In the ACPF-based method, local representations fail to capture complex appearances effectively. Global representations are not robust to partial occlusions. Errors tend to accumulate during object model updates, leading to tracking failures under occlusions. The complexity and cost of global search pose an issue [14]. In the QPGPF-based method, the complexity of feature selection strategies is mentioned as a

burden on computational efficiency. The position updating equation has some limitations in optimizing the particle distributions with imbalanced video data. This method incurs higher computational costs despite improving target positioning accuracy [15].

Table 7. Computational time (in seconds per 100 frames of UCSD video data-1 and 2)

Methods	UCSD-1 [32]	UCSD-2 [32]
PF [1]	493.756	176.107
MGWPF [13]	119.621	54.758
ACPF [14]	145.072	67.198
QPGPF [15]	404.676	168.769
FWPF (Proposed)	55.993	25.775

6.4. Discussion

The proposed method is not meant for real-time tracking. Instead, it is used for limited, unlabelled real-world imbalanced video datasets. The limited video data from CCTV cameras for a particular time duration can be of important investigation by the law-enforcement authorities. Further, weakly labelled training data is generated using weakly supervised object detection. The weighted sum rule fusion of complementary trackers can further refine the localization of multiple objects. The proposed method tries to convert imbalanced tracking with a biased proposal pdf into balanced tracking with an unbiased proposal pdf in limited, unlabelled, real-world imbalanced video datasets. The conditions of extreme data imbalance with noisy input are generally found in the limited video data footage obtained from the video surveillance camera for a particular time duration. Weakly supervised object detection and refined localization by fusion of complementary trackers result in a better tracking model that deals with the imbalance in the limited video data footage [36]. As a minimum variance unbiased estimator, the proposed method exhibits bias-variance correction as an imbalance reduction factor, as shown in Tables 5 and 6.

The higher values of MOTP and MOTA are attained by the proposed FWPF method because it fuses the fractional calculus-based whale optimization algorithm and the standard particle filter under weighted sum rule fusion. Integrating the FWPF with an iterative Gaussian mixture model (GMM) with unbiased sample variance and sample mean allows the proposal or importance pdf to adapt to the imbalanced nature of surveillance video data. Thus, there are two trackers, one based on particle filter with a biased proposal or importance pdf and another based on FWOA with the adaptive proposal or importance pdf. Since these two trackers operate on the same video frame (data) distribution, they are not independent but are uncorrelated due to their complementary nature. These trackers deal with two complementary video data distributions that are mutually exclusive. These two trackers' weighted sum rule fusion leads the FWPF to a minimum variance unbiased estimation for effectively detecting and tracking multiple objects in the imbalanced surveillance video data. Since multiple objects are extracted with human intervention in the initial frame, making them weakly labelled (object or non-objects), the weighted sum rule fusion improves both MOTP and MOTA (with lowest TDC) over the frames due to the minimum variance achieved by these trackers being complementary. The FWPF, being an unbiased estimator, also results in the higher MOTP and MOTA overcoming the challenges of the imbalanced surveillance videos. The proposed method solves the problems of sample degeneration and impoverishment along with other tracking problems, such as scale change of objects tracked, occlusion, tracking resumption, etc [30, 31]. To justify the claim of the proposed method acting as a minimum variance unbiased estimator, the sample mean obtained from the multiple observations with different numbers of objects (from one to eight) is determined to be close to zero (unbiased) with minimum variance. The proposed method justifies the imbalance reduction through the bias-variance correction based on sample mean and sample variance obtained for multiple observations. The bias-variance correction (imbalance reduction) is attributed to an unbiased proposal pdf in the imbalanced FWPF-based tracking.

7. Conclusions

This paper fuses the fractional calculus-based whale optimization algorithm (FWOA) and the standard particle filter under weighted sum rule fusion. The fractional calculus up to the first four terms, known as adaptive multiple variable polynomials, makes the FWOA a local and global search operator with inherent memory property. Integrating the FWPF with an iterative Gaussian mixture model (GMM) with unbiased sample variance and mean allows the proposal or importance pdf to be adaptive to deal with the imbalanced surveillance video data, which is highly nonlinear and non-Gaussian. Thus, there are two trackers, one based on particle filter with a biased proposal or importance pdf and another based on FWOA with the adaptive proposal or importance pdf. In view of operating on the same video frame (data) distribution, these trackers are not independent but are uncorrelated due to their complementary nature. These two complementary trackers' weighted sum rule fusion leads the FWPF to a minimum variance unbiased estimation for effectively detecting and tracking multiple objects in the imbalanced surveillance video data. Since multiple objects are extracted with human intervention in the initial frame are weakly labelled (object or non-objects), the weighted sum rule fusion of these complementary trackers causes minimum variance, improving both MOTP and MOTA over the frames. The proposed method handles sample degeneration, impoverishment, and other tracking challenges, such as scale change

of objects tracked, occlusion, tracking resumption, etc. The resultant tracker, FWPF, achieves a higher MOTP of 92.71% (91.43%), MOTA of 90.73% (89.58%), TDC of 0.0038 (0.0035) for tracking of eight multiple objects in the imbalanced UCSD surveillance video data-1 (2) and outperforms the existing methods. The consistent tracking distance correlation near zero values infers the bias-variance correction for efficient imbalance reduction by the proposed method.

Acknowledgement

I confirm that the presented research work is the author's individual work. I also attest to the validity and legitimacy of the data and its interpretation. There are no conflicts of interest present, either with any individual or the organization.

References

- [1] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002, doi: 10.1109/78.978374.
- [2] O. Cappe, S. J. Godsill, and E. Moulines, "An overview of existing methods and recent advances in sequential Monte Carlo," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 899–924, 2007, doi: 10.1109/JPROC.2007.893250.
- [3] M. Isard and A. Blake, "CONDENSATION-Conditional Density Propagation for Visual Tracking," *Int J Comput Vis*, vol. 29, no. 1, pp. 5–28, 1998.
- [4] G. M. Rao and Ch. Satyanarayana, "Visual Object Target Tracking Using Particle Filter: A Survey," *International Journal of Image, Graphics and Signal Processing*, vol. 5, no. 6, pp. 57–71, May 2013.
- [5] F. Gustafsson et al., "Particle Filters for Positioning, Navigation, and Tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 425–437, 2002.
- [6] S. Wang, C. Shao, S. Xu, X. Yang, and H. Yu, "MSFSS: A whale optimization-based multiple sampling feature selection stacking ensemble algorithm for classifying imbalanced data," *AIMS Mathematics*, vol. 9, no. 7, pp. 17504–17530, 2024, doi: 10.3934/math.2024851.
- [7] E. M. Hassib, A. I. El-Desouky, L. M. Labib, and E. S. M. El-kenawy, "WOA + BRNN: An imbalanced big data classification framework using Whale optimization and deep neural network," *Soft comput*, vol. 24, no. 8, pp. 5573–5592, Apr. 2020, doi: 10.1007/s00500-019-03901-y.
- [8] S. Park, J. P. Hwang, E. Kim, and H. J. Kang, "A new evolutionary particle filter for the prevention of sample impoverishment," *IEEE Transactions on Evolutionary Computation*, vol. 13, no. 4, pp. 801–809, 2009, doi: 10.1109/TEVC.2008.2011729.
- [9] S. Mirjalili and A. Lewis, "The Whale Optimization Algorithm," *Advances in Engineering Software*, vol. 95, pp. 51–67, May 2016, doi: 10.1016/j.advengsoft.2016.01.008.
- [10] Y. Wu and T. S. Huang, "Robust Visual Tracking by Integrating Multiple Cues Based on Co-Inference Learning," *Int J Comput Vis*, vol. 58, no. 1, pp. 55–71, 2004.
- [11] J. Berclaz, F. Fleuret, E. Türetken, and P. Fua, "Multiple object tracking using k-shortest paths optimization," *IEEE Trans Pattern Anal Mach Intell*, vol. 33, no. 9, pp. 1806–1819, 2011, doi: 10.1109/TPAMI.2011.21.
- [12] X. Hai-Xia, W. Yao-Nan, Z. Wei, Z. Jiang, and Y. Xiao-Fang, "Multi-object visual tracking based on reversible jump Markov chain Monte Carlo," *IET Computer Vision*, vol. 5, no. 5, pp. 282–290, Sep. 2011, doi: 10.1049/iet-cvi.2010.0086.
- [13] M. Narayana, H. Nenavath, S. Chavan, and L. Koteswara Rao, "Intelligent visual object tracking with particle filter based on Modified Grey Wolf Optimizer," *Optik - International Journal for Light and Electron Optics*, vol. 193, pp. 1–21, Sep. 2019, doi: 10.1016/j.ijleo.2019.06.013.
- [14] M. Firouznia, J. A. Koupaei, K. Faez, G. A. Trunfio, and H. Aminadavar, "Adaptive chaotic sampling particle filter to handle occlusion and fast motion in visual object tracking," *Digital Signal Processing: A Review Journal*, vol. 134, pp. 1–13, Apr. 2023, doi: 10.1016/j.dsp.2023.103933.
- [15] X. Yuqi, W. Yongjun, and Y. Fan, "A scale adaptive generative target tracking method based on modified particle filter," *Multimed Tools Appl*, vol. 82, no. 20, pp. 31329–31349, Aug. 2023, doi: 10.1007/s11042-023-14901-4.
- [16] S. Xu and H. Chen, "Visual Tracking via a Novel Adaptive Anti-occlusion Mean Shift Embedded Particle Filter," *Circuits Syst Signal Process*, pp. 1–26, Feb. 2024, doi: 10.1007/s00034-024-02882-0.
- [17] J. Li, Z. Cao, F. Liu, Y. Fu, X. Li, and F. Tian, "An adaptive biogeography-based optimization with integrated covariance matrix learning for robust visual object tracking," *Expert Syst Appl*, vol. 234, pp. 1–24, Dec. 2023, doi: 10.1016/j.eswa.2023.121110.
- [18] Y. Xiao and Y. Wu, "Robust visual tracking based on modified mayfly optimization algorithm," *Image Vis Comput*, vol. 135, pp. 1–8, Jul. 2023, doi: 10.1016/j.imavis.2023.104691.
- [19] M. Tian, Y. Bo, Z. Chen, P. Wu, and C. Yue, "Multi-target tracking method based on improved firefly algorithm optimized particle filter," *Neurocomputing*, vol. 359, pp. 438–448, Sep. 2019, doi: 10.1016/j.neucom.2019.06.003.
- [20] Z. Li, J. Chen, and J. Bi, "Multiple Object Tracking With Appearance Feature Prediction and Similarity Fusion," *IEEE Access*, vol. 11, pp. 52492–52500, 2023, doi: 10.1109/ACCESS.2023.3279868.
- [21] F. Wang, Y. Wang, J. He, F. Sun, X. Li, and J. Zhang, "Visual object tracking via iterative ant particle filtering," *IET Image Process*, vol. 14, no. 8, pp. 1636–1644, Jun. 2020, doi: 10.1049/iet-ipr.2019.0967.
- [22] S. Arora, T. Mathur, S. Agarwal, K. Tiwari, and P. Gupta, "Applications of fractional calculus in computer vision: A survey," *Neurocomputing*, vol. 489, pp. 407–428, Jun. 2022, doi: 10.1016/j.neucom.2021.10.122.
- [23] T. P. Morris, I. R. White, J. R. Carpenter, S. J. Stanworth, and P. Royston, "Combining fractional polynomial model building with multiple imputation," *Stat Med*, vol. 34, no. 25, pp. 3298–3317, Nov. 2015, doi: 10.1002/sim.6553.
- [24] M. S. Couceiro, R. P. Rocha, N. M. F. Ferreira, and J. A. T. Machado, "Introducing the fractional-order Darwinian PSO," *Signal Image Video Process*, vol. 6, no. 3, pp. 343–350, Sep. 2012, doi: 10.1007/s11760-012-0316-2.
- [25] E. J. Solteiro Pires, J. A. Tenreiro MacHado, P. B. De Moura Oliveira, J. Boaventura Cunha, and L. Mendes, "Particle swarm optimization with fractional-order velocity," *Nonlinear Dyn*, vol. 61, no. 1–2, pp. 295–301, Jul. 2010, doi: 10.1007/s11071-009-9649-y.

- [26] M. Dunnhofer, K. Simonato, and C. Micheloni, "Combining complementary trackers for enhanced long-term visual object tracking," *Image Vis Comput*, vol. 122, pp. 1–10, Jun. 2022, doi: 10.1016/j.imavis.2022.104448.
- [27] W. Dai, L. Jin, and G. Li, "Long-term adaptive tracking via complementary trackers," *IET Image Process*, vol. 13, no. 9, pp. 1569–1577, Jul. 2019, doi: 10.1049/iet-ipr.2018.6142.
- [28] Y. Xie, L. Peng, Z. Chen, B. Yang, H. Zhang, and H. Zhang, "Generative learning for imbalanced data using the Gaussian mixed model," *Applied Soft Computing Journal*, vol. 79, pp. 439–451, Jun. 2019, doi: 10.1016/j.asoc.2019.03.056.
- [29] M. Javvad ur Rehman, S. C. Dass, and V. S. Asirvadam, "A weighted likelihood criteria for learning importance densities in particle filtering," *EURASIP J Adv Signal Process*, vol. 2018, no. 1, pp. 1–19, Dec. 2018, doi: 10.1186/s13634-018-0557-5.
- [30] R. Gurajala, P. B. Choppala, J. S. Meka, and P. D. Teal, "A fast and unbiased minimalistic resampling approach for the particle filter," in *2021 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, Kuala Terengganu, Malaysia, Sep. 2021, pp. 227–232.
- [31] C. Liu, Y. Wang, D. Zhou, and X. Shen, "Minimum-Variance Unbiased Unknown Input and State Estimation for Multi-Agent Systems by Distributed Cooperative Filters," *IEEE Access*, vol. 6, pp. 18128–18141, Mar. 2018, doi: 10.1109/ACCESS.2018.2815662.
- [32] "UCSD Anomaly Detection Dataset." [Online]. Available: www.svcl.ucsd.edu/projects/anomaly/dataset.htm
- [33] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: The CLEAR MOT metrics," *EURASIP J Image Video Process*, vol. 2008, pp. 1–10, 2008, doi: 10.1155/2008/246309.
- [34] Y. Park, L. M. Dang, S. Lee, D. Han, and H. Moon, "Multiple object tracking in deep learning approaches: A survey," *Electronics (Switzerland)*, vol. 10, no. 19, pp. 1–31, Oct. 2021, doi: 10.3390/electronics10192406.
- [35] S. Feng, J. Keung, P. Zhang, Y. Xiao, and M. Zhang, "The impact of the distance metric and measure on SMOTE-based techniques in software defect prediction," *Inf Softw Technol*, vol. 142, pp. 1–14, Feb. 2022, doi: 10.1016/j.infsof.2021.106742.
- [36] X. Yang and L. J. Latecki, "Weakly Supervised Shape Based Object Detection with Particle Filter," in *2010 11th European Conference on Computer Vision (ECCV)*, Springer Berlin Heidelberg, 2010, pp. 757–770.

Authors' Profiles



Avinash Ratre received his Ph.D. in Electronics and Communication Engineering from the Indian Institute of Technology, Roorkee, India. Presently, he is working as an Assistant Professor in the Department of Electronics and Communication Engineering at Delhi Technological University, Delhi, India (ORCID ID: <https://orcid.org/0000-0001-9803-2665>). His research interests include computer vision, machine learning, signal processing, and wireless communication. He can be contacted at email: avinashratre@dtu.ac.in.

How to cite this paper: Avinash Ratre, "GMM-based Imbalanced Fractional Whale Particle Filter for Multiple Object Tracking in Surveillance Videos", *International Journal of Computer Network and Information Security(IJCNIS)*, Vol.17, No.2, pp.34-50, 2025. DOI:10.5815/ijcnis.2025.02.03