*Article*

# UAV Complex-Scene Single-Target Tracking Based on Improved Re-Detection Staple Algorithm

Yiqing Huang [1,2], He Huang [1,2], Mingbo Niu [3,*], Md Sipon Miah [3,4], Huifeng Wang [1] and Tao Gao [5]

1   School of Electronics and Control Engineering, Chang'an University, Xi'an 710064, China; 2021132005@chd.edu.cn (Y.H.); huanghe@chd.edu.cn (H.H.); hfwang@chd.edu.cn (H.W.)
2   Xi'an Key Laboratory of Intelligent Expressway Information Fusion and Control, Chang'an University, Xi'an 710064, China
3   IVR Low-Carbon Research Institute, School of Energy and Electrical Engineering, Chang'an University, Xi'an 710064, China; mmiah@ing.uc3m.es
4   Department of Signal Theory and Communications, University Carlos III of Madrid, Leganes, 28903 Madrid, Spain
5   School of Information Engineering, Chang'an University, Xi'an 710064, China; gaotao@chd.edu.cn
*   Correspondence: ivr.niu@chd.edu.cn

**Abstract:** With the advancement of remote sensing technology, the demand for the accurate monitoring and tracking of various targets utilizing unmanned aerial vehicles (UAVs) is increasing. However, challenges such as object deformation, motion blur, and object occlusion during the tracking process could significantly affect tracking performance and ultimately lead to tracking drift. To address this issue, this paper introduces a high-precision target-tracking method with anomaly tracking status detection and recovery. An adaptive feature fusion strategy is proposed to improve the adaptability of the traditional sum of template and pixel-wise learners (Staple) algorithm to changes in target appearance and environmental conditions. Additionally, the Moth Flame Optimization (MFO) algorithm, known for its strong global search capability, is introduced as a re-detection algorithm in case of tracking failure. Furthermore, a trajectory-guided Gaussian initialization technique and an iteration speed update strategy are proposed based on sexual pheromone density to enhance the tracking performance of the introduced re-detection algorithm. Comparative experiments conducted on UAV123 and UAVDT datasets demonstrate the excellent stability and robustness of the proposed algorithm.

**Keywords:** moth-flame optimization; re-detection; remote sensing; staple; target tracking

## 1. Introduction

Recent advancements in the field of computer vision, specifically unmanned aerial vehicle (UAV) target-tracking technology, have promoted developments in tracking accuracy and speed. These developments have far-reaching implications, encompassing applications such as disaster detection, remote sensing inspection, traffic management, and agricultural protection [1–7]. However, effectively maintaining tracking performance remains a significant challenge in highly challenging task scenarios involving scale variations, low resolutions, and partial occlusions. Moreover, the changes in UAV flight attitude and camera shaking can result in target deformation, impacting tracking quality. Therefore, researching efficient and robust target tracking algorithms holds widespread significance for UAV applications.

Consequently, deep-learning-based methods have garnered significant attention in the field of image processing owing to their exceptional performance [8]. Neural network technology has also led to notable progress in target tracking [9,10]. However, these methods require high computational power, creating obstacles in fulfilling the real-time target tracking requirements on UAV platforms. In contrast, the correlation filter-based

tracking methods can achieve efficient computation. In recent years, many researchers have made significant achievements in the field of object tracking by adopting correlation filter methods [11–15]. The basic principle of correlation filter tracking is to use the Fourier transform to calculate the correlation between the state space model and the target candidate region in the frequency domain, select the location of the maximum response value in the candidate region as the location of the tracked object at the current moment, and continuously track the target by repeating this process.

Bolme et al. [16] proposed the MOSSE algorithm, which, for the first time, introduced the target appearance-adaptive correlation filter method to the field of target tracking. Henriques et al. [17] proposed a tracking algorithm based on a circular tracking structure using fast Fourier transform and kernel methods, designing the CSK tracker based on detection tracking with kernel functions. Subsequently, they combined the multi-channel histogram of oriented gradients with the MOSSE algorithm to establish the kernel correlation filter KCF [18], significantly improving tracking performance. Martin et al. [19] introduced a discriminative scale space representation and proposed the DSST algorithm, which enables precise scale estimation in visual tracking by adapting to changes in target scale. However, when confronted with tracking drift issues arising from interferences like target occlusion or scale variations during long-term UAV tracking, there remains large room for enhancing the failure re-detection capabilities of various correlation filter trackers.

In view of such problems, this paper proposes a long-term UAV target tracking algorithm that is more suitable for challenges such as target deformation and partial occlusion. Firstly, an adaptive feature fusion strategy is proposed to address the limitations of the sum of template with pixel-wise learners (Staple) algorithm [20] in achieving the optimal fusion of the advantages of two tracking models due to fixed weights in position filtering. In addition, this paper has introduced a re-detection indicator to indicate the confidence level of tracking. Moreover, an improved multi-strategy moth flame optimization (MFO) algorithm is integrated into the tracking framework as a re-detection algorithm. In cases where the tracking result is deemed unreliable, the re-detection algorithm is employed to adjust the tracking target, ensuring long-term stability and resistance against tracking drift. In conclusion, the main contributions of this work can be summarized as follows.

1. Based on the improved Staple algorithm, a novel re-detection target-tracking framework is proposed to achieve the long-term tracking of UAVs. In particular, the algorithm adjusts the feature weights adaptively by detecting the response differences between the filter model and the histogram model. Additionally, the improved MFO algorithm is introduced as a re-detection mechanism to enhance the stability of the tracker.
2. A refined swarm intelligence algorithm (MFO) employing diverse strategies is proposed to mitigate tracking failures in target tracking. To swiftly correct inaccuracies in unreliable tracking scenarios, a trajectory-driven population initialization method is advocated. Furthermore, the iteration process of the population's position is enhanced by integrating the influence of sex pheromone concentration on individual moths, thereby optimizing the tracking algorithm's performance.
3. We conducted experiments on well-acknowledged tracking datasets, which demonstrate the outstanding performance of the proposed tracking algorithm. Compared to traditional tracking algorithms, the proposed method exhibits significant improvements in accuracy and robustness, making it effective for tracking aerial photography scenes.

The structure of the subsequent sections in this paper is as follows. Previous research relevant to this study is briefly reviewed in Section 2. Section 3 introduces the staple tracker framework and its related algorithms. The analyses of experiments and results based on the benchmark datasets UAV123 and UAVDT are presented in Section 4. The final conclusions of the algorithm and future task directions are given in Section 5.

## 2. Related Works

In this section, we briefly review the recent progress of relevant filtering algorithms and meta-heuristic algorithms for target-tracking-related applications.

### 2.1. Correlation Filtering Tracking Method

Since the introduction of the MOSSE filter algorithm by Bolme et al. [16], discriminative correlation filter-based target tracking algorithms have undergone extensive development. Li et al. [21] proposed multi-feature fusion to enhance the performance of the tracker and utilized a diverse sample scale pool to achieve adaptive target tracking, effectively addressing the fixed sample size issue in traditional trackers. Galoogahi et al. [22] addressed the inefficiency of using background information in correlation filters by incorporating high-confidence positive and negative samples into the learning and detection process, further enhancing the filter's ability to discriminate between background and target. To address tracking drift caused by environmental changes, Zhang et al. [14] proposed an improved discriminative correlation filter (DCF) method, proposing a learning model with a regularization term that adjusts the position of the target object by computing the difference in the feature representations of the target object in adjacent frames. Deng et al. [23] improved DCF by utilizing dynamic spatial regularization weights and the alternative direction method of multipliers to suppress interfering factors. Kumar et al. [24] established a tracking model based on discriminative correlation filters and a motion estimation model employing Kalman filtering while concurrently integrating CH matrices for motion position compensation of the tracked targets.

### 2.2. Meta-Heuristic Algorithm Solution

From the point of view of numerical optimization, the purpose of target tracking is to leverage the shape and appearance information calculated by the mathematical model in the previous frame to pinpoint the location of the tracked object when the cost function converges to the optimal value within the high-dimensional search space in the current frame. By applying swarm intelligence algorithms, one may consider each individual within the population as a potential tracking solution. Through iterative processes and movement strategies, these individuals gradually converge toward the optimal solution, thereby achieving accurate tracking results.

In recent years, there has been a growing trend among researchers to adopt swarm intelligence methods to enhance the accuracy and robustness of target tracking. Gao et al. [25] proposed a firefly algorithm for target tracking, establishing a generic optimization-based tracking framework. Ong et al. [26] introduced an improved flower pollination algorithm for motion target tracking, with experimental results on benchmark test videos indicating excellent performance. Castro et al. [27] proposed a stochastic frog leaping algorithm for dynamic optimization problems, which delimited the solution space by the double exponential smoothing to achieve target tracking.

Additionally, researchers have explored combining the advantages of different swarm intelligence algorithms to improve the performance of trackers. For example, Sardari et al. [28] integrated particle filtering with an enhanced galactic-based search algorithm to propose an occlusion-free object tracking method, which is capable of handling variations in object appearance and occlusion detection based on appearance models. Kang et al. [29] combine the advantages of particle swarm optimization and gravitational search algorithms to propose a novel hybrid gravitational search algorithm with integrated convolutional neural network features to achieve superior performance in online target tracking. Moghaddasi et al. [30] introduced a reduced particle filter based on a genetic algorithm to address the issues of sample impoverishment and target occlusion.

## 3. Proposed Approach

The overall execution flow of the algorithm proposed in this paper is illustrated in Figure 1. After obtaining the target information and image feature at the $t-1$ frame, the

Staple algorithm [20] uses both the correlation filter and color histogram filter to obtain the target response values. These responses are then fused using adaptive feature fusion coefficients that vary with the environment. The target position is predicted based on the maximum fused response. Determining the tracking status by assessing the confidence of detection results, and if it is below a threshold defined by the algorithm, the re-detection algorithm is employed to search for the target solution with the highest fitness based on the image features of the previous frame in the given solution space.
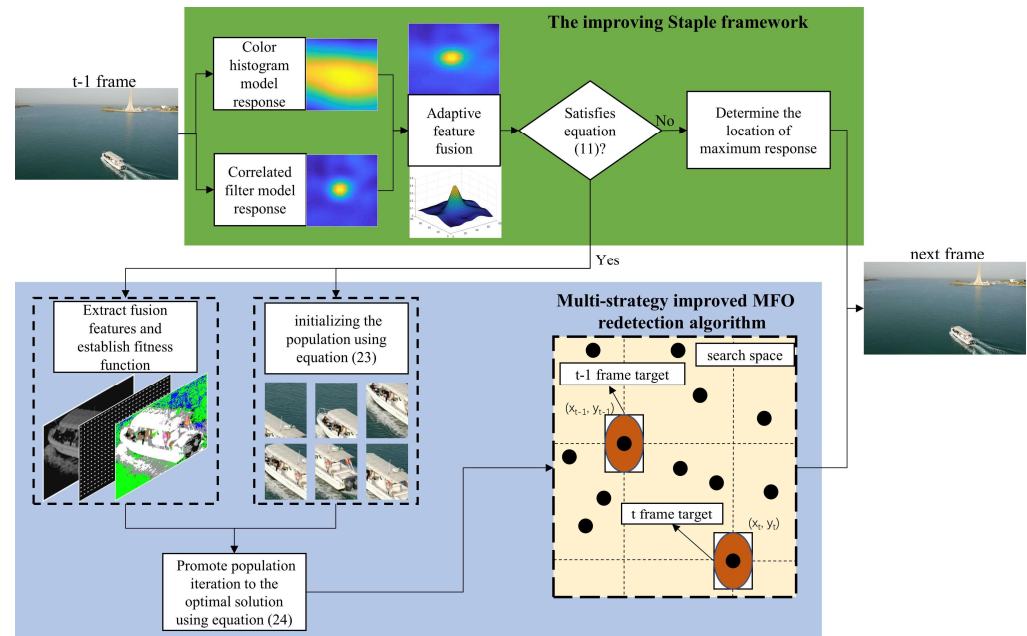


**Figure 1.** The overall structure of the proposed tracking algorithm.

### 3.1. Tracking Algorithm and Its Improvement

Section 3.1.1 gives the implementation details of the Staple algorithm for target tracking. However, due to the boundary effects caused by the circular shift operation during the tracking process, this algorithm may degrade the long-term tracking performance of the UAV and still has limitations. The proposed improvements on adaptive feature weight and tracking confidence detection mechanism will be discussed in Sections 3.1.2 and 3.1.3, respectively.

### 3.1.1. Staple Filtering Algorithm

In the Staple algorithm, target position prediction is achieved through the joint training of two tracking models: the correlation filter and the color histogram.

$$response = (1 - \alpha) \cdot response\_cf + \alpha \cdot response\_pwp \tag{1}$$

where $\alpha$ is the interpolation parameter, and $response\_cf$ and $response\_pwp$ are the filter response score and color histogram response score, respectively.

1. Correlation filter model based on HOG feature

To generate training samples, image patches containing the target and surrounding information are collected, followed by performing circular shifting operations. The $d$-dimensional multi-channel HOG features of the target and its surrounding area are extracted, and the correlation filter $h$ is obtained by solving the target model with a ridge regression equation. The optimal filter $h$ is obtained by minimizing the following loss function:

$$\varepsilon = \left\| \sum_{l=1}^{d} h^l * f^l - g \right\|^2 + \lambda \sum_{l=1}^{d} \left\| h^l \right\|^2 \tag{2}$$

where $h^l$ represents the filter in each dimension, $f^l$ represents the *l*-th dimensional HOG feature vector, $g$ is the expected output associated with $f^l$, and $\lambda$ is the weight of the regularization coefficient. The optimal solution can be obtained by converting the above formula to the frequency domain by Fourier transform as follows:

$$H^l = \frac{\overline{G} \odot F^l}{\sum\limits_{l=1}^{d} \overline{F}^l \odot F^l + \lambda} \tag{3}$$

where the uppercase form of the letter represents the discrete Fourier transform value of the corresponding parameter, the symbol $\odot$ signifies element-wise multiplication, and $\overline{G}$ represents the complex conjugation of the corresponding parameter.

The filter model adopts the linear difference mode to update the filter of each frame, and the calculation method is as follows:

$$\begin{cases} A^{t+1} = (1 - \eta)A^{t-1} + \eta A^t \\ B^{t+1} = (1 - \eta)B^{t-1} + \eta B^t \end{cases} \tag{4}$$

where $\eta$ is the learning rate, and the response score of the new frame $y$ is calculated by Formula (5).

$$y = F^{-1} \left\{ \frac{\sum\limits_{l=1}^{d} \overline{A}^l Z^l}{B + \lambda} \right\}. \tag{5}$$

2. Bayesian classifier model based on color histogram feature

To enhance the description of an RGB image, in this paper, we divide the pixel values into $M$ intervals, treating each interval as a feature dimension. By extracting color features from the three channels, one can use linear regression to independently establish a loss function for all feature pixels through the following:

$$h_{hist}(\psi, p, \beta) = \text{argmin}_h \{ \sum\limits_{j=1}^{M} \left( \frac{N^j(O)}{|O|} (h^j - 1)^2 + \right. \tag{6}$$
$$\left. \frac{N^j(B)}{|B|} (h^j)^2 \right) + \frac{1}{2} \lambda_{hist} \|h\|^2 \}$$

where $\psi$ represents the extracted color histogram features from the target and background images, $p$ represents the ideal response where the target region has a value of 1 and the background region has a value of 0, and $|O|$ and $|B|$ represent the total number of pixels in the foreground slice and background slice, respectively. Solving the above formula, a Bayesian classifier for the $j$ feature can be obtained as follows:

$$\beta^j_{hist} = \frac{\rho^j(O)}{\rho^j(O) + \rho^j(B) + \lambda_{hist}} \tag{7}$$

where $\rho^j(A)$ represents the ratio of the sum of pixels in the $j$-th square column in region $A$ to the total number of pixels in sampling region $A$. Therefore, the following formula is obtained to update the color histogram model with a fixed learning rate $\eta$ as

$$\begin{cases} \rho_{t+1}(O) = (1 - \eta)\rho_{t-1}(O) + \eta\rho_t(O) \\ \rho_{t+1}(B) = (1 - \eta)\rho_{t-1}(B) + \eta\rho_t(B) \end{cases} \tag{8}$$

### 3.1.2. Anomaly Tracking Status Detection

When tracking targets against a complex and changing background in low-altitude airspace, factors like background clutter, target occlusion, and low resolution can cause disturbances in the target's position and motion, leading to tracking drift and affecting the updating of the feature model, thus resulting in tracking failures. Therefore, this paper adopts two types of re-detection metrics, namely, the maximum response peak

and the average peak-to-correlation energy (APCE), proposed by Wang [31], to reflect the confidence evaluation of the response of the correlation filter model in tracking. The criterion for determining whether the tracking state is anomaly is reflected by the magnitude of the re-detection indicator.

The maximum response peak value can be expressed as

$$F_{max} = \max F(x', y') \tag{9}$$

where $F(x', y')$ is the response of horizontal displacement $x'$ and vertical displacement $y'$ in the detection range of the target.

The expression of the *APCE* value is

$$APCE = \frac{|F_{max} - F_{min}|^2}{\text{mean}(\sum\limits_{x,y}(F(x', y') - F_{min})^2)} \tag{10}$$

where $F_{min}$ is the minimum response value within the detection range of the target.

By comparing Figure 2a,b, it can be observed that after the tracked target undergoes deformation, the response changes from a single peak that smoothly rises from the center to the appearance of side lobes, and the contrast between the main peak and the side lobes decreases. As shown in Figure 2c, it is observed that the APCE value progressively increases during the initial phase of the tracking process, indicating the high confidence and stable performance of the Staple algorithm in detecting objects amidst background interference. However, a notable decrease and fluctuation in the APCE is evident during the latter phase of the tracking process. This suggests that the Staple algorithm, employing a frame-by-frame update strategy, is susceptible to drifting in scenarios involving occlusion or motion blur, necessitating further refinement.
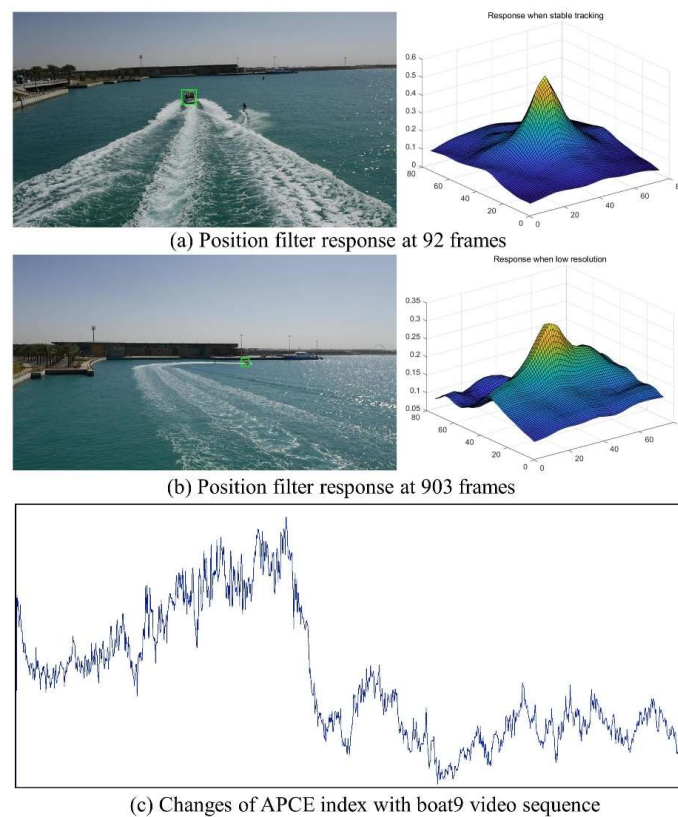


(a) Position filter response at 92 frames



(b) Position filter response at 903 frames



(c) Changes of APCE index with boat9 video sequence

**Figure 2.** Response and APCE results during the tracking process by Staple algorithm.

The response confidence degree $R_\alpha$ of the improved filtering algorithm is

$$R_\alpha = \begin{cases} 1, APCE^t > \frac{\lambda_{APCE}}{t-1}\sum_{i=1}^{t-1} APCE_i \cap F_{max}^t > \frac{\lambda_F}{t-1}\sum_{i=1}^{t-1}(F_{max})_i \\ 0, \text{ else} \end{cases} \tag{11}$$

where $\lambda_{APCE}$ and $\lambda_F$ represent the respective adjustment coefficients of the two types of confidence judgments, respectively. If $R_\alpha = 1$, it signifies that the results calculated by the correlation filter model are reliable.

### 3.1.3. Adaptive Feature Fusion Strategy

In the Staple algorithm, the fixed-weight tracking method shown in Formula (1) can easily cause the tracking algorithm to ignore the correct prediction results of low-weight features. To address this challenge, an adaptive feature fusion strategy is suggested in this study. This approach entails dynamically adjusting the weights assigned to primary and secondary features in response to variations in the target's appearance or environmental conditions. Consequently, the algorithm demonstrates improved performance in tracking challenges such as appearance distortions and fluctuations in lighting conditions during UAV target tracking tasks.

During the tracking process, based on the tracking quality evaluation index APCE, the reliability evaluation of the tracking feature response results derived from the filter model and the histogram model can be realized. To maximize the advantages of each type of feature in the following frame, the weights can be used to adaptively adjust the role of features in the fusion template as

$$\omega_{pwp} = \frac{F_{max}^{pwp} \cdot APCE^{pwp}}{F_{max}^{pwp} \cdot APCE^{pwp} + F_{max}^{cf} \cdot APCE^{cf}} \tag{12}$$

where $F_{max}^{pwp}$ and $F_{max}^{cf}$ represent the maximum value of the response of the filter model and the histogram model, respectively. Finally, the weight of the model is updated based on the learning rate via

$$\omega_{pwp}^t = (1 - \delta) \cdot \omega_{pwp}^{t-1} + \delta \cdot \omega_{pwp} \tag{13}$$

and

$$\omega_{cf}^t = 1 - \omega_{pwp}^t. \tag{14}$$

The parameter $\delta$, which represents the weight learning rate during the tracking process, is predefined as 0.045. Therefore, the improved response is

$$response = \omega_{cf}^t \cdot response\_cf + \omega_{pwp}^t \cdot response\_pwp. \tag{15}$$

### 3.2. Object Re-Detection Algorithm and Improvement
### 3.2.1. Moth Flame Optimization

As a population-based intelligent algorithm, the Moth Flame Optimization (MFO) algorithm leverages the collective intelligence of individuals within the population to effectively explore solution spaces. By incorporating a spiral search paradigm and adaptive adjustment capabilities, MFO demonstrates remarkable robustness in tackling intricate optimization challenges. This unique combination of features has led to its widespread adoption across various domains, making it a valuable tool for addressing complex real-world problems [32,33]. In this paper, the MFO algorithm is selected as the basis of the tracking framework. When the tracking result is considered unreliable at some time, this algorithm is used as the re-detection algorithm.

The moth $M$ in the algorithm symbolizes the candidate solutions for the optimization problem, while the flame $F$ denotes the fitness of these candidates. During the positional iteration of the MFO algorithm, the logarithmic spiral function is employed to simulate the motion of moths.

$$M_i = D_i^{k-1} \cdot e^{b\tau} \cdot \cos(2\pi\tau) + F_j^{k-1} \tag{16}$$

where $D_i^{k-1}$ represents the Euclidean distance between the *i*-th moth and the *i*-th flame at the $k-1$ iteration, *b* is the constant for simulating the spiral motion of the moth logarithm, and $\tau$ is the random number between $[r, 1]$ with *r* defined as

$$r(k) = -1 - \frac{k}{K} \tag{17}$$

where *K* represents the maximum iteration count, and *k* denotes the current iteration count with $k = (1,2,\ldots,K)$. The iterative process of the moth individual approaching the center of the flame is depicted in Figure 3 as a spiral curve, with each blue dot representing the position of the moth at each iteration.
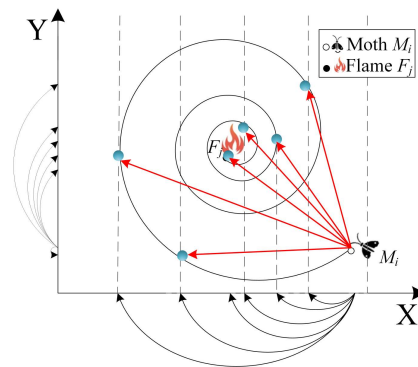


**Figure 3.** The spiral update of the moth towards the flame.

Meanwhile, to prevent the algorithm from being constrained by locally optimal solutions during the generation process and ensure the efficiency of searching for the optimal solution, the number of flames can be dynamically reduced by the number of iterations through

$$n_f(k) = \text{round}(n - \frac{k}{K}(n-1)) \tag{18}$$

where $n_f(k)$ is the maximum number of flames in the *k*-th iteration, and round(.) denotes the integer operation.

3.2.2. Feature Template Extraction and Update

After obtaining the position information of the tracked object by fusing the response results based on two-class filters in the Staple algorithm, this paper introduces an approach that incorporates color name (CN) features, gray features, and the fast histogram of oriented gradients (FHOG) features to extract feature information of tracked objects under diverse tracking backgrounds. Additionally, during the tracking process, the algorithm may contaminate the feature information of the currently tracked object due to factors such as target occlusion. Therefore, this paper combines sparser updating and anomaly tracking status detection schemes to ensure the acquisition of effective feature information under different tracking statuses while maintaining the tracking speed, and the corresponding formula is given as

$$\phi^t = \begin{cases} \alpha_{sta}\phi^t + (1 - \alpha_{sta})\phi^{t-1}, & \text{mod}(t,4) = 0 \cap R_a = 1 \\ \alpha_{ano}\phi^t + (1 - \alpha_{ano})\phi^{t-1}, & \text{mod}(t,4) = 0 \cap R_a = 0 \\ \phi^{t-1}, & \text{else} \end{cases} \tag{19}$$

and

$$\alpha_{ano} = \frac{\gamma_1}{1 + \gamma_2 \exp(\gamma_3(\frac{1}{N}\sum_{i=1}^{N} Fit[M_i] - \gamma_4))} \tag{20}$$

where $\phi^t$ represents the feature information template updated at time *t* of the current frame, $\alpha_{sta}$ and $\alpha_{ano}$ represent the feature learning rate of the template when Staple algorithm and

re-detection algorithm are respectively enabled, $Fit[M_i]$ represents the fitness of the i-th moth, $N$ is the number of moth population, $\alpha_{sta}$ is set to 0.03, and adjustment parameters $\gamma_1$, $\gamma_2$, $\gamma_3$, and $\gamma_4$ are set to 0.04, 0.1, $-45$, and 0.75, respectively. In addition, the dimensions of selected features and the fusion process are shown in Figure 4 below.
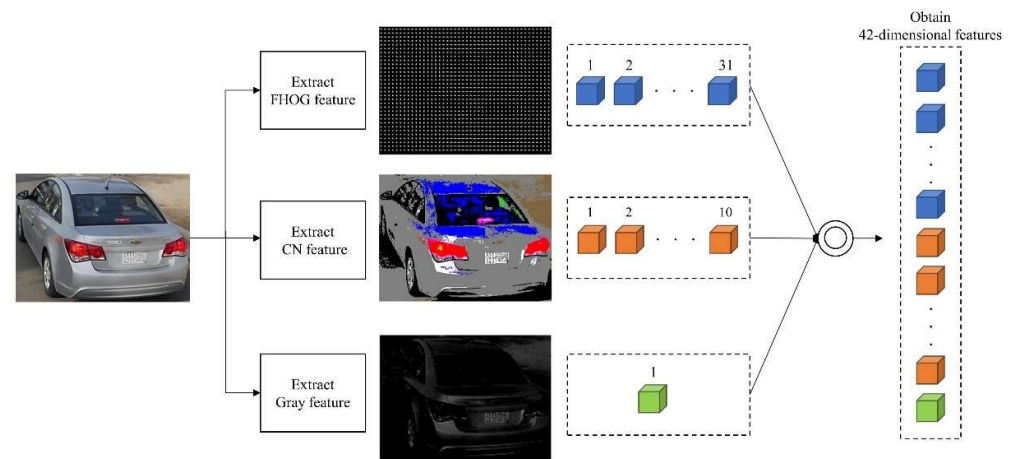


**Figure 4.** The fusion process of tracking target feature information.

### 3.2.3. Establish Fitness Function

During the tracking process in the MFO algorithm, the feature information of the moth individuals' positions during the iterative process is compared with the fused feature template of previous tracking results to calculate the fitness value. This paper adopts the cosine similarity method as the fitness calculation formula, measuring the difference between two individuals by computing the cosine value of the angle between two vectors.

### 3.2.4. Trajectory-Guided Gaussian Initialization

In order to address the slow convergence and long-term trapping in local optima issues that could arise from traditional random initialization methods in target tracking, this paper proposes a trajectory-guided Gaussian initialization method that leverages the tracked object's positions in consecutive frames of a video sequence to predict the potential motion trajectory of the target. This method utilizes a Gaussian distribution initialization approach to guide the population to be located as close as possible to the predicted area of the target during the initial stage, thereby shortening the time required for global optimization. The moth $M_i^t$ meets the multivariate Gaussian distribution, and the initialization formulas are given as follows:

$$v^t = P^{t-1} - P^{t-2}v \tag{21}$$

$$\mu^t = P^{t-1} + v^t \tag{22}$$

and

$$p(M_i^t) = \frac{\exp\left(-\frac{\left(M_i^t - \mu^t\right)^T\left(M_i^t - \mu^t\right)}{2\Sigma^t}\right)}{\sqrt{(2\pi)^t\left|\Sigma^t\right|}}. \tag{23}$$

At time $t$, $v^t$ represents the model's motion speed, $P^t$ denotes the center point of the tracking region obtained by the algorithm at time $t$, and $\Sigma^t$ and $\mu^t$ represent the covariance matrix and mean of the initialized population, respectively. Additionally, the initial weight values provided by the Gaussian distribution enable a few moths to spread across a larger search space, which allows for the continued tracking of the target when encountering unexpected changes such as occlusion or fast movement of the target.

3.2.5. Population Iteration Velocity Dominated by Sex Pheromone Density

The parameter $\tau$ in Formula (16) determines the distance of the moth to the flame after iteration, and the step size is determined solely by the iteration count $k$. Although this method can guarantee a certain degree of global optimization performance in the early stages and local optimization performance in the later stages, it can affect the optimization efficiency of the algorithm. In real-life scenarios, female moths of reproductive age produce and emit sex pheromones to attract male moths for mating. Varying concentrations of these pheromones in the population result in distinct flight speeds and behavioral tendencies among individual moths. Based on this phenomenon, this paper improves the process of population location iteration in MFO algorithm. The new step size is influenced by the fitness of the moth $Fit[M_i]$ and can be calculated by

$$a1 = -2 + \frac{1}{N}\sum_{i=1}^{N} \frac{1}{1 + e^{(-10 \cdot (1-Fit[M_i])+5)}} \tag{24}$$

where $\gamma$ is the random number between [a1, 1], $M_i = D_i^{k-1} \cdot e^{b\gamma} \cdot \cos(2\pi\gamma) + F_i^{k-1}$, and $N$ is the number of moth population.

Algorithm 1 summarizes the main process of the proposed improving Staple framework with re-detection algorithm.

---

**Algorithm 1:** The proposed improving re-detection Staple algorithm

---

**Input:** The initial frame $t_0$ with the corresponding object ground truth bounding box *B0* ($x_0$, $y_0$, $w_0$, $h_0$);

**Output:** The predicted bounding box *Bt* ($x_t$, $y_t$, $w_t$, $h_t$) of the $t$th frame;

1 Initialize the correlation filter model and histogram model for the Staple algorithm;

2 **for** $i$ = 1, 2, ..., $n$ **do**

3　　Extract features of the current frame image to obtain responses from the relevant filter and histogram classifier;

4　　Determine the target position by blending response maps, calculate $APCE^t$ and $F_{\max}^t$ at the current frame moment;

5　　**if** $APCE^t > \frac{\lambda_{APCE}}{t-1}\sum_{i=1}^{t-1} APCE_i \cap F_{max}^t > \frac{\lambda_F}{t-1}\sum_{i=1}^{t-1}(F_{max})_i$ **then**

6　　　　Extract CN, Fhog, and Gray features from the current frame image and merge them into a 42-dimensional feature vector;

7　　　　Build the corresponding feature template at the current frame and establish the fitness function using Equations (19) and (20);

8　　　　Initialize the population in the MFO algorithm using Equation (23);

9　　　　Iterate through the population to obtain the individual with the best fitness as the detected target box;

10　　　　$Bt = Bt_{\text{MFO}}$;

11　　**else**

12　　　Select the position corresponding to the maximum value in the blended response map as the detected target box;

13　　　$Bt = Bt_{\text{Staple}}$;

14　　Update the correlation filter model and histogram model.

---

## 4. Experimental Results

To validate the proposed methodology, a comprehensive experiment was con-ducted, including an assessment of overall performance, evaluation of video sequence attributes performance, and visual evaluation of tracking results. The evaluation was carried out on the widely recognized UAV vision datasets, UAV123 [34] and UAVDT [35], known for their diverse scenarios and target categories. These datasets cover a wide range of scenarios and tracking objects, enabling the evaluation of tracking algorithms across various sizes, appearance changes, and motion patterns of low-altitude UAVs. As representative benchmarks in the field, these datasets ensure a rigorous assessment of the proposed approach.

*4.1. Experiment Setup*

From a perspective of comprehensively assessing the robustness and performance of the proposed algorithm in UAV target tracking tasks, comparative experiments were conducted between the proposed algorithm, the proposed re-detection Staple algorithm without improvement strategy (Staple-MF) and mainstream tracking algorithms (EMCF [14], IBRI [15], Staple [20], SAMF [21], MSCF [36], ARCF [37], SRDCF [38], AutoTrack [39], CSR-DCF [40], CACF [41], ReCF [42], and DRCF [43]) in testing experiments on open datasets. We used a testing platform with the following hardware specifications: 11th Gen Intel(R) Core(TM) i7-11800H 2.30 GHz processor, RTX3050 GPU, and 16 GB of RAM. The software platform was Matlab R2019a. For the purpose of balancing operational speed and tracking precision in the target re-detection algorithm, the number of individual moths in the population is set to 25 with a maximum iteration count of 25.

In this study, the evaluation criteria for the tracking effectiveness of various algorithms on test datasets are based on two important indicators: success rate and accuracy.

① Precision: Center location error (CLE) represents the ratio of the distance between the algorithm-calibrated coordinates of the target's center point and the real target center point coordinates in the frame, as calculated by the tracker, and is below a certain threshold. In this paper, the threshold is set to 20. The CLE is defined as

$$CLE = \sqrt{(x_{track} - x_{truth})^2 + (y_{track} - y_{truth})^2} \tag{25}$$

where $(x_{\text{track}}, y_{\text{track}})$ represents the central point coordinates of the target calculated by the tracker, and $(x_{\text{truth}}, y_{\text{truth}})$ represents the central point coordinates of the real target area.

② Success rate: Defining the success rate of the target tracking task as the proportion of frames in which the intersection over union (IoU) value between the predicted bounding box, calculated by the tracker, and the ground truth bounding box exceeds a specified threshold. The IoU per frame is given by

$$IoU = \left| \frac{I_{truth} \cap I_{track}}{I_{truth} \cup I_{track}} \right| \tag{26}$$

where $I_{\text{truth}}$ is the actual target region, and $I_{\text{track}}$ is the target region calculated by the algorithm.

*4.2. Quantitative Experimental Results*

4.2.1. UAV123 Benchmark

The UAV123 dataset consists of 123 low-altitude UAV video sequences taken from different angles and heights, covering a variety of tracking scenes and objects. In addition, the video sequence in the dataset contains a total of 12 attributes, such as illumination change, partial occlusion, etc., which provides valuable information for analyzing and understanding the experimental results in the dataset.

(1) Overall evaluation

As illustrated in Figure 5, the outcomes of the proposed algorithm (OURS) compared to other trackers in the OPE evaluation on the UAV123 dataset. The figure illustrates that the precision and success rate of the algorithm proposed in this paper are 0.709 and 0.472, respectively, which has the best accuracy among all the comparisons. The algorithm proposed in this paper is 2.0% better than the previous most advanced tracker AutoTrack in terms of precision. Compared to the Staple tracker, OURS achieves performance improvements of 4.3% in precision and 2.7% in success rate, respectively. These results can be attributed to the inclusion of the adaptive feature fusion strategy, which enhances the utilization of dominant weights in the algorithm. Additionally, through the re-detection mechanism for abnormal states, one can enable the improved MFO algorithm with Gaus-

sian initialization based on target motion speed and dynamic adjustment of population iteration rate, ensuring effective tracking in challenging scenarios.
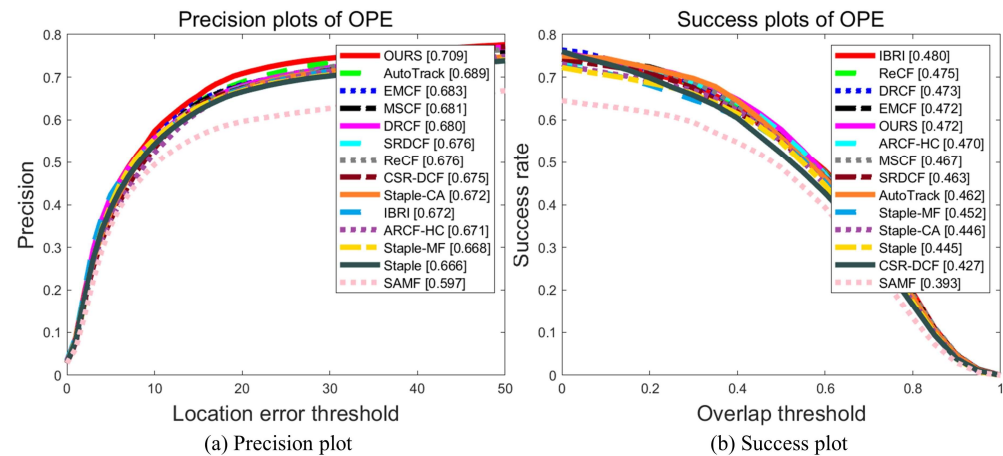


**Figure 5.** Precision and success comparison of various trackers on the UAV123 dataset.

(2) Attribute evaluation

In order to measure the performance improvement of the proposed algorithm in different scenarios and analyze the generalization ability of the trackers, we have also utilized the UAV123 dataset to perform an attribute-based comparison between the proposed algorithm and other trackers. The precision and success rate values for each attribute category, as well as the corresponding tracker data, are presented in Figures 6 and 7. The figure demonstrates that the improved algorithm we proposed exhibits a performance enhancement of more than 10% across all 12 attributes when compared to the Staple tracker. The adaptive weight fusion of features has enhanced the description capability of the target appearance to a certain extent, enabling OURS to better maintain tracking stability when facing challenges such as aspect ratio changes and viewpoint changes that cause deformation. Additionally, in the presence of occlusions, the anomaly tracking status detection and recovery strategy can maximize the avoidance of tracking drift.

4.2.2. UAVDT Benchmark

The UAVDT dataset consists of more than 100 video sequences composed of 80,000 images taken by the drone platform, including a variety of scenes, such as toll booths, intersections, and main roads, with a focus on vehicle tracking in drone movement scenarios. Compared with the UAV123 dataset, this dataset can realize the performance comparison of eight attributes in different scenarios in single target tracking.

(1) Overall evaluation

Figure 8 presents the results of our proposed algorithm compared to other trackers in the OPE evaluation on the UAVDT dataset. The precision and success rates were 0.735 and 0.480, marking a 5.7% and 7.8% increase, respectively, compared to the Staple baseline tracker. In terms of precision, OURS performance ranks among the top in comparison to various advanced trackers. As to success rate, OURS has an 8.0% performance improvement compared with the Staple-CA algorithm. The incorporation of a re-detection mechanism in the Staple algorithm enables the proposed algorithm to effectively address environmental interference and target deformation. This is substantiated by the notable improvements in target position prediction accuracy and scale estimation performance observed in the dataset calculations.
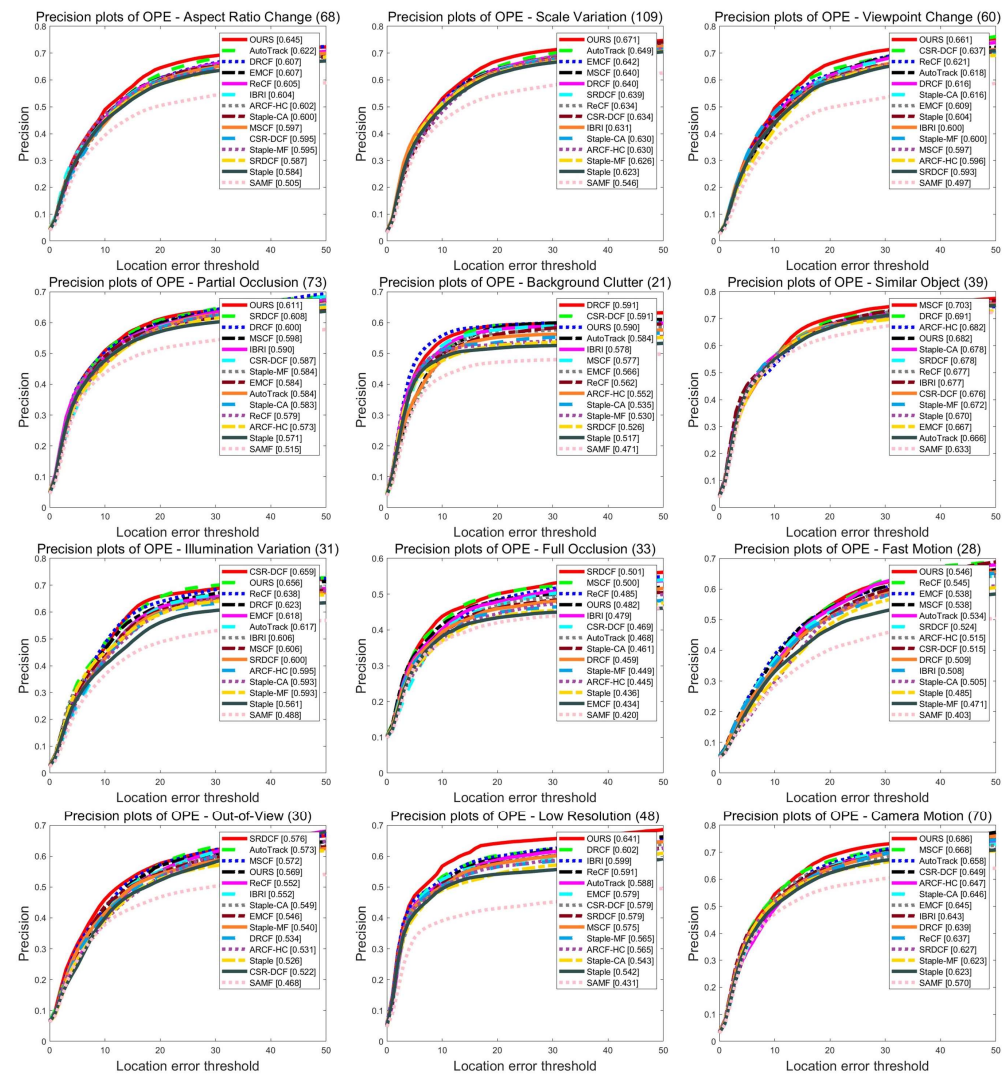
**Figure 6.** Precision comparison of 12 attributes for the UAV123 dataset.

(2) Overall evaluation

Figure 9 visually depicts a performance comparison of different trackers across the eight attributes of UAVDT. As depicted in the figure, the calculations presented in this article exhibit superior performance compared to the Staple baseline tracker across all attributes. Furthermore, OURS achieves the highest precision and success rate rankings in the SV, OM, IV, and OB attributes.

*4.3. Qualitative Experimental Results*

To empirically evaluate the efficacy of the proposed algorithm for UAV tracking in diverse and complex scenarios, this study conducted qualitative analysis on four representative video sequences from UAV123 and UAVDT datasets, encompassing maritime vessels, nocturnal vehicular movement, fast-moving characters, and long-term vehicular tracking. The categories of challenge attributes in various video sequences are provided in Table 1, and the visualization results of the five trackers in the video sequence are shown in Figure 10.
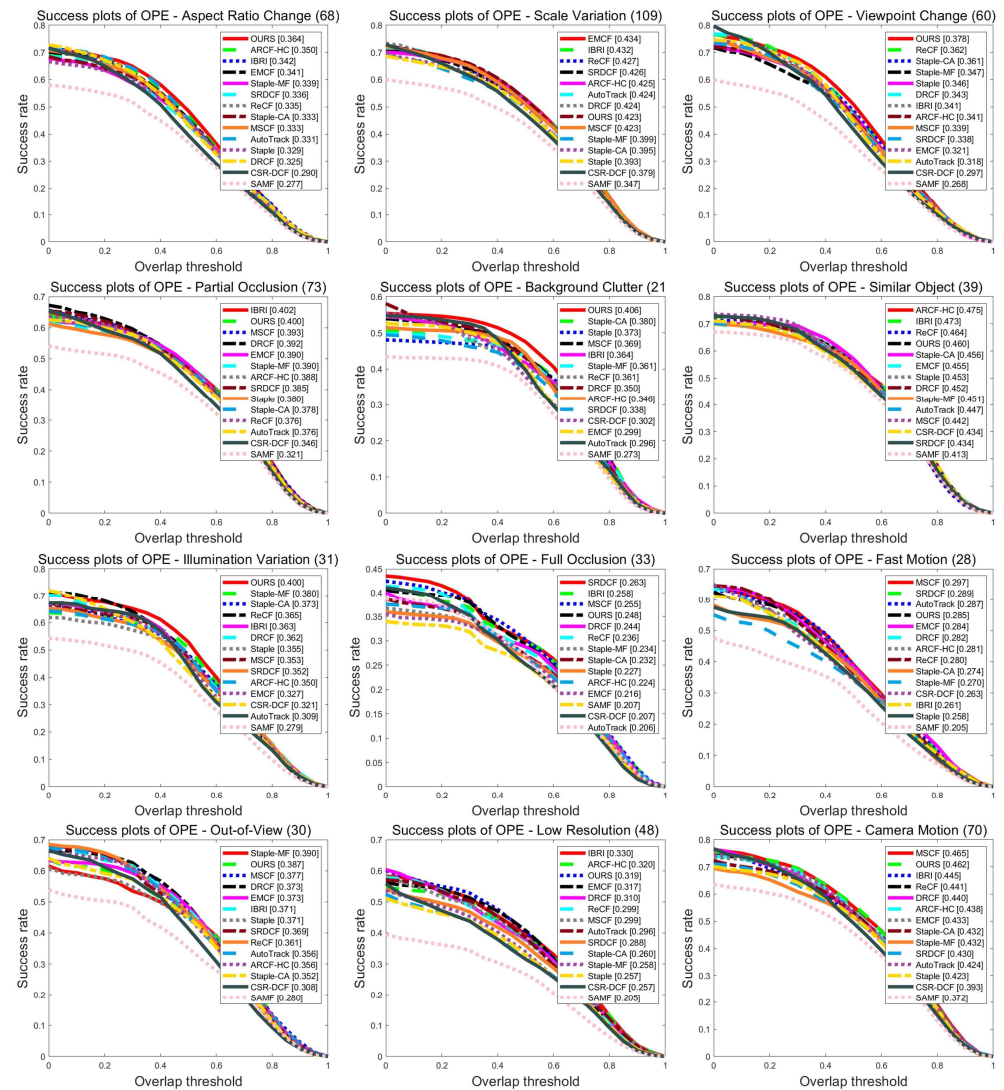
**Figure 7.** Success comparison of 12 attributes for the UAV123 dataset.
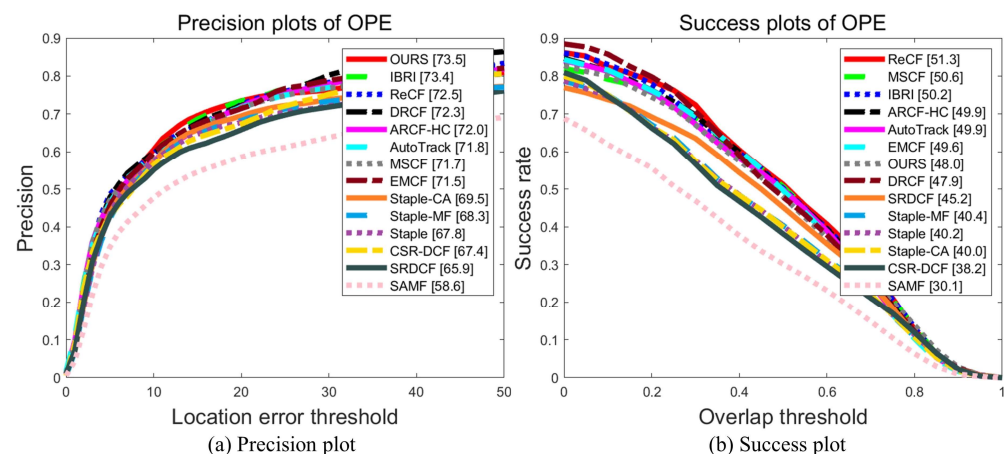


(a) Precision plot

(b) Success plot

**Figure 8.** Precision and success comparison of various trackers on the UAVDT dataset.

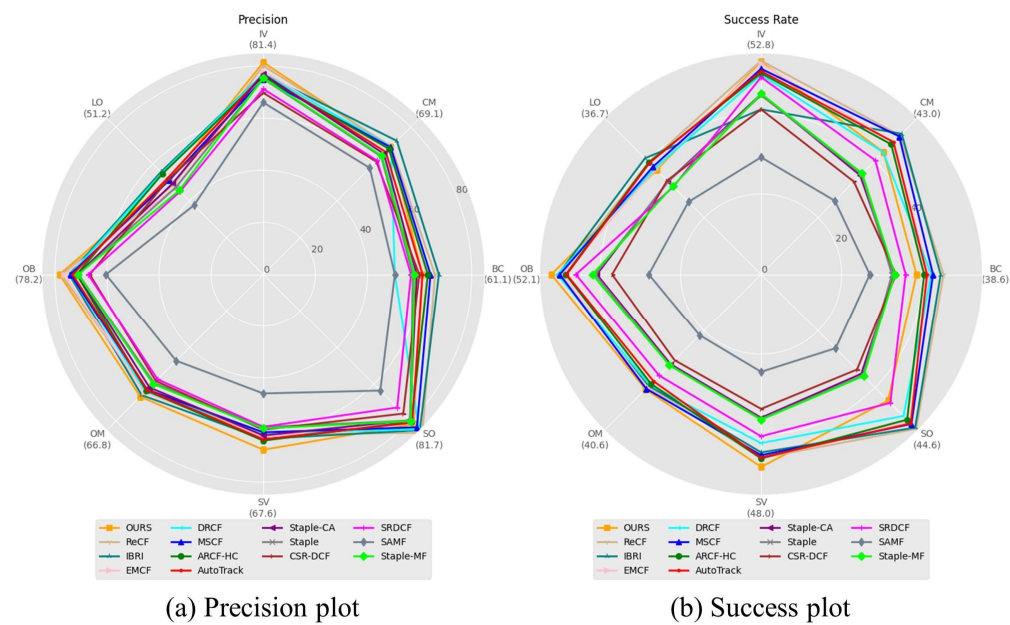(a) Precision plot        (b) Success plot

**Figure 9.** Evaluation of tracker attributes on the UAVDT dataset.

**Table 1.** Specific attributes of selected datasets video sequence.

| Video Sequence | Attributes |
|---|---|
| boat9 | SV, ARC, LR, POC, VC |
| wakeboard5 | SV, ARC, LR, FM, POC, IV, VC, CM |
| S0801 | BC, CM, OM, SV, LO |
| S1201 | BC, CM, OM, SV, LO, IV |
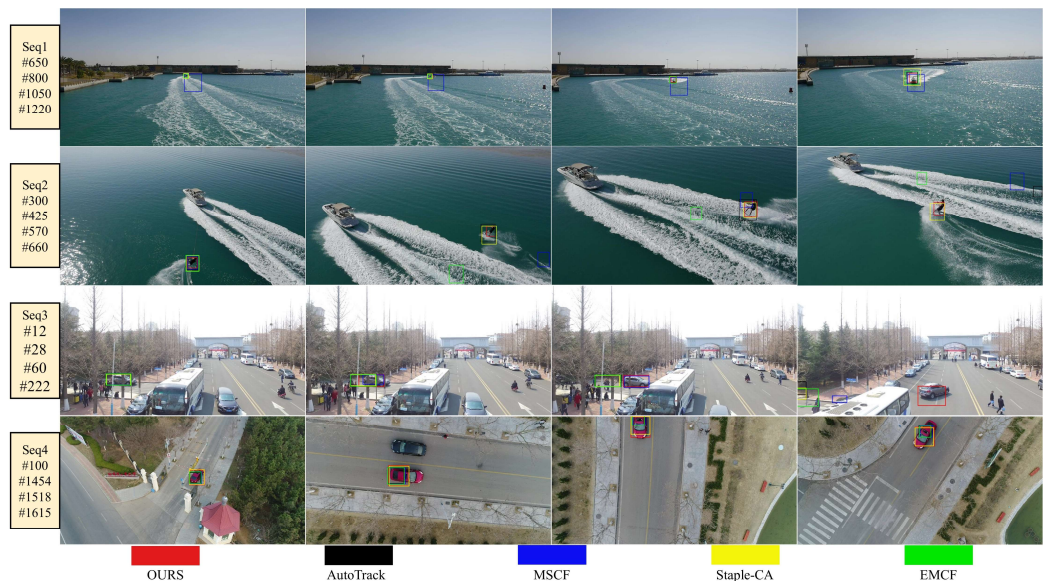


**Figure 10.** Visualization results from five trackers on tracking experiment.

In Seq1, the primary challenge arises from the tracking difficulties induced by the drones' long-distance shooting. As the target kept moving away, the MSCF algorithm experienced scale drift at 650 frames. Furthermore, at 1220 frames, Staple-CA failed to adapt to the scale change over a long time, leading to tracking drift. OURS, AtuoTrack, and EMCF could realize the long-term tracking of tracking targets with ever-changing scales, but only the proposed algorithm could maintain high tracking performance continuously.

In Seq2, the main challenges are fast-moving targets and light interference. At 425 frames, the MSCF and EMCF algorithms failed to distinguish between waves and people. At 660 frames, in addition to OURS, only Staple-CA could continuously track the target, but the Staple-CA algorithm could not calculate the scale transformation of the target more accurately.

In Seq3, the main challenges are background clutter and complete occlusion. As can be seen from frames 28 and 60, background interference made the tracking boxes of the Staple-CA, AutoTrack, and EMCF algorithms barely move, making it impossible to track the vehicle. In frame 222, in addition to our proposed algorithm, the rest of the trackers were unable to recover tracking when the tracking object reappeared.

Seq4 shows a low altitude perspective for the UAV to track the video of car movement and turning. The primary challenges encountered included tracking occlusion and scale changes. As depicted in the figure, all trackers demonstrated stable tracking in the presence of partial occlusion 100 frames prior and with similar targets at 1454 frames. However, at 1518 frames, AutoTrack, MSCF, and EMCF could only track less of the target due to camera rotation. Notably, only the algorithm proposed in this study closely approximated the actual target tracking framework when the tail reappeared at 1615 frames.

## 5. Conclusions

In this work, a novel and high-precision method is introduced for UAV target tracking. This approach enhances the existing Staple algorithm by implementing an adaptive feature fusion strategy to ensure accurate tracking in complex environments. Additionally, two strategies are proposed, namely trajectory-guided Gaussian initialization and population iteration velocity dominated by sex pheromone density, which improve the overall performance of the enhanced MFO tracking framework. To address abnormal tracking scenarios, the enhanced MFO is incorporated as a re-detection algorithm for stable tracking. Through extensive experimental studies, the effectiveness of our proposed algorithm is confirmed with high tracking accuracy and reliable discrimination of target scales in complex scenes. Furthermore, strong adaptability is demonstrated in major remote tracking challenges, such as motion blur and complex backgrounds. In the future research, the impact of remote tracking will be investigated with more comprehensive adverse weather conditions, such as rain and fog, as well as their impacts on the quality of tracking images.

**Author Contributions:** Conceptualization, Y.H. and M.N.; data curation, H.H., M.S.M. and H.W.; methodology, Y.H., M.N. and H.H.; software, Y.H.; validation, Y.H., H.H. and M.N.; formal analysis, M.N.; investigation, H.H. and T.G.; resources, M.S.M. and H.H.; writing—original draft preparation, Y.H. and H.H.; writing—review and editing, M.N.; visualization, Y.H., M.N. and M.S.M.; supervision, M.N.; project administration, M.N., T.G. and M.S.M.; funding acquisition, H.H. and M.N. All authors have read and agreed to the published version of the manuscript.

**Data Availability Statement:** The data presented in this study are available upon request from the corresponding author. The data are not publicly available due to privacy restrictions.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Ruan, W.; Chen, J.; Wu, Y.; Wang, J.; Liang, C.; Hu, R.; Jiang, J. Multi-Correlation Filters with Triangle-Structure Constraints for Object Tracking. *IEEE Trans. Multimed.* **2019**, *21*, 1122–1134. [CrossRef]

2. Rautaray, S.S.; Agrawal, A. Vision Based Hand Gesture Recognition for Human Computer Interaction: A Survey. *Artif. Intell. Rev.* **2015**, *43*, 1–54. [CrossRef]

3. Zhang, J.; Zhang, X.; Huang, Z.; Cheng, X.; Feng, J.; Jiao, L. Bidirectional Multiple Object Tracking Based on Trajectory Criteria in Satellite Videos. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–14. [CrossRef]

4. Yan, H.; Xu, X.; Jin, G.; Hou, Q.; Geng, Z.; Wang, L.; Zhang, J.; Zhu, D. Moving Targets Detection for Video SAR Surveillance using Multilevel Attention Network Based on Shallow Feature Module. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–18. [CrossRef]

5. Li, B.; Fu, C.; Ding, F.; Ye, J.; Lin, F. All-Day Object Tracking for Unmanned Aerial Vehicle. *IEEE Trans. Mobile Comput.* **2023**, *22*, 4515–4529. [CrossRef]

6. Hu, S.; Yuan, X.; Ni, W.; Wang, X.; Jamalipour, A. Visual camouflage and online trajectory planning for unmanned aerial vehicle-based disguised video surveillance: Recent advances and a case study. *IEEE Veh. Technol. Mag.* **2023**, *18*, 48–57. [CrossRef]

7. Gao, G.; Yao, L.; Li, W.; Zhang, L.; Zhang, M. Onboard information fusion for multisatellite collaborative observation: Summary, challenges, and perspectives. *IEEE Geosc. Remote Sens. Mag.* **2023**, *11*, 40–59. [CrossRef]

8. Wen, Y.; Gao, T.; Zhang, J.; Li, Z.; Chen, T. Encoder-free Multi-axis Physics-aware Fusion Network for Remote Sensing Image Dehazing. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–15.

9. Wei, H.; Wan, G.; Ji, S. ParallelTracker: A Transformer Based Object Tracker for UAV Videos. *Remote Sens.* **2023**, *15*, 2544. [CrossRef]

10. Zhang, S.; Zhuo, L.; Zhang, H.; Li, J. Object Tracking in Unmanned Aerial Vehicle Videos via Multifeature Discrimination and Instance-Aware Attention Network. *Remote Sens.* **2020**, *12*, 2646. [CrossRef]

11. Bian, Z.; Xu, T.; Chen, J.; Ma, L.; Cai, W.; Li, J. Auto-Learning Correlation-Filter-Based Target State Estimation for Real-Time UAV Tracking. *Remote Sens.* **2022**, *14*, 5299. [CrossRef]

12. Li, Y.; Fu, C.; Huang, Z.; Zhang, Y.; Pan, J. Intermittent Contextual Learning for Keyfilter-Aware UAV Object Tracking using Deep Convolutional Feature. *IEEE Trans. Multimed.* **2021**, *23*, 810–822. [CrossRef]

13. Zhang, F.; Ma, S.; Yu, L.; Zhang, Y.; Qiu, Z.; Li, Z. Learning Future-Aware Correlation Filters for Efficient UAV Tracking. *Remote Sens.* **2021**, *13*, 4111. [CrossRef]

14. Zhang, F.; Ma, S.; Zhang, Y.; Qiu, Z. Perceiving temporal environment for correlation filters in real-time uav tracking. *IEEE Signal Process. Lett.* **2021**, *29*, 6–10. [CrossRef]

15. Fu, C.; Ye, J.; Xu, J.; He, Y.; Lin, F. Disruptor-aware interval-based response inconsistency for correlation filters in real-time aerial tracking. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 6301–6313. [CrossRef]

16. Bolme, D.S.; Beveridge, J.R.; Draper, B.A.; Lui, Y.M. Visual Object Tracking using Adaptive Correlation Filters. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 2544–2550.

17. Henriques, J.F.; Caseiro, R.; Martins, P.; Batista, J. Exploiting the Circulant Structure of Tracking-by-Detection with Kernels. In Proceedings of the 12th European Conference on Computer Vision, Florence, Italy, 7–13 October 2012; pp. 702–715.

18. Henriques, J.F.; Caseiro, R.; Martins, P.; Batista, J. High-speed Tracking with Kernelized Correlation Filters. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *37*, 583–596. [CrossRef] [PubMed]

19. Danelljan, M.; Häger, G.; Khan, F.; Felsberg, M. Accurate Scale Estimation for Robust Visual Tracking. In Proceedings of the British Machine Vision Conference, Nottingham, UK, 1–5 September 2014; Bmva Press: Newcastle, UK, 2014.

20. Bertinetto, L.; Valmadre, J.; Golodetz, S.; Miksik, O.; Torr, P.H. Staple: Complementary Learners for Real-time Tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1401–1409.

21. Li, Y.; Zhu, J. A Scale Adaptive Kernel Correlation Filter Tracker with Feature Integration. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 5–12 September 2014; pp. 254–265.

22. Kiani Galoogahi, H.; Fagg, A.; Lucey, S. Learning Background-aware Correlation Filters for Visual Tracking. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 1144–1152.

23. Deng, C.; He, S.; Han, Y.; Zhao, B. Learning Dynamic Spatial-temporal Regularization for UAV Object Tracking. *IEEE Signal Process. Lett.* **2021**, *28*, 1230–1234. [CrossRef]

24. Kumar, R.; Deb, A.K. Pedestrian Tracking in UAV Images with Kalman Filter Motion Estimator and Correlation Filter. *IEEE Aerosp. Electron. Syst. Mag.* **2023**, *38*, 4–19. [CrossRef]

25. Gao, M.-L.; Li, L.-L.; Sun, X.-M.; Yin, L.-J.; Li, H.-T.; Luo, D.-S. Firefly Algorithm (FA) Based Particle Filter Method for Visual Tracking. *Optik* **2015**, *126*, 1705–1711. [CrossRef]

26. Ong, K.M.; Ong, P.; Sia, C.K.; Low, E.S. Effective Moving Object Tracking using Modified Flower Pollination Algorithm for Visible Image Sequences under Complicated Background. *Appl. Soft Comput.* **2019**, *83*, 105625. [CrossRef]

27. Castro, E.C.d.; Salles, E.O.T.; Ciarelli, P.M. A New Approach to Enhanced Swarm Intelligence Applied to Video Target Tracking. *Sensors* **2021**, *21*, 1903. [CrossRef] [PubMed]

28. Sardari, F.; Moghaddam, M.E. A Hybrid Occlusion Free Object Tracking Method using Particle Filter and Modified Galaxy Based Search Meta-heuristic Algorithm. *Appl. Soft Comput.* **2017**, *50*, 280–299. [CrossRef]

29. Kang, K.; Bae, C.; Yeung, H.W.F.; Chung, Y.Y. A Hybrid Gravitational Search Algorithm with Swarm Intelligence and Deep Convolutional Feature for Object Tracking Optimization. *Appl. Soft Comput.* **2018**, *66*, 319–329. [CrossRef]

30. Moghaddasi, S.S.; Faraji, N. A Hybrid Algorithm based on Particle Filter and Genetic Algorithm for Target Tracking. *Expert Syst. Appl.* **2020**, *147*, 113188. [CrossRef]
31. Wang, M.; Liu, Y.; Huang, Z. Large Margin Object Tracking with Circulant Feature Maps. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4021–4029.
32. Mirjalili, S. Moth-flame Optimization Algorithm: A Novel Nature-inspired Heuristic Paradigm. *Knowl.-Based Syst.* **2015**, *89*, 228–249. [CrossRef]
33. Abd El Aziz, M.; Ewees, A.A.; Hassanien, A.E. Whale Optimization Algorithm and Moth-flame Optimization for Multilevel Thresholding Image Segmentation. *Expert Syst. Appl.* **2017**, *83*, 242–256. [CrossRef]
34. Mueller, M.; Smith, N.; Ghanem, B. A Benchmark and Simulator for uav Tracking. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 445–461.
35. Du, D.; Qi, Y.; Yu, H.; Yang, Y.; Duan, K.; Li, G.; Zhang, W.; Huang, Q.; Tian, Q. The Unmanned Aerial Vehicle Benchmark: Object Detection and Tracking. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 370–386.
36. Zheng, G.; Fu, C.; Ye, J.; Lin, F.; Ding, F. Mutation Sensitive Correlation Filter for Real-time UAV Tracking with Adaptive Hybrid Label. In Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 30 May–5 June 2021; pp. 503–509.
37. Huang, Z.; Fu, C.; Li, Y.; Lin, F.; Lu, P. Learning Aberrance Repressed Correlation Filters for Real-time UAV Tracking. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 2891–2900.
38. Danelljan, M.; Hager, G.; Shahbaz Khan, F.; Felsberg, M. Learning Spatially Regularized Correlation Filters for Visual Tracking. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 4310–4318.
39. Li, Y.; Fu, C.; Ding, F.; Huang, Z.; Lu, G. AutoTrack: Towards High-performance Visual Tracking for UAV with Automatic Spatio-temporal Regularization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11923–11932.
40. Lukezic, A.; Vojir, T.; Čehovin Zajc, L.; Matas, J.; Kristan, M. Discriminative Correlation Filter with Channel and Spatial Reliability. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6309–6318.
41. Mueller, M.; Smith, N.; Ghanem, B. Context-aware Correlation Filter Tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1396–1404.
42. Lin, F.; Fu, C.; He, Y.; Xiong, W.; Li, F. ReCF: Exploiting response reasoning for correlation filters in real-time UAV tracking. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 10469–10480. [CrossRef]
43. Fu, C.; Xu, J.; Lin, F.; Guo, F.; Liu, T.; Zhang, Z. Object saliency-aware dual regularized correlation filter for real-time aerial tracking. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 8940–8951. [CrossRef]