Computer Modeling in
Engineering & Sciences

Tech Science Press

https://doi.org/10.32604/cmes.2026.085413

ARTICLE

Adaptive Correlation Filter Learning with Motion Smoothing for UAV Tracking

Yu-Feng Yu[1,*], **Xiaoying Tan**[1], **Qirong Wu**[1] **and Guoxia Xu**[2]

[1]Department of Statistics, Guangzhou University, Guangzhou, China

[2]School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing, China

*Corresponding Author: Yu-Feng Yu. Email: yufengyu@gzhu.edu.cn

Received: 11 May 2026; Accepted: 23 July 2026; Published: 28 August 2026

ABSTRACT: To tackle critical visual tracking difficulties arising in UAV tracking tasks, including frequent target occlusion and abrupt fast motion during high-altitude inspection, we propose an adaptive correlation filter tracking algorithm incorporating a motion smoothing module and adaptive residual regularization, named MACF. The tracker is constructed via multi-strategy fusion of two elaborately designed components at the algorithmic modeling level. First, we design a Motion Smoothing Module (MSM) that conducts weighted fusion of historical motion trends in the modeling pipeline. It suppresses search window jitter arising from instantaneous positioning errors and lowers target drift risk by providing precise spatial priors for search center prediction. Second, we embed an Adaptive Residual (AR) regularization term into the objective function of the correlation filter model, which analyzes temporal feature fluctuations to suppress contextual interference and avoid misleading caused by sporadic distractors. The synergistic effect of the motion smoothing module and adaptive residual regularization enables the filter to adapt to various feature changes, thereby maintaining tracking continuity under target occlusion, rapid motion and blurring.

KEYWORDS: UAV tracking; image processing; motion smoothing; correlation filter

1 Introduction

Target tracking is an important research direction in the field of computer vision and has received extensive attention in both academic research and practical applications in recent years. Owing to its excellent performance in long sequence tracking, target tracking has been deployed in several key areas, such as underwater detection [1], autonomous driving [2], security surveillance [3], and medical visual tracking [4]. With the expanding application scope of Unmanned Aerial Vehicles (UAV), UAV tracking has gradually become a research focus for many scholars. The use of UAVs in certain military conflicts further demonstrates their capability to execute designated tasks in complex environments. Currently, object tracking models can be primarily categorized into three types: trackers based on Convolutional Neural Networks (CNN), trackers based on Vision Transformers (ViT), and trackers based on Discriminative Correlation Filters (DCF).

The Efficient Hybrid Linear Self-Attention CNN-Transformer-based Tracker (ECTTrack) [5] employs CNN to extract local information, while the Spatial Scale Attention Tracker (SSTrack) [6] adopts the Transformer framework, thereby achieving more robust and accurate tracking. Although CNN- and Transformer-based tracking methods can achieve favorable accuracy performance, they tend to suffer from relatively large model sizes and considerable computational overhead, which imposes high requirements on hardware resources. For this reason, such methods may not be the most suitable option for resource-limited UAV onboard platforms. Discriminative Correlation Filter (DCF)-based trackers are inherently lightweight and can achieve an elegant balance between computational efficiency and tracking accuracy, rendering them well-suited for real-time visual tracking tasks on UAV onboard platforms.

In recent years, significant progress has been made in the development of correlation filter algorithms. Among them, minimum output sum of squared error (MOSSE) [7] is the first method that applies correlation filters to visual tracking tasks. Subsequently, the spatio-temporal regularized correlation filter (STRCF) [8] and the background-aware correlation filter (BACF) [9] have effectively enhanced the performance of DCF-based trackers through continuous improvements. Although traditional DCF trackers themselves are already capable of handling issues such as illumination variations reasonably well, these successors have still brought significant advancements.

However, UAVs differ from stationary conventional ground cameras. They typically operate at certain altitudes and move at varying speeds. Such operational characteristics give rise to more severe boundary effects, which make it difficult for traditional DCF models to accurately capture targets and achieve stable tracking. Optimizing spatial constraints to adjust the appearance model [10] and enhancing background learning [11] are feasible and practical approaches to address boundary effects. Nevertheless, the value of background information varies over time and is not constant. Furthermore, during UAV tracking, the target is often temporarily occluded by structures like trees and utility poles. It is also susceptible to interference originating from similar objects and complex environments. Under such circumstances, target features can easily be lost, leading to tracking interruption and target drift.

To address the aforementioned issues, this paper proposes a novel method termed the adaptive correlation filter tracking algorithm incorporating a motion smoothing module (MACF), with its overall framework illustrated in Fig. 1. The proposed approach enhances the tracker performance from two perspectives: feature optimization and model improvement. In terms of feature optimization, a Motion Smoothing Module (MSM) is introduced prior to feature extraction to update the search region, thereby yielding higher-quality feature representations. Unlike existing motion prediction methods, MSM employs pre-defined weights to achieve a balanced fusion of historical displacement information across multiple temporal scales. This design provides the tracker with continuous positional prior information, thereby strengthening its robustness. Regarding model improvement, an Adaptive Residual regularization term (AR) is incorporated into the baseline model, which consists of an adaptive matrix and a feature residual component. Distinct from conventional strategies, the adaptive matrix constructed by AR imposes stronger spatial penalties on regions with large historical response variances, thus suppressing the filter's overfitting to noise. Conversely, penalties are reduced in regions with stable responses to preserve the integrity of target features. Meanwhile, AR leverages background information through the feature residual term, enabling the tracker to maintain stable and uninterrupted tracking even in complex scenarios. The main contributions of this paper are summarized as follows.

1. A Motion Smoothing Module (MSM) is proposed. This module performs weighted fusion of historical motion trends. This fusion effectively suppresses severe jitter in the search window caused by instantaneous positioning errors. In this way, the risk of target drift is significantly reduced.
2. Embed the Adaptive Residual regularization term (AR) into the objective function. The model proposed in this paper introduces an adaptive residual regularization term to reduce the impact of context changes.
3. The MSM and the AR are integrated into the baseline model to construct the MACF. This tracker employs a multi-strategy fusion mechanism. As a result, the impact of rapid motion on tracking performance is effectively mitigated. At the same time, the tracker robustly handles temporary target occlusion.

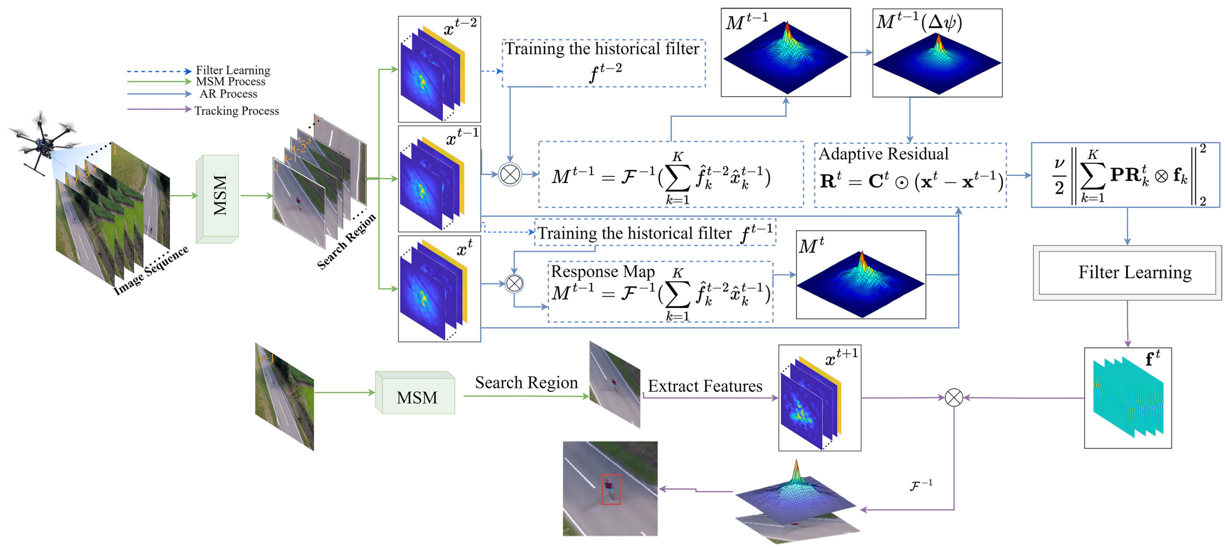

Figure 1: Framework of MACF. MACF begins with a preprocessing stage, where the input is processed by MSM. The MSM improves the localization accuracy of the search region and effectively reduces occlusion-related interference. Subsequently, in the filter training stage, MACF integrates the AR into the model. AR significantly reduces interference caused by abrupt environmental changes and further suppresses boundary effects in complex scenarios. The synergistic integration of these techniques enhances the stability of the tracker.

The rest of this paper is organized as follows. Section 2 reviews the related work. Section 3 details the proposed MACF tracker model. In Section 4, we present an extensive experimental analysis. Finally, Section 5 summarizes the paper.

2 Related Work

2.1 Trackers with Correlation Filters Methods

MOSSE is one of the first methods to introduce correlation filtering into visual tracking. Due to its simple and efficient computation, it is extremely fast, but it is highly susceptible to interference. By introducing the circulant matrix to enable fast Fourier transform acceleration, the kernelized correlation filter (KCF) [12] significantly improves its tracking speed. However, this design also introduces the challenge of boundary effects. To solve the boundary effects, spatially regularized discriminative correlation filters (SRDCF) [13] introduces the concept of spatial regularization by adding a location-dependent weight term to the DCF objective function to penalize filter coefficients, thereby improving tracking quality through regularization. STRCF [8] applies spatial-temporal regularization terms during online learning, and dual regularized correlation filter (DRCF) [14] uses dynamic regularization terms to penalize the background. These optimization mechanisms impose constraints on the background, thereby highlighting the target.

With the widespread application of UAVs, traditional trackers often struggle to perform effectively in practical tasks. To address this issue, some researchers have attempted to incorporate historical information into trackers to better cope with rapid motion in UAV scenarios. To adapt to continuous variations in target appearance, the Interval-Based Response Inconsistency (IBRI) [15] integrates historical response information into the model. Rather than simply stacking historical frames, this method leverages historical response maps to subtly embed information across multiple frames during filter training. The mutation-sensitive correlation filter (MSCF) [16] embeds historical information into the label construction process to enhance the tracker's ability to handle sudden target mutations. Although such label optimization strategies can effectively alleviate target drift, they still struggle to balance the differences between mutation and non-mutation scenarios. To address this limitation, this paper adopts a multi-strategy fusion approach. By integrating historical information more effectively into the filter training process, this approach achieves more accurate target tracking.

2.2 Tracker with Feature Optimization

High quality feature representations convert raw image pixels into discriminative mathematical descriptions, enabling trackers to accurately identify and locate targets from the background. Efficient convolution operators (ECO) [17] employs a dimensionality reduction matrix to compress high-dimensional features, effectively reducing overfitting. DCF-based trackers typically rely on single handcrafted features such as histogram of oriented gradients (HOG) [18] or color names (CN) [19] to perform target tracking tasks. The main drawback of this approach is the lack of comprehensive feature information, which prevents trackers from accurately recognizing targets. To address this issue, researchers have adopted feature fusion strategies to obtain more comprehensive feature representations and thereby enhance tracker performance. DeepSTRCF [8] introduces deep features, and aberrance-repressed correlation filter (ARCF) [20] utilizes multi-channel features. LiteTrack [21] optimizes the feature extraction process at the architectural level by caching template features, which fundamentally avoids redundant computation during inference. This improvement effectively enhances the efficiency of deep feature extraction. Furthermore, the fusion of handcrafted and deep features [22] further improves the reliability of samples during model update. Although the aforementioned approaches effectively enhance feature representation, they also inevitably increase computational complexity.

Feature preprocessing is also an important approach to improving feature quality. It refers to the process of enhancing feature quality by applying simple operations to input features or images. Learning dynamic sensitivity enhanced correlation filter (LDECF) [23] and robust correlation filter learning (RCFL) [24] apply different denoising operations to images, thereby enhancing the discriminative capability of target contours. Channel attentional correlation filter (CACF) [25] adopts a weighting scheme to optimize the importance of different feature channels, obtaining more robust feature representations. Enhanced robust spatial feature selection and correlation filter learning (EFSCF) [26] performs feature enhancement on the target region to highlight the target against the background, thereby strengthening the tracker's discriminative ability. K-nearest neighbor correlation filter learning with p-Laplacian Regularization (KNN-CFL) [27] accomplishes feature enhancement by leveraging feature channel attention. This approach first constructs a similarity matrix among different feature channels, fully utilizing the shared information contained in each feature. Through multi-scale simulated transformations of HOG and CN features, the Multi-Scale Enhanced Features Correlation Filters (MSEFCF) [28] effectively emphasize the target region, which in turn improves the tracker's robustness in complex scenarios, especially under continuously varying target scales. Most of the above improvement schemes are based on image denoising or feature optimization, which impose high requirements on the quality of the feature extraction region. To address this issue, this paper designs the MSM. By introducing a displacement compensation mechanism for consecutive frames, the MSM dynamically corrects the target center position, thereby effectively optimizing the search region and alleviating tracking interruptions caused by occlusion.

2.3 Trackers Handling Background Interference

Background-Aware Correlation Filters (BACF) [9] incorporates background patches into filter training through a cropping matrix, thereby enhancing the filter's ability to handle background information. Spatial Reliability Enhanced Correlation Filter (SRECF) [29] proposes the spatial reliability enhancement learning strategy that dynamically identifies and suppresses unreliable background regions during operation, achieving extremely high robustness and efficiency at a very low computational cost. Aba-ViT [30] proposes a background-aware token halting strategy that preferentially discards background tokens to reduce inference overhead while retaining target tokens to ensure tracking accuracy.

On the other hand, environmental changes themselves can be viewed as dynamic background information, and leveraging background residuals appropriately can also help reduce background interference. Environmental mutation-insensitive correlation filters (EMCF) [11] introduces a background residual term, significantly enhancing the filter's adaptability to environmental changes. Similarly, the bidirectional incongruity-aware correlation filter (BiCF) [31] introduces a bidirectional incongruity error term to efficiently learn appearance variations. However, these methods still fail to fully account for the spatio-temporal inconsistency of environmental changes. Considering that different environmental residuals play distinct roles in filter learning, response-weighted background residual and spatio-temporal regularization correlation filters (RBSCF) [32] optimizes the background residual term using response information, further enhancing the filter's robustness in complex environments. However, the impact of environmental residuals on the filter varies, leaving room for further optimization. The model proposed in this paper introduces an adaptive residual regularization term to mitigate the impact of environmental changes.

3 Methodology

The high maneuverability of UAVs poses significant challenges to visual tracking. Rapid motion and deformation of the target within the field of view often lead to a sharp decline in the performance of traditional DCF-based trackers. Furthermore, the dual motion of both the UAV and the target places additional strain on tracking accuracy and continuity. To enhance tracker adaptability in complex UAV scenarios, this paper proposes the MACF. Specifically, the MSM leverages historical motion information to effectively reduce the risk of target loss, while the AR utilizes background information to achieve adaptive adjustment of the tracker. The synergistic effect of MSM and AR significantly strengthens the anti-interference capability of MACF.

3.1 Motion Smoothing Module

In long-distance imaging scenarios such as UAV high-altitude patrol, the target occupies only a very small number of pixels on the image plane, exhibiting small target characteristics. This results in scarce appearance features available for recognition, making the target highly vulnerable to interference from complex background noise. Since accurate estimation of the search region can effectively compensate for the lack of appearance information, this paper introduces the MSM before the position update stage.

Traditional correlation filter-based trackers typically determine the search region in the current frame by directly utilizing the displacement vector from the previous frame. However, this strategy has inherent limitations when dealing with continuous occlusion. Once an estimation error occurs in a particular frame, subsequent motion compensation relies on incorrect historical information, which tends to cause cumulative errors and lead to target drift. In contrast, the proposed MSM focuses on the recursive processing of temporal motion information. This module estimates the target position based on historical locations rather than simply replicating the single frame displacement. This approach effectively reduces noise interference. The mechanism provides a reasonable motion inertia prior for the target, enabling the tracker to maintain stable position estimation even when detection fails. This strategy effectively prevents instantaneous drift from disrupting tracking continuity. The specific framework of the MSM is illustrated in Fig. 2. The proposed MSM integrates historical information into the feature preprocessing stage via dynamic smoothing. Unlike existing methods, MACF performs this integration at the feature preprocessing level. Specifically, MSM employs a set of pre-defined weights to balance the influence of historical displacement information across different temporal scales.

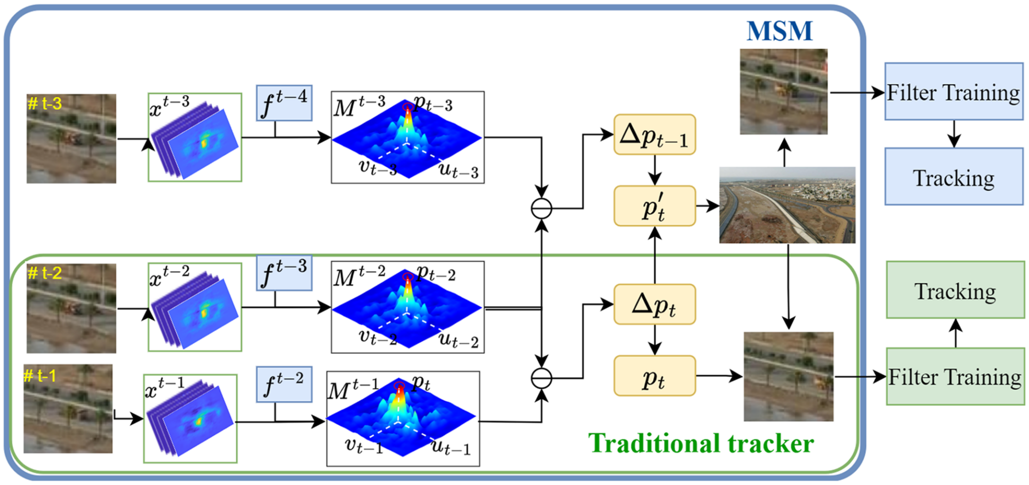

Figure 2: Framework of MSM. The motion smoothing is achieved jointly by the motion compensation of the t-th frame Δp_t and the motion compensation of the $(t-1)$-th frame Δp_{t-1}. This strategy effectively reduces the risk of misleading offset motion compensation, thereby generating reliable motion estimation results and achieving precise adjustment of the target center.

We denote the optimized center position of the search window as $p'_t = (u'_t, v'_t)$. This position is obtained by adding the motion compensation vector $\Delta p_t = (\Delta u_t, \Delta v_t)$ to the target center of the previous frame, $p_{t-1} = (u_{t-1}, v_{t-1})$. Here, the center position p_{t-1} of the previous frame is selected as the location corresponding to the highest peak of the response map in that frame.

$$\begin{cases} u'_t = u_{t-1} + \Delta u_t \\ v'_t = v_{t-1} + \Delta v_t \end{cases} \tag{1}$$

The motion compensation vector $\Delta p_t = (\Delta u_t, \Delta v_t)$ for frame t is calculated using MSM. The specific formula is as follows:

$$\begin{cases} \Delta u_t & = \alpha_2 * (u_{t-1} - u_{t-2}) + (1 - \alpha_2) * \Delta u_{t-1} \\ \Delta v_t & = \alpha_2 * (v_{t-1} - v_{t-2}) + (1 - \alpha_2) * \Delta v_{t-1} \end{cases} \tag{2}$$

where $p_{t-1} = (u_{t-1}, v_{t-1})$, $p_{t-2} = (u_{t-2}, v_{t-2})$ represents the target centers of frames $t-1$ and $t-2$, respectively. $\Delta p_{t-1} = (\Delta u_{t-1}, \Delta v_{t-1})$ denotes the motion compensation vector for frame $t-1$.

Based on the center point coordinates $p'_t = (u'_t, v'_t)$ obtained from MSM, this paper dynamically adjusts and updates the search region. Within the optimized search region, this paper extracts the target feature representation. For clarity and consistency in subsequent discussions, $\mathbf{x}$ in the following content of this paper refers to the target feature extracted within the search region updated by MSM.

3.2 Adaptive Residual Regularization Term

From an aerial perspective, the tracking environment is dynamically changing. Complex ground traffic, cluttered urban streets, similar distractors, and drastic lighting variations exist in such diverse background environments. These factors can adversely affect the tracker. Specifically, they may cause multi-peak phenomena in the response map or lead to confidence fluctuation. Diverse targets such as vehicles, animals, and UAV impose higher requirements on the algorithm's generalization capability.

To address the aforementioned issues, this paper introduces AR. Unlike existing methods, AR leverages response-map-derived weights to adaptively modulate the influence of feature residuals on filter updates. These weights effectively encode historical information, ensuring better alignment with the temporal consistency constraint. By operating directly on feature residuals, our approach suppresses background interference while preserving target-relevant cues.

The core idea of the adaptive residual regularization term is to distinguish the target from the background by analyzing the temporal volatility of the response map. Regions containing the true target typically exhibit higher response values. By contrast, background regions or distractors behave differently. They often display drastic fluctuations in response. Based on this, we construct an adaptive matrix. For regions with high historical response variance, stronger spatial penalties are applied to suppress filter learning in unreliable areas. Conversely, for regions with stable responses, penalties are reduced to preserve the integrity of target features. The specific formulation of the adaptive residual regularization term is given in Eq. (3).

$$\left\| \sum_{k=1}^{K} \mathbf{P}\mathbf{R}_k^t \otimes \mathbf{f}_k \right\|_2^2 \tag{3}$$

where $\mathbf{R}_k^t$ represents the adaptive residual, expressed as follows.

$$\mathbf{R}_k^t = \mathbf{C}^t \odot (\mathbf{x}_k^t - \mathbf{x}_k^{t-1}) \tag{4}$$

where $\odot$ denotes element-wise multiplication, $\mathbf{x}^t$ represents the t-th filter, and $\mathbf{C}^t$ denotes the adaptive matrix. The adaptive matrix $\mathbf{C}^t$ is constructed from the response variation map, which intuitively illustrates the temporal volatility of the response map. The calculation of the adaptive matrix primarily consists of three steps.

First, compute the response maps $\mathbf{M}^t$ and $\mathbf{M}^{t-1}$ for t-th and $t-1$-th frames. The calculation process for $\mathbf{M}^t$ is defined as:

$$\mathbf{M}^t = \mathcal{F}^{-1}\left(\frac{1}{K}\sum_{k=1}^{K} \hat{\mathbf{f}}_k^{t-1} \odot \hat{\mathbf{x}}_k^t\right) \tag{5}$$

where $\mathcal{F}^{-1}$ denotes the inverse Fourier transform, and K is the number of feature channels. Next, construct the response variation map $\Delta\mathbf{C}^t$ by aligning the consecutive response maps and computing their difference:

$$\Delta\mathbf{C}^t = \mathbf{M}^t - \mathbf{M}^{t-1}(\Delta\psi) \tag{6}$$

where $\Delta\psi$ denotes the offset required to align the peaks of consecutive response maps. Finally, normalize $\Delta\mathbf{C}^t$ to obtain the adaptive matrix $\mathbf{C}^t$ using min-max scaling:

$$\mathbf{C}^t = \frac{\max(\Delta\mathbf{C}^t) - \Delta\mathbf{C}^t}{\max(\Delta\mathbf{C}^t) - \min(\Delta\mathbf{C}^t)} \tag{7}$$

3.3 MACF Tracker

To present the proposed method more clearly, we first provide a brief introduction to BACF, which serves as the baseline tracker. Let the feature of the k channel be denoted as $\mathbf{x}_k \in \mathbb{R}^{D\times N}$, and the class label predefined by a Gaussian function be $\mathbf{y} \in \mathbb{R}^{D\times N}$. The objective function of BACF is defined as follows:

$$\varphi(\mathbf{f}) = \frac{1}{2}\left\|\mathbf{y} - \sum_{k=1}^{K} \mathbf{P}\mathbf{x}_k^t \otimes \mathbf{f}_k^t\right\|_2^2 + \sum_{k=1}^{K} \|\mathbf{f}_k\|_2^2 \tag{8}$$

where $\mathbf{f} = [\mathbf{f}_1, \mathbf{f}_2, \cdots, \mathbf{f}_K] \in \mathbb{R}^{D\times N\times K}$ denotes the filters for each feature channel, while $\mathbf{f}_k \in \mathbb{R}^{D\times N}$ is the filter for the k channel. $\mathbf{x}_k^t \in \mathbb{R}^{D\times N}$ undergoes further processing via the binary matrix $\mathbf{P}$ to crop the central region of the feature $\mathbf{x}_k^t$. $\otimes$ represents the correlation operator.

The MACF tracker proposed in this paper considers the temporal continuity property of video sequences. To this end, it incorporates an adaptive residual regularization term into the model to enhance the filter's adaptive capability to sudden changes. During the problem-solving process, the weight of the residual term continuously evolves to adapt to these mutations. The complete representation of the MACF objective function is given in Eq. (9).

$$\varphi(\mathbf{f}) = \frac{1}{2}\left\|\mathbf{y} - \sum_{k=1}^{K} \mathbf{P}\mathbf{x}_k^t \otimes \mathbf{f}_k\right\|_2^2 + \frac{\lambda}{2}\sum_{k=1}^{K} \|\mathbf{f}_k\|_2^2 + \frac{\varepsilon}{2}\left\|\sum_{k=1}^{K} \mathbf{P}\mathbf{R}_k^t \otimes \mathbf{f}_k\right\|_2^2 \tag{9}$$

where $\mathbf{R}_k^t$ denotes the adaptive feature residual calculated by Eq. (4). λ denotes the regularization hyperparameter, and ε is a hyperparameter.

3.4 ADMM Optimization

First, we introduce the auxiliary variable $\mathbf{g} = \mathbf{P}^{\mathrm{T}}\mathbf{f}$, where $\mathbf{g} = [\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_K]$. The objective function in Eq. (9) is then formulated in the augmented Lagrangian form as:

$$\varphi(\mathbf{f}, \mathbf{g}, \mathbf{s}) = \frac{1}{2}\left\|\mathbf{y} - \sum_{k=1}^{K} \mathbf{x}_k^t \otimes \mathbf{g}_k\right\|_2^2 + \frac{\lambda}{2}\sum_{k=1}^{K} \|\mathbf{f}_k\|_2^2 + \frac{\varepsilon}{2}\left\|\sum_{k=1}^{K} \mathbf{R}_k^t \otimes \mathbf{g}_k\right\|_2^2 + \sum_{k=1}^{K}\left(\mathbf{g} - \mathbf{P}^{\mathrm{T}}\mathbf{f}\right)^{\mathrm{T}}\mathbf{s} + \frac{\delta}{2}\sum_{k=1}^{K}\left\|\mathbf{g}_k - \mathbf{P}^{\mathrm{T}}\mathbf{f}_k\right\|_2^2 \tag{10}$$

where $\mathbf{s}$ is the Lagrange multiplier and δ is the penalty factor. Introducing the scaled multiplier $\mathbf{h} = \frac{1}{\delta}\mathbf{s}$ yields the equivalent formulation:

$$\varphi(\mathbf{f}, \mathbf{g}, \mathbf{h}) = \frac{1}{2}\left\|\mathbf{y} - \sum_{k=1}^{K} \mathbf{x}_k^t \otimes \mathbf{g}_k\right\|_2^2 + \frac{\lambda}{2}\sum_{k=1}^{K} \|\mathbf{f}_k\|_2^2 + \frac{\varepsilon}{2}\left\|\sum_{k=1}^{K} \mathbf{R}_k^t \otimes \mathbf{g}_k\right\|_2^2 + \frac{\delta}{2}\sum_{k=1}^{K}\left\|\mathbf{g}_k - \mathbf{P}^{\mathrm{T}}\mathbf{f}_k + \mathbf{h}_k\right\|_2^2 \tag{11}$$

Using Parseval's theorem, we transform the problem into the Fourier domain, where the objective function (11) becomes:

$$\varphi(\hat{\mathbf{f}}, \hat{\mathbf{g}}, \hat{\mathbf{h}}) = \frac{1}{2T}\left\|\hat{\mathbf{y}} - \sum_{k=1}^{K} \hat{\mathbf{x}}_k^t \odot \hat{\mathbf{g}}_k\right\|_2^2 + \frac{\lambda}{2}\sum_{k=1}^{K} \|\hat{\mathbf{f}}_k\|_2^2 + \frac{\varepsilon}{2T}\left\|\sum_{k=1}^{K} \hat{\mathbf{R}}_k^t \odot \hat{\mathbf{g}}_k\right\|_2^2 + \frac{\delta}{2}\sum_{k=1}^{K}\left\|\hat{\mathbf{g}}_k - \hat{\mathbf{f}}_k + \hat{\mathbf{h}}_k\right\|_2^2 \tag{12}$$

where the hat symbol $\hat{\cdot}$ denotes the representation of the corresponding variable in the discrete Fourier domain, obtained via the discrete Fourier transform (DFT).

To facilitate calculation, Eq. (12) is converted into a matrix expression.

$$\varphi(\hat{\mathbf{f}}, \hat{\mathbf{g}}, \hat{\mathbf{h}}) = \frac{1}{2T}\left\|\hat{\mathbf{y}} - (\hat{\mathbf{x}}^t)^{\mathrm{T}} \odot \hat{\mathbf{g}}\right\|_2^2 + \frac{\lambda}{2}\left\|\hat{\mathbf{f}}\right\|_2^2 + \frac{\varepsilon}{2T}\left\|(\hat{\mathbf{R}}^t)^{\mathrm{T}} \odot \hat{\mathbf{g}}\right\|_2^2 + \frac{\delta}{2}\left\|\hat{\mathbf{g}} - \hat{\mathbf{f}} + \hat{\mathbf{h}}\right\|_2^2 \tag{13}$$

Subsequently, we employ the alternating direction method of multipliers algorithm to solve Eq. (13). The optimization problem is decomposed into three tractable subproblems, as presented below:

$$\begin{cases} \hat{\mathbf{f}}^{(i+1)} = \arg\min\limits_{\hat{\mathbf{f}}} \left\{ \dfrac{\lambda}{2} \left\| \hat{\mathbf{f}} \right\|_2^2 + \dfrac{\delta}{2} \left\| \hat{\mathbf{g}} - \hat{\mathbf{f}} + \hat{\mathbf{h}} \right\|_2^2 \right\} \\ \hat{\mathbf{g}}^{(i+1)} = \arg\min\limits_{\hat{\mathbf{g}}} \left\{ \dfrac{1}{2T} \left\| \hat{\mathbf{y}} - (\hat{\mathbf{x}}^t)^{\mathrm{T}} \odot \hat{\mathbf{g}} \right\|_2^2 + \dfrac{\varepsilon}{2T} \left\| (\hat{\mathbf{R}}^t)^{\mathrm{T}} \odot \hat{\mathbf{g}} \right\|_2^2 + \dfrac{\delta}{2} \left\| \hat{\mathbf{g}} - \hat{\mathbf{f}} + \hat{\mathbf{h}} \right\|_2^2 \right\} \\ \hat{\mathbf{h}}^{(i+1)} = \arg\min\limits_{\hat{\mathbf{h}}} \left\{ \hat{\mathbf{g}}^{(i+1)} - \hat{\mathbf{f}}^{(i+1)} + \hat{\mathbf{h}}^{(i)} \right\} \end{cases} \tag{14}$$

Subproblem for $\mathbf{f}$: Given $\hat{\mathbf{g}}$ and $\hat{\mathbf{h}}$, the optimal solution for $\hat{\mathbf{f}}^{(i+1)}$ is derived in closed form as:

$$\hat{\mathbf{f}} = \frac{\delta \hat{\mathbf{g}} + \delta \hat{\mathbf{h}}}{\lambda + \delta} \tag{15}$$

Subproblem for $\mathbf{g}$: Given $\hat{\mathbf{f}}$ and $\hat{\mathbf{h}}$, $\hat{\mathbf{g}}^{(i+1)}$ can be derived as follows:

$$\hat{\mathbf{g}} = \left(\hat{\mathbf{x}}^t (\hat{\mathbf{x}}^t)^{\mathrm{T}} + \varepsilon \hat{\mathbf{R}}^t (\hat{\mathbf{R}}^t)^{\mathrm{T}} + \delta T \mathbf{I} \right)^{-1} \left(\hat{\mathbf{x}}^t \hat{\mathbf{y}} + \delta T \hat{\mathbf{f}} - \delta T \hat{\mathbf{h}} \right) \tag{16}$$

The following approximate solution is obtained using the Sherman-Morrison formula.

$$\hat{\mathbf{g}} = -\frac{\hat{\mathbf{x}}^t (\hat{\mathbf{x}}^t)^{\mathrm{T}} + \varepsilon \hat{\mathbf{R}}^t (\hat{\mathbf{R}}^t)^{\mathrm{T}}}{\delta T \left(\delta T + (\hat{\mathbf{x}}^t)^{\mathrm{T}} \hat{\mathbf{x}}^t + \varepsilon (\hat{\mathbf{R}}^t)^{\mathrm{T}} \hat{\mathbf{R}}^t \right)} \left(\hat{\mathbf{x}}^t \hat{\mathbf{y}} + \delta T \hat{\mathbf{f}} - \delta T \hat{\mathbf{h}} \right) + \frac{\hat{\mathbf{x}}^t \hat{\mathbf{y}} - \delta T \hat{\mathbf{h}} + \delta T \hat{\mathbf{f}}}{\delta T} \tag{17}$$

Subproblem for $\mathbf{h}$: Updates the Lagrange multiplier $\mathbf{h}$ by

$$\hat{\mathbf{h}}^{(i+1)} = \left\{ \hat{\mathbf{g}}^{(i+1)} - \hat{\mathbf{f}}^{(i+1)} + \hat{\mathbf{h}}^{(i)} \right\}. \tag{18}$$

The penalty factor δ is adaptively updated during iterations to improve convergence:

$$\delta^{(i+1)} = \min\left(\beta \delta^i, \delta^{\max} \right) \tag{19}$$

where $\delta^{\max}$ denotes the maximum of δ and β is the scale step factor.

Finally, we update the appearance model $\mathbf{x}_{k,\mathrm{model}}^t$ using a linear interpolation strategy:

$$\hat{\mathbf{x}}_{k,\mathrm{model}}^t = (1 - \theta) \hat{\mathbf{x}}_{k,\mathrm{model}}^{t-1} + \theta \hat{\mathbf{x}}^t \tag{20}$$

where θ represents the learning rates. This online dynamic update provides a stable and robust feature representation for subsequent tracking.

The detailed procedure of the proposed MACF tracker is summarized in Algorithm 1.

Algorithm 1: MACF tracker

Input: Historical Target Center: $p_{t-1} = (u_{t-1}, v_{t-1})$, $p_{t-2} = (u_{t-2}, v_{t-2})$;
Historical motion compensation: $\Delta p_{t-1} = (\Delta u_{t-1}, \Delta v_{t-1})$
Output: Correlation filter: $\mathbf{f}_k^t$
Initialization: Initialization $\hat{\mathbf{g}}_k^t(0)$, $\hat{\mathbf{f}}_k^t(0)$, $\hat{\varepsilon}_k(0)$;
Set maximum ADMM iteration count L;

(Continued)

Algorithm 1 (continued)

1: Update $p'_t = (u'_t, v'_t)$ by Eqs. (1) and (2);
2: Update searching area by $p'_t = (u'_t, v'_t)$;
3: Update $\mathbf{R}_k^t$ by Eqs. (4)–(7);
4: for $i = 1$ to K do
5: Update $\mathbf{f}^{(i+1)}$ by Eq. (15);
7: Update $\hat{\mathbf{g}}^{(i+1)}$ by Eq. (17);
8: Update $\hat{\mathbf{h}}^{(i+1)}$ by Eq. (18);
9: Update $\delta^{(i+1)}$ by Eq. (19);
10: End for.
11: Update $\hat{\mathbf{x}}_{k,\mathrm{model}}^t$ by Eq. (20).

4 Experimental Results and Analysis

To evaluate the performance of the proposed MACF, this paper adopts three representative benchmark datasets: DTB70 [33], UAV123@10FPS [34] and UAVDT [35]. On these benchmark datasets, MACF is compared with various state-of-the-art trackers to verify that the improvements are effective. These three datasets include multiple scenarios such as urban and forest environments. Moreover, they cover key challenges including aspect ratio variation, occlusion, and similar object interference. Therefore, a rigorous and practical evaluation of MACF is enabled.

4.1 Implementation Details

In this work, MATLAB 2021a was employed as the software platform for evaluating the MACF tracker, and all experiments were conducted on a computer equipped with an Intel Core i3-1115G4 processor (CPU, 3.00 GHz). The performance of MACF was assessed under the One-Pass Evaluation (OPE) protocol, with three primary metrics considered: precision (Prec.), success rate (Succ.), and frames per second (FPS). Specifically, precision and success rate serve as key indicators of tracking accuracy and robustness, while FPS is adopted as the criterion for real-time performance evaluation.

Considering that parameter settings have a significant impact on model performance, this paper provides the specific parameter values required by the algorithm: $\alpha_2 = 0.7$, $\lambda = 100$, $\varepsilon = 50$. The learning rate θ is set as 0.0382.

4.2 Experimental Results on the DTB70 Benchmark

On the DTB70 dataset, we compare the proposed MACF tracker with 19 state-of-the-art trackers, including ECO [17], BACF [9], STRCF [8], ARCF [20], DRCF [14], IBRI [15], tracking for UAV with automatic spatio-temporal regularization (AutoTrack) [36], BiCF [31], response reasoning-based correlation filter (ReCF) [37], multi-regularized correlation filter (MRCF) [38], MSCF [16], EMCF [11], CACF [25], RCFL [24], RBSCF [32], enhanced robust spatial feature selection and correlation filter learning (EFSCF) [26], MSE-FCF [28], SRECF [29] and LDECF [23]. Experimental results demonstrate that the proposed MACF achieves excellent tracking performance.

As shown in Fig. 3, MACF ranks first in both precision and success rate metrics. Specifically, its tracking precision reaches 76.3%, surpassing the second-ranked LDECF by 1.6% and exceeding BACF by 17.3%. In terms of success rate, MACF achieves a score of 52.1%, outperforming LDECF by 1.3% and leading BACF by 11.9%. They indicate that the proposed MACF achieves substantial improvements in tracking performance.

This is because the combination of MSM and AR enables the tracker to maintain robust target localization and tracking continuity.

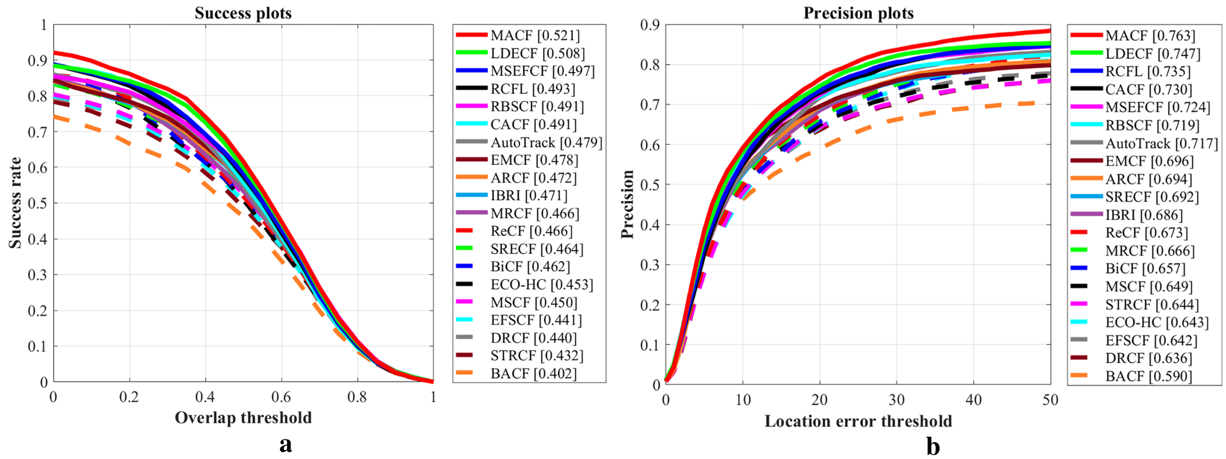

Figure 3: Success plots (**a**) and precision plots (**b**) of the proposed MACF and other trackers on the DTB70 database.

To verify the statistical reliability of the performance improvements achieved by MACF, we employ the Wilcoxon signed-rank test to conduct pairwise comparisons between MACF and the six mainstream trackers with the closest performance to MACF, as well as the baseline method BACF on the DTB70 dataset. Table 1 shows the comparison results. The results demonstrate that MACF achieves significant or marginally significant improvements over most of the compared methods in terms of both Success rate and Precision, which further supports the effectiveness of our proposed enhancements.

Table 1: Wilcoxon signed-rank test results for performance comparison of MACF against state-of-the-art trackers on the DTB70 dataset.

Metric	Compared Tracker	p-Value	Significance (α = 0.1)
Succ.	LDECF	0.037	Significant ($*$)
	MSEFCF	0.046	Significant ($*$)
	RCFL	0.123	Not Significant
	RBSCF	0.069	Marginally Significant ($\cdot$)
	CACF	0.133	Not Significant
	AutoTrack	0.015	Significant ($*$)
	BACF	0.000	Significant ($***$)
Prec.	LDECF	0.091	Marginally Significant ($\cdot$)
	MSEFCF	0.081	Marginally Significant ($\cdot$)
	RCFL	0.616	Not Significant
	RBSCF	0.032	Significant ($*$)
	CACF	0.595	Not Significant
	AutoTrack	0.285	Not Significant
	BACF	0.000	Significant ($***$)

Note: indicates marginal significance ($0.05 < p \leq 0.1$), * indicates significance ($0.01 < p \leq 0.05$), and *** indicates high significance ($p \leq 0.001$).

Specifically, MACF significantly outperforms the second-ranked tracker LDECF in both Success rate (52.1% vs. 50.8%, $p = 0.037$) and Precision (76.3% vs. 74.7%, $p = 0.091$), confirming that the observed performance advantages are not attributable to random fluctuations. Meanwhile, we note that the improvements of MACF over RCFL and CACF do not reach statistical significance. This can be attributed to the inherent nature of the Wilcoxon test, which evaluates the consistency of pairwise differences rather than the absolute magnitude of mean differences. Taking RCFL as an example, although MACF maintains a lead in overall Precision, this advantage is primarily concentrated on a limited number of sequences where MACF achieves substantial gains, while the two trackers exhibit comparable or even reversed performance on the majority of the remaining sequences. This leads to insufficient consistency in both the direction and ranking of pairwise differences, thereby yielding higher p-values. In contrast, although the improvement of MACF over LDECF is more modest in terms of the average margin, it is consistently observed in a stable manner across most sequences, thus passing the significance test. This comparison does not undermine the effectiveness of MACF; rather, it provides a nuanced characterization of how the stability of its advantages varies across different baseline methods. On the other hand, MACF achieves highly significant improvements over the baseline BACF in both Success rate and Precision ($p < 0.001$), which directly validates the fundamental effectiveness of our core technical contributions. Overall, the statistical test results provide evidence for the overall effectiveness of MACF.

To comprehensively evaluate the performance of MACF, we conduct comparative experiments across various attributes. Fig. 4 presents the comparison results for selected attributes. As shown in Fig. 4, MACF demonstrates superior performance across multiple attributes. In terms of Background Clutter (BC), Fast Camera Motion (FCM), In-Plane Rotation (IPR), Motion Blur (MB), Occlusion (OCC) and Similar Objects Around (SOA), MACF achieves the highest scores in both success rate and precision. This outstanding performance stems from the motion smoothing module and the residual regularization term. The former incorporates sufficient historical motion information to accurately lock onto the search region. The latter imposes dynamic environmental penalties on the filter, which enables the tracker to better adapt to cluttered environments and achieve robust tracking. These results indicate that the improved tracker exhibits stable performance in diverse challenging scenarios.

4.3 Experimental Results on the UAV123@10fps Benchmark

This paper evaluates the performance of MACF on the UAV123@10fps dataset, comparing it against 19 trackers. As shown in Fig. 5, MACF achieves the best overall performance, ranking first in both success rate and precision. Specifically, in terms of success rate, MACF outperforms the second-ranked tracker (EMCF) by 0.8% and significantly surpasses the baseline method (BACF) by 8.1%. In terms of precision, MACF exceeds EMCF by 1.1% and achieves a substantial improvement of 12.6% over BACF. This outstanding performance is primarily attributed to the synergistic effect of the MSM and AR modules. The MSM leverages the temporal inertia of the motion trajectory. By doing so, it effectively reduces the risk of localization drift caused by single-frame velocity anomalies or occlusion. The AR suppresses interfering responses from background noise while enhancing the saliency of target features. This effectively alleviates the confusion of the tracker caused by similar background textures.

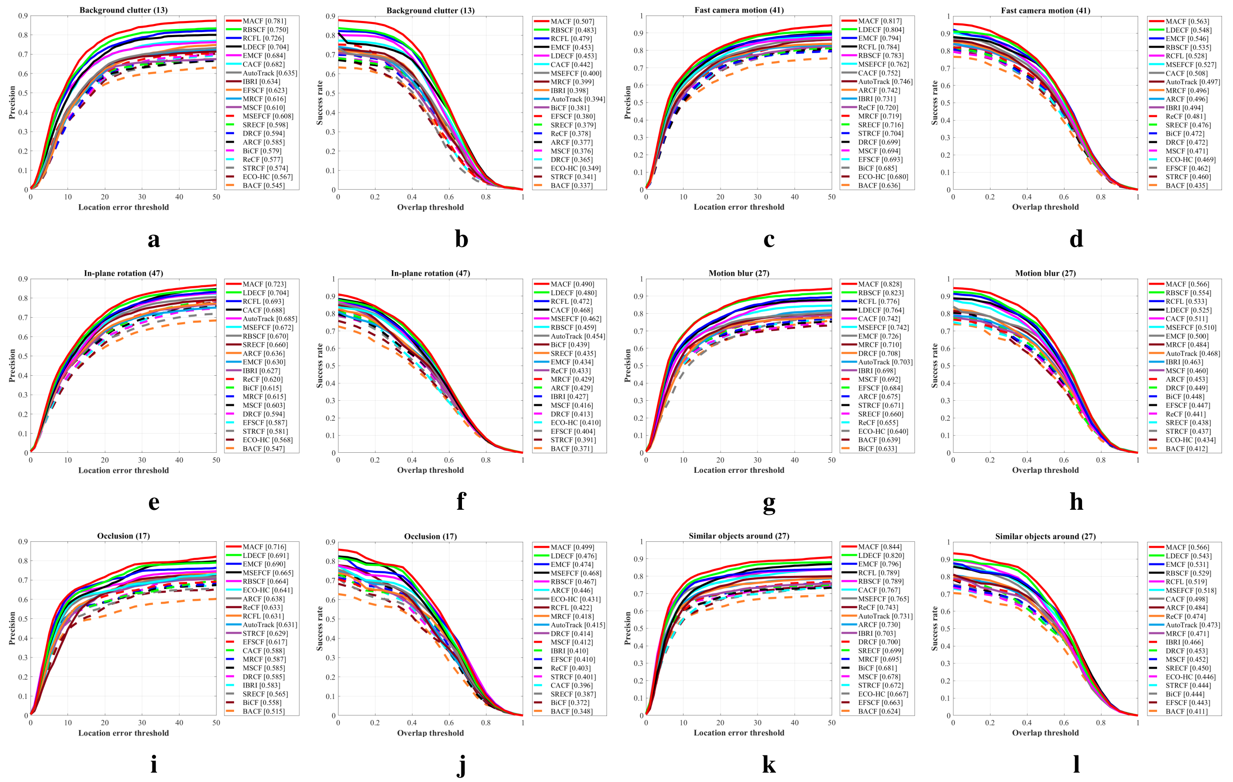

Figure 4: Success plots (**a**,**c**,**e**,**g**,**i**,**k**) and precision plots (**b**,**d**,**f**,**h**,**j**,**l**) of MACF and other trackers on partial attributes of the DTB70 dataset.

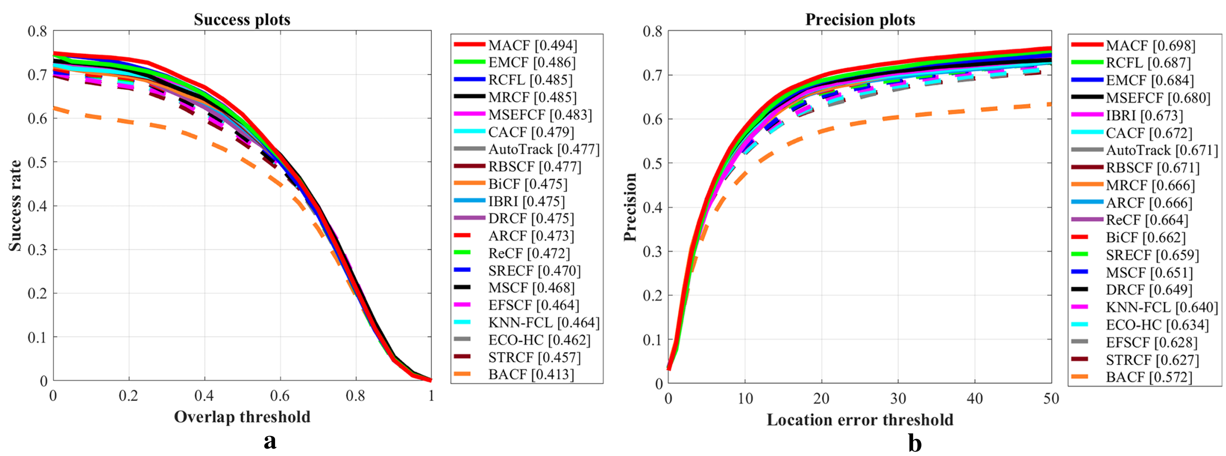

Figure 5: Success plots (**a**) and precision plots (**b**) on the UAV123@10fps database.

To comprehensively evaluate the performance of the proposed MACF tracker, we conduct comparative experiments across 12 highly challenging attribute scenarios. Table 2 summarizes the precision results of 19 comparative trackers and MACF on several key attributes. As shown in the table, MACF ranks first in precision for attributes such as ARV, Illumination Variation (IV), and Viewpoint Change (VC). The core reason for this achievement lies in the introduction of the adaptive residual regularization term. This term enables the tracker to adaptively adjust feature responses when the target appearance undergoes dynamic changes, thereby effectively improving tracking accuracy. Meanwhile, MACF also demonstrates excellent performance on more challenging attributes such as Camera Motion (CM), Fast Motion (FM), and Partial Occlusion (PO). This performance is attributed to MSM. It optimizes the center positioning of the search region based on historical displacement information. This effectively mitigates drift issues caused by occlusion and rapid target motion. By dynamically adjusting the position of the search window, the tracker maintains continuous locking on the target in complex scenarios. Furthermore, MACF achieves the best performance under Low Resolution (LR) conditions. This is attributed to its strong background suppression capability. Additionally, MACF is able to extract high-quality discriminative features under such conditions. Overall, MACF exhibits considerable robustness and adaptability when evaluated across various challenging attributes.

Table 2: Success rate comparison of the proposed MACF and 19 trackers across various attributes on the UAV123@10fps dataset. The top three methods for each attribute are highlighted in Red, Green, and Blue, respectively. An upward arrow ⇑ indicates superior performance compared to the baseline.

Methods	ARV⇑	CM⇑	FM⇑	IV⇑	LR⇑	PO⇑	SV⇑	VC⇑	BC⇑	OV⇑
ECO	0.558	0.609	0.487	0.507	0.527	0.556	0.587	0.548	0.511	0.522
BACF	0.478	0.532	0.407	0.430	0.431	0.467	0.525	0.491	0.425	0.421
STRCF	0.524	0.602	0.488	0.493	0.509	0.559	0.580	0.537	0.477	0.523
EFSCF	0.532	0.592	0.479	0.494	0.515	0.560	0.581	0.526	0.475	0.518
DRCF	0.553	0.610	0.520	0.556	0.522	0.591	0.604	0.566	0.486	0.533
MSCF	0.567	0.598	0.517	0.543	0.521	0.582	0.606	0.578	0.431	0.545
MRCF	0.575	0.622	0.510	0.541	0.553	0.594	0.623	0.577	0.460	0.544
BiCF	0.578	0.625	0.484	0.581	0.534	0.591	0.619	0.581	0.515	0.533
ARCF	0.580	0.610	0.516	0.552	0.561	0.582	0.623	0.573	0.445	0.513
EMCF	0.608	0.646	0.510	0.572	0.572	0.583	0.644	0.602	0.486	0.542
ReCF	0.584	0.632	0.512	0.592	0.543	0.590	0.621	0.581	0.514	0.526
IBRI	0.592	0.623	0.547	0.548	0.559	0.595	0.631	0.590	0.443	0.539
AutoTrack	0.598	0.647	0.525	0.550	0.532	0.584	0.629	0.588	0.502	0.554
MSEFCF	0.605	0.653	0.539	0.567	0.533	0.571	0.640	0.593	0.492	0.509
CACF	0.600	0.630	0.544	0.553	0.533	0.577	0.630	0.590	0.459	0.537
RBSCF	0.583	0.631	0.503	0.554	0.554	0.562	0.629	0.597	0.486	0.522
SRECF	0.588	0.625	0.513	0.545	0.495	0.593	0.615	0.582	0.460	0.565
RCFL	0.624	0.655	0.552	0.593	0.573	0.609	0.647	0.607	0.533	0.556
KNN-FCL	0.537	0.610	0.473	0.502	0.501	0.579	0.594	0.557	0.473	0.552
MACF	0.627	0.675	0.550	0.616	0.574	0.609	0.659	0.609	0.518	0.562

4.4 Experimental Results on the UAVDT Benchmark

To verify the effectiveness of the proposed MACF method, this paper conducts comparative experiments on three categories of different trackers: trackers based on discriminative correlation filter (DCF), trackers based on convolutional neural network (CNN), and trackers based on Vision Transformer (ViT). The evaluated DCF-based trackers include BACF [9], DeepSTRCF [8], Unsupervised deep tracking (UDT+) [39], MetricNe _ECO (MN_ECO) [40], hand-crafted features-based tracker (CCF) [41], learning background-aware and spatial-temporal regularized correlation filters (BSTCF) [42], MSEFCF [28], SRECF [29] and KNN-CFL [27]. The CNN-based trackers include LUDT+ [43], Scale-Equivariant

SiamFC (SE-SiamFC) [44], Multi-Stream Faster R-CNN [45] and tracking with query guided Redetection (QRDT) [46]. The ViT-based trackers include Spatio-temporal transformer network for visual tracking (STARK-ST101) [47], hierarchical vision transformers (HiT) [48], transformer tracking framework based on the successful mixFormer tracker (Mixformerv2) [49], Aba-ViT [30], motion blur robust vision transformers with dynamic early exit (BDTrack-ViT) [50], LiteTrack [21], target aware vision transformers (TATrack-ViT) [51], MMKLTrack [52] and Aba-ViTrack++ [51].

As shown in Table 3, comparative results are provided regarding success rate, precision, and tracking speed on GPU and CPU platforms. Experimental results indicate that MACF outperforms all other DCF-based trackers in both precision and success rate, even surpassing some deep learning-based trackers. This improvement is primarily attributed to the MSM providing more accurate search region localization. However, it must be pointed out that although MACF achieves state-of-the-art performance among traditional correlation filter-based trackers, its overall performance still lags behind mainstream ViT-based trackers. We attribute this performance gap primarily to the inherent linear template update mechanism of DCF, which struggles to cope with drastic appearance variations. In contrast, the global attention mechanism of ViT enables more effective modeling of long-range dependencies. This limitation also suggests a promising direction for our future work: to organically integrate the computational efficiency of MACF with the powerful representation capability of Transformers.

4.5 Ablation Study

This study improves the tracker from two main aspects: feature preprocessing and model construction. To comprehensively evaluate the contribution of each improvement to overall performance, we design an ablation study. By comparing four models on the DTB70 dataset, the impact of different optimization strategies on tracking performance can be clearly identified. The experimental setup is as follows: the first model is $Baseline$, namely BACF; the second model ($Baseline_{MSM}$) introduces only MSM; the third model ($Baseline_{AR}$) adds only AR; the fourth model ($Baseline_{MSM+AR}$) is the completely improved version, integrating both MSM and AR. Table 4 presents the comparison results of various performance metrics, including precision and success rate.

The experimental results on the DTB70 dataset demonstrate that the improved $Baseline_{MSM+AR}$ achieves a precision of 76.3% and a success rate of 52.1%, both of which significantly outperform the baseline model. As shown in Table 4, $Baseline_{MSM+AR}$ consistently surpasses $Baseline$ in terms of both precision and success rate. Furthermore, compared with $Baseline_{MSM}$, $Baseline_{MSM+AR}$ exhibits further improvements in both tracking precision and success rate. These findings confirm the critical role of the AR module in enhancing tracker performance. It is worth noting that both the precision and success rate of $Baseline_{MSM}$ are lower than those of $Baseline$, indicating that the MSM alone fails to improve the baseline model's performance on the DTB70 dataset. This degradation may be attributed to the fact that the MSM, when deployed individually, imposes excessive smoothness, which in turn impairs the discriminative power of the baseline.

We further conduct a comprehensive comparison across multiple challenging attributes, with the results summarized in Table 4. It is observed that $Baseline_{MSM+AR}$ outperforms the other three competing trackers in terms of both precision and success rate on the vast majority of attributes, with particularly notable improvements under challenges such as deformation, scale variation, and out-of-view scenarios. The superior performance of $Baseline_{MSM+AR}$ validates the effective complementarity between MSM and AR. MSM provides a motion smoothing prior to enhance trajectory continuity, while AR suppresses anomalous responses at the response map level. The synergy between the two modules effectively addresses appearance variations and motion uncertainties in complex tracking scenarios. Collectively, these experimental results corroborate the robustness of MACF.

Table 3: Comparisons of performance on the UAVDT dataset between MACF and other deep trackers. The top three methods in Prec, Succ, and FPS are highlighted by using different colors: Red, Green, and Blue. The superscript § indicates that the tracker utilizes deep features.

Method		Venue	Prec.	Succ.	GPU	CPU
DCF-based	BACF	17'ICCV	0.706	0.442	–	51.9
	DeepSTRCF§	18'CVPR	0.667	0.437	6.8	–
	UDT+§	19'CVPR	0.674	0.442	73.3	–
	MN_ECO§	20'ACM	0.691	0.435	30.6	–
	fECO§	20'TIP	0.699	0.415	20.6	–
	CCF§	23'TCSVT	0.736	0.467	45	–
	BSTCF§	23'AI	0.685	0.441	19	–
	MSEFCF	24'TIV	0.733	0.448	–	30
	SRECF	24'TMM	0.717	0.450	–	–
	KNN-CFL	25'TST	0.693	0.439	–	–
	MACF	**Ours**	**0.744**	**0.451**	–	64.49
CNN-based	LUDT+	21'IJCV	0.701	0.406	59.4	–
	Multi-Stream Faster R-CNN	21'RS	0.710	0.452	–	–
	SE-SiamFC	22'IJCNN	0.794	0.555	255.4	51.6
	QRDT	24'TIM	0.606	0.477	–	–
ViT-based	STARK-ST101	21'ICCV	0.704	0.469	30.2	–
	HiT	23'ICCV	0.623	0.471	237.7	43.2
	Mixformerv2	23'NIPS	0.578	0.421	184.6	37.2
	Aba-ViT	23'ICCV	0.834	0.599	–	–
	BDTrack-ViT	24'arXiv	0.789	0.573	235.2	56.9
	TATrack-ViT	24'TGRS	0.805	0.586	203.3	47.7
	LiteTrack	24'ICRA	0.818	0.596	203.3	47.7
	MMKLTrack	26'ESWA	0.791	0.573	–	–
	Aba-ViTrack++	26'TCSVT	0.847	0.614	–	–

Table 4: Ablation study of MACF on DTB70 (precision/success).

Tracker	Prec.	Suc.	SV	SOA	ARV	BC	Def	FCM	IPR	MB	OV	FPS
Baseline (BACF)	0.590	0.402	0.533 /0.392	0.624 /0.411	0.392 /0.273	0.545 /0.337	0.448 /0.302	0.636 /0.435	0.547 /0.371	0.639 /0.412	0.650 /0.419	47.20
Baseline$_{MSM}$	0.573	0.384	0.525 /0.383	0.638 /0.410	0.463 /0.321	0.514 /0.343	0.388 /0.264	0.648 /0.424	0.521 /0.356	0.598 /0.397	0.558 /0.362	63.33
Baseline$_{AR}$	0.737	0.503	0.635 /0.462	0.843 /0.560	0.588 /0.395	0.747 /0.492	0.638 /0.429	0.787 /0.542	0.692 /0.473	0.810 /0.552	0.607 /0.402	64.00
Baseline$_{MSM+AR}$ (MACF)	0.763	0.521	0.682 /0.486	0.844 /0.566	0.623 /0.419	0.781 /0.507	0.686 /0.453	0.817 /0.563	0.723 /0.490	0.828 /0.566	0.690 /0.458	63.43

We evaluate the FPS of MACF and its variants on the DTB70 dataset (Table 4). While MACF introduces additional modules (MSM and AR), it achieves 63.43 FPS, which is 16.23 FPS higher than the BACF baseline (47.20 FPS). Although adding modules usually increases latency, the observed speed-up is attributed to the MSM-provided accurate search region, which accelerates the convergence of the ADMM optimizer. Meanwhile, AR operates on lightweight feature residuals, introducing negligible overhead. This confirms that MACF not only improves accuracy but also maintains excellent real-time performance for UAV deployment.

4.6 Parameter Analysis

This paper integrates historical motion information into the MSM through weighted fusion, where the weight parameter α_2 is a critical hyperparameter. To evaluate its impact, a sensitivity analysis is conducted on α_2. On the DTB70 dataset, parameter α_2 is evaluated within the range [0.1, 1.0) with a step size of 0.1. The results (as shown in Fig. 6) indicate that the value of α_2 significantly affects the precision of MACF and the success rate.

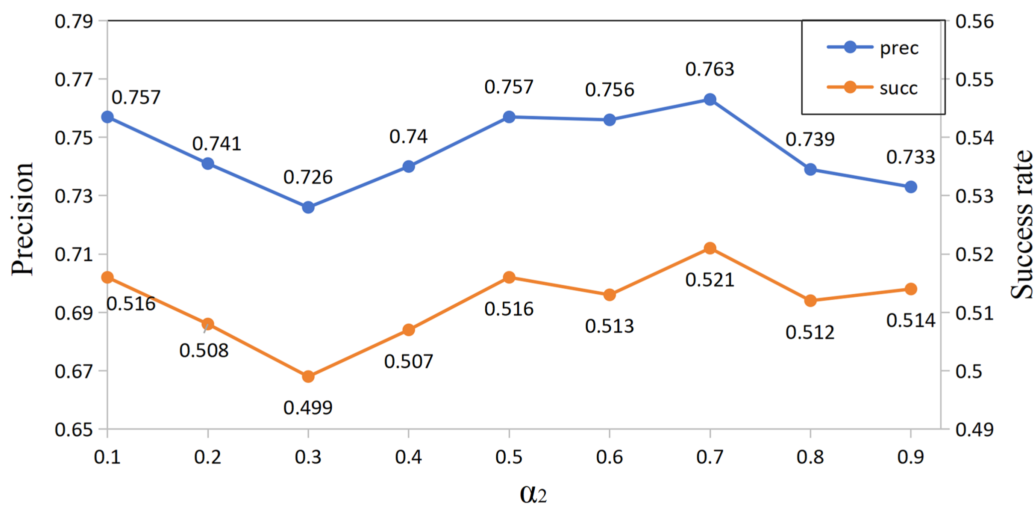

Figure 6: Analysis of the key parameter α_2 in MACF.

A small α_2 (i.e., $\alpha_2 < 0.3$) causes severe motion prediction lag and compromises tracking accuracy. A large α_2 (i.e., $\alpha_2 > 0.8$) makes motion compensation overly sensitive, amplifying single-frame localization noise into oscillations that undermine tracking stability. When α_2 is set to 0.7, an optimal balance is achieved. At this value, the motion prediction provided by MSM closely follows the target's dynamic changes. Simultaneously, it effectively filters out high-frequency noise, thereby maximizing tracking precision and success rate. Therefore, this paper sets α_2 to 0.7.

4.7 Complexity Analysis

To provide a fair and objective assessment of the computational complexity and real-time capability of our tracker, we conduct a modular runtime analysis on the Skiing1 sequence, which is selected from the DTB70 dataset due to its representative challenging attributes. The per-frame processing latency is employed as the primary metric, allowing us to suppress system-level fluctuations and reliably characterize the steady-state efficiency of the algorithm during routine tracking operations.

Table 5 presents the runtime breakdown of each module in MACF. Notably, the motion compensation based on MSM requires only 0.029 ms/frame, consuming negligible computational resources, which underscores the inherent lightweight design advantage of this module. In contrast, the ADMM-based filter training still incurs a runtime of 2.14 ms/frame. Nevertheless, the complete tracking pipeline achieves an average processing time of 19.8 ms per frame, indicating that the overall system already meets the fundamental requirements for real-time tracking.

Table 5: Per-frame runtime breakdown of MACF on the DTB70 Skiing1 sequence using the CPU platform.

Component	Runtime
MSM motion compensation	0.029 ms/frame
ADMM filter training	2.14 ms/frame
Total pipeline	19.8 ms/frame

To gain a thorough understanding of the algorithm's resource footprint, we evaluate two complementary metrics: the model size and the runtime peak memory consumption. All statistics are computed over the full DTB70 dataset comprising 70 sequences, and are reported as mean ± standard deviation to reflect both the typical performance and its stability.

Table 6 reveals that MACF occupies a model size of merely 2.05 MB, which substantiates its compact and lightweight design. Throughout the tracking pipeline, the average runtime peak memory is measured at 2.70 MB. Notably, even when evaluated on 70 sequences covering a wide range of challenging scenarios—including scale changes and rapid object movement—the peak memory consumption of MACF remains highly stable, with a standard deviation as low as 0.32 MB. In summary, the above experimental results demonstrate that MACF exhibits favorable lightweight characteristics in terms of both computational time and memory footprint.

Table 6: Memory footprint of MACF on the DTB70 dataset.

Metric	Value (Mean ± Std)
Model Size (Static)	2.05 ± 0.00 MB
Peak Memory Consumption (Runtime)	2.70 ± 0.32 MB

5 Conclusions

This paper presents the MACF tracker, which integrates a Motion Smoothing Module (MSM) and an Adaptive Residual Regularization (AR) mechanism for robust UAV-based visual tracking. The MSM optimizes the search region using historical tracking information and motion trends from response maps, enabling stable tracking and rapid target re-capture under high-speed motion and brief occlusions, which is critical for public security scenarios. The AR module dynamically adjusts feature weights and regularization constraints to suppress background noise and handle appearance variations caused by viewpoint, illumination, or motion changes, improving adaptability in complex environments. Experimental results on multiple UAV datasets demonstrate that MACF achieves competitive tracking performance. The tracking framework integrating MSM and AR outperforms existing mainstream methods in both precision and success rate. However, there remains room for improvement in combining this method with deep learning techniques. Future research will further explore the integration of deep learning approaches into the DCF tracking framework, aiming to achieve greater improvements in both accuracy and real-time performance.

Acknowledgement: None.

Funding Statement: This work was supported in part by the Guangdong Basic and Applied Basic Research Foundation under Grants 2024A1515012040 and 2026A1515011166, and in part by the Science and Technology Planning Project of Guangzhou under Grant 2024A03J0401.

Author Contributions: The authors confirm their contribution to the paper as follows: Yu-Feng Yu, Xiaoying Tan and Qirong Wu contributed to writing—original draft and literature research; Yu-Feng Yu and Guoxia Xu contributed to conceptualization, supervision, and project administration; Yu-Feng Yu, Xiaoying Tan, Qirong Wu and Guoxia Xu contributed to writing—review & editing. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The datasets analyzed in this study are publicly available from their original sources cited in the article.

Ethics Approval: Not applicable. This study used publicly available datasets and did not involve new studies with human participants or animals performed by any of the authors.

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

References

1. Zou L, Liang B, Cheng X, Li S, Lin C. Sonar image target detection for underwater communication system based on deep neural network. Comput Model Eng Sci. 2023;137(3):2641–59. doi:10.32604/cmes.2023.028037.
2. Yang SG, Kim J, Lim SC. High-accuracy path tracking for unmanned vehicle navigation: a hierarchical deep reinforcement learning approach. IEEE Trans Veh Technol. 2026;75(2):3342–7.
3. Abba S, Bizi AM, Lee JA, Bakouri S, Crespo ML. Real-time object detection, tracking, and monitoring framework for security surveillance systems. Heliyon. 2024;10(15):e34922. doi:10.1016/j.heliyon.2024.e34922.
4. Ye S, Yang B, Quan H, Liu S, Tang M, Tian J. A method for ultrasound servo tracking for puncture needle. Comput Model Eng Sci. 2025;144(2):2287–306. doi:10.32604/cmes.2025.066195.
5. Xu L, Guo Z, Wang L. Efficient hybrid linear self-attention based visual object tracking with LoRA. Neurocomputing. 2025;640(3):130335. doi:10.1016/j.neucom.2025.130335.
6. Ma S, Wan Z, Hu B, Zhang J, Hou Z, Zhao X. SSTrack: joint scale-aware temporal prompts and spatio-temporal prior transformer for visual object tracking. Knowl Based Syst. 2026;337:115370.
7. Bolme DS, Beveridge JR, Draper BA, Lui YM. Visual object tracking using adaptive correlation filters. In: Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2010 Jun 13–18; San Francisco, CA, USA. p. 2544–50.

8. Li F, Tian C, Zuo W, Zhang L, Yang MH. Learning spatial-temporal regularized correlation filters for visual tracking. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2018 Jun 18–23; Salt Lake City, UT, USA. p. 4904–13.
9. Kiani Galoogahi H, Fagg A, Lucey S. Learning background-aware correlation filters for visual tracking. In: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV); 2017 Oct 22–29; Venice, Italy. p. 1135–43.
10. An Z, Wang X, Li B, Fu J. Learning spatial regularization correlation filters with the hilbert-schmidt independence criterion in RKHS for UAV tracking. IEEE Trans Instrum Meas. 2023;72:5011612. doi:10.1109/tim.2023.3265106.
11. Zhang F, Ma S, Zhang Y, Qiu Z. Perceiving temporal environment for correlation filters in real-time UAV tracking. IEEE Signal Process Lett. 2022;29:6–10. doi:10.1109/lsp.2021.3120943.
12. Henriques JF, Caseiro R, Martins P, Batista J. High-speed tracking with kernelized correlation filters. IEEE Trans Pattern Anal Mach Intell. 2015;37(3):583–96. doi:10.1109/tpami.2014.2345390.
13. Danelljan M, Häger G, Khan FS, Felsberg M. Learning spatially regularized correlation filters for visual tracking. In: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV); 2015 Dec 7–13; Santiago, Chile. p. 4310–8.
14. 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(12):8940–51. doi:10.1109/tgrs.2020.2992301.
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. 2020;59(8):6301–13. doi:10.1109/tgrs.2020.3030265.
16. 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); 2021 May 30–Jun 5; Xi'an, China. p. 503–9.
17. Danelljan M, Bhat G, Shahbaz Khan F, Felsberg M. ECO: efficient convolution operators for tracking. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017 Jul 21–26; Honolulu, HI, USA. p. 6638–46.
18. Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05); 2005 Jun 20–25; San Diego, CA, USA. p. 886–93.
19. Van De Weijer J, Schmid C, Verbeek J, Larlus D. Learning color names for real-world applications. IEEE Trans Image Process. 2009;18(7):1512–23. doi:10.1109/tip.2009.2019809.
20. Huang Z, Fu C, Li Y, Lin F, Lu P. Learning aberrance repressed correlation filters for real-time UAV tracking. In: Proceedings of the 2019 IEEE International Conference on Computer Vision (ICCV); 2019 Oct 27–Nov 2; Seoul, Republic of Korea. p. 2891–900.
21. Wei Q, Zeng B, Liu J, He L, Zeng G. LiteTrack: layer pruning with asynchronous feature extraction for lightweight and efficient visual tracking. In: Proceedings of the 2024 IEEE International Conference on Robotics and Automation (ICRA); 2024 May 13–17; Yokohama, Japan. p. 4968–75.
22. Ma S, Zhao Z, Hou Z, Zhang L, Yang X, Pu L. Correlation filters based on multi-expert and game theory for visual object tracking. IEEE Trans Instrum Meas. 2022;71:2510814. doi:10.1109/tim.2022.3178482.
23. Yu YF, Chen Z, Zhang Y, Zhang C, Ding W. Learning dynamic-sensitivity enhanced correlation filter with adaptive second-order difference spatial regularization for UAV tracking. IEEE Trans Intell Transp Syst. 2025;26(5):7211–30. doi:10.1109/tits.2025.3533953.
24. Zhang Y, Yu YF, Chen L, Ding W. Robust correlation filter learning with continuously weighted dynamic response for UAV visual tracking. IEEE Trans Geosci Remote Sens. 2023;61(15):4705814. doi:10.1109/tgrs.2023.3325337.
25. Zhang Y, Yu YF, Huang KK, Wang Y. Channel attentional correlation filters learning with second-order difference for UAV tracking. IEEE Geosci Remote Sens Lett. 2023;20:8001905. doi:10.1109/lgrs.2023.3311441.
26. Wen J, Chu H, Lai Z, Xu T, Shen L. Enhanced robust spatial feature selection and correlation filter learning for UAV tracking. Neural Netw. 2023;161(1):39–54. doi:10.1016/j.neunet.2023.01.003.
27. Yu YF, Tan X, Zhang Y, Chen L, Ding W. K-nearest neighbor correlation filters learning with p-Laplacian regularization for visual tracking. Tsinghua Sci Technol. 2025;38(3):100482. doi:10.26599/tst.2025.9010049.

28. Yu YF, Zhang Y, Chen L, Ge P, Chen CLP. Multi-scale enhanced features correlation filters learning with dual second-order difference for UAV tracking. IEEE Trans Intell Veh. 2024;9(2):3232–45. doi:10.1109/tiv.2024.3355171.
29. Fu C, Jin J, Ding F, Li Y, Lu G. Spatial reliability enhanced correlation filter: an efficient approach for real-time UAV tracking. IEEE Trans Multimed. 2024;26:4123–37.
30. Li S, Yang Y, Zeng D, Wang X. Adaptive and background-aware vision transformer for real-time UAV tracking. In: Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV); 2023 Oct 1–6; Paris, France. p. 13989–4000.
31. Lin F, Fu C, He Y, Guo F, Tang Q. BiCF: learning bidirectional incongruity-aware correlation filter for efficient UAV object tracking. In: Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA); 2020 May 31–Aug 31; Paris, France. p. 2365–71.
32. Lin J, Peng J, Chai J. Real-time UAV correlation filter based on response-weighted background residual and spatio-temporal regularization. IEEE Geosci Remote Sens Lett. 2023;20:6005405. doi:10.1109/lgrs.2023.3272522.
33. Li S, Yeung DY. Visual object tracking for unmanned aerial vehicles: a benchmark and new motion models. In: Proceedings of the 2017 Association for the advancement of artificial intelligence (AAAI); 2017 Feb 4–9; San Francisco, CA, USA. p. 4140–6.
34. Mueller M, Smith N, Ghanem B. A benchmark and simulator for UAV tracking. In: 2016 European Conference on Computer Vision (ECCV). Cham, Switzerland: Springer; 2016. p. 445–61.
35. Du D, Qi Y, Yu H, Yang Y, Duan K, Li G, et al. The unmanned aerial vehicle benchmark: object detection and tracking. In: 2018 European Conference on Computer Vision (ECCV). Cham, Switzerland: Springer; 2018. p. 370–86.
36. 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 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 13–19; Seattle, WA, USA. p. 11920–9.
37. 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(8):10469–80.
38. Ye J, Fu C, Lin F, Ding F, An S, Lu G. Multi-regularized correlation filter for UAV tracking and self-localization. IEEE Trans Ind Electron. 2021;69(6):6004–14. doi:10.1109/tie.2021.3088366.
39. Wang N, Song Y, Ma C, Zhou W, Liu W, Li H. Unsupervised deep tracking. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 Jun 15–20; Long Beach, CA, USA. p. 1308–17.
40. Zhao J, Dai K, Wang D, Lu H, Yang X. Online filtering training samples for robust visual tracking. In: Proceedings of the 2020 ACM International Conference on Multimedia (ACM MM); 2020 Oct 12–16; Seattle, WA, USA. p. 1488–96.
41. Wang B, Li W, Zhang B, Liu Y, Du J. Correlation filters for UAV online tracking based on complementary appearance model and reversibility reasoning. IEEE Trans Circuits Syst Video Technol. 2024;34(5):3983–97. doi:10.1109/tcsvt.2023.3325672.
42. Zhang J, He Y, Feng W, Wang J, Xiong NN. Learning background-aware and spatial-temporal regularized correlation filters for visual tracking. Appl Intell. 2023;53(7):7697–712. doi:10.1007/s10489-022-03868-8.
43. Wang N, Zhou W, Song Y, Ma C, Liu W, Li H. Unsupervised deep representation learning for real-time tracking. Int J Comput Vis. 2021;129(2):400–18. doi:10.1007/s11263-020-01357-4.
44. Sosnovik I, Moskalev A, Smeulders A. Scale equivariance improves siamese tracking. In: Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision (WACV); 2021 Jan 3–8; Waikoloa, HI, USA. p. 2764–73.
45. Avola D, Cinque L, Diko A, Fagioli A, Foresti GL, Mecca A, et al. MS-Faster R-CNN: multi-stream backbone for improved faster R-CNN object detection and aerial tracking from UAV images. Remote Sens. 2021;13(9):1670.
46. Xue Y, Shen T, Jin G, Tan L, Wang N, Wang L, et al. Handling occlusion in UAV visual tracking with query-guided redetection. IEEE Trans Instrum Meas. 2024;73(86):5030217. doi:10.1109/tim.2024.3440378.

47. Yan B, Peng H, Fu J, Wang D, Lu H. Learning spatio-temporal transformer for visual tracking. In: Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV); 2021 Oct 10–17; Montreal, QC, Canada. p. 10428–37.
48. Kang B, Chen X, Wang D, Peng H, Lu H. Exploring lightweight hierarchical vision transformers for efficient visual tracking. In: Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV); 2023 Oct 1–6; Paris, France. p. 9578–87.
49. Cui Y, Song T, Wu G, Wang L. Mixformerv2: efficient fully transformer tracking. Adv Neural Inf Process Syst. 2023;36:58736–51.
50. Wu Y, Wang X, Zeng D, Ye H, Xie X, Zhao Q, et al. Learning motion blur robust vision transformers with dynamic early exit for real-time UAV tracking. arXiv:2407.05383. 2024.
51. Li S, Yang X, Wang X, Zeng D, Ye H, Zhao Q. Learning target-aware vision transformers for real-time UAV tracking. IEEE Trans Geosci Remote Sens. 2024;62:4705718. doi:10.1109/tgrs.2024.3417400.
52. Yin K, Feng J, Dong S, Long Y. MMKLTrack: illumination-robust UAV tracking with progressive localization. Expert Syst Appl. 2026;300:130207.