Foreground Aware Correlation Filter with Adaptive Feature Response Fusion for Real-Time UAV Tracking

Zhuo Xiao
SKLSVMS, School of Aerospace
Xi'an Jiaotong University
Xi'an China
Zhuoxiao@stu.xjtu.edu.cn

Yi Yang
SKLSVMS, School of Aerospace
Xi'an Jiaotong University
Xi'an China
jiafeiyy@mail.xjtu.edu.cn

Sixian Zhang
SKLSVMS, School of Aerospace
Xi'an Jiaotong University
Xi'an China
touchmeteor@stu.xjtu.edu.cn

Wenbiao Li
SKLSVMS, School of Aerospace
Xi'an Jiaotong University
Xi'an China
li1476021027@stu.xjtu.edu.cn

Pengrong Bao
SKLSVMS, School of Aerospace
Xi'an Jiaotong University
Xi'an China
baobao111@stu.xjtu.edu.cn

Deqiang Han
School of Automation Science and Engineering
Xi'an Jiaotong University
Xi'an China
deqhan@gmail.com

***Abstract*—Background Aware Correlation Filter (BACF) tracker achieves accurate tracking result in visual object tracking by mitigating boundary effects, yet is limited in challenging scenarios especially in viewpoint change and illumination variation, which are frequently encountered in Unmanned Aerial Vehicle (UAV) tracking tasks. To address the shortcomings, we propose a Foreground Aware Correlation Filter with adaptive feature response fusion (FACF). In this paper, we use saliency detection to generate foreground prior knowledge in training phase for suppressing potential noise. Furthermore, recognizing the limitation of BACF, which relies on a single feature, a novel adaptive fusion strategy is designed to fuse multiple feature responses during the detection phase. This strategy aims to enhance the robustness of the tracker. Extensive experiments have been conducted on three challenging benchmarks. The tracking results show that the proposed tracker performs accurate and robust tracking result and satisfies real-time requirement with 48.28fps.**

Keywords—correlation filter, saliency detection, foreground-aware, adaptive feature response fusion

I. INTRODUCTION

Aerial visual object tracking is currently garnering a plethora of attention and undergoing rapid advancements within the realm of remote sensing technology[1]. However, numerous challenges arise when applying it to UAV tracking scenarios. UAVs obtain substantial video data in the course of executing tasks such as monitoring and navigation. Due to limitations of computing resources, UAV platforms are incapable of processing such large-scale data in real-time. Moreover, in comparison to conventional tracking scenarios, aerial visual object tracking tasks usually encounter more severe challenges, including but not limited to viewpoint change and illumination variation, which significantly impact tracking reliability. Therefore, there is an essential demand for algorithms that exhibit not only high performance but also low computational cost to meet these challenges.

Discriminative Correlation Filter (DCF) [2] based trackers have long been popular in visual object tracking due to their successful application of cyclic matrix theory, which allows them to strike an admirable balance between tracking speed and accuracy. Currently, Deep Learning (DL) techniques are widely used in visual object tracking for its powerful capability of feature representation [3]. However, due to their high computational cost, meeting real-time requirements in resource-constrained UAV platforms remains a challenge for DL-based solutions.

Since suffering from boundary effects, the performance of DCF-based trackers is limited. BACF improve negative sample quality through consideration of spatial relationships, which effectively mitigate boundary effects and gain superior tracking performance [4]. However, only HOG [5] feature is adopted into BACF, the tracker tend to degrade and perform poorly. Moreover, the fixed two-dimensional matrix mask in BACF tends to neglect potential noise in foreground region, which may leads to model drift and further tracking failure.

To conquer the limitations of BACF tracker mentioned above and improve the overall performance of UAV tracking, a Foreground Aware Correlation Filter with adaptive feature response fusion (FACF) is proposed in this paper. The main contributions are summarized as follows:

- A foreground aware mask based on saliency detection is proposed to make the tracker focus on tracked object and suppress potential noise introduced within the foreground region, thereby addressing challenges such as viewpoint change.
- A novel response fusion strategy is designed, which adaptively adjusts weights of response maps corresponding to different features and effectively alleviates the decrease of tracking performance caused by illumination variation, etc.
- Extensive experiments on three mainstream UAV tracking benchmarks (i.e., UAV123@10fps [6], DTB70 [7] and UAV112 [8]) with 305 UAV video sequences show advancement of our proposed trackers against other state-of-the-art methods.

The rest of this paper is organized subsequently: Section II reviews DCF trackers with multi-features and saliency. Section III describes the details of our tracker. Section IV

demonstrates our implementation and experimental results in existing mainstream on three benchmarks. Section V presents the conclusion.

II. Related Works

A. DCF Tracking with multi-features

DCF-based trackers are constructed by ridge regression in training phase and predict the object center position based on the maximum of the response in detection phase. Suffering from boundary effects, the performance of DCF-based trackers is still limited. For the sake of mitigating inaccurate negative training samples, Spatially Regularized Correlation Filter (SRDCF) [9] designs a spatial regularization to penalize the distribution of energy at edge and make filter focus on center region. BACF [4] introduces a two-dimensional cropping matrix to distinguish positive and negative samples, which further alleviate boundary effects. However, the single HOG [5] feature used in tracker makes BACF incapable of keeping stable tracking when facing challenging scenarios (i.e. illumination variation). Therefore, the combination of multi-features in correlation filters has attracted much research attention [10][11]. Online Multi Feature Learning (OMFL) [12] designs a strategy to adaptively fuse responses of multiple features based on BACF and enhanced the performance of the tracker. Although the feature response fusion strategy effectively mitigates model degradation, it inadequately account for temporal information in feature responses. When a distortion in a feature response yields high weight in the fusion process, the risk of tracking failure would increase.

B. DCF tracking with saliency

Saliency detection demonstrates superior performance when dealing with irregular object appearances especially in viewpoint change, leading to its widespread adoption in DCF-based trackers. OMFL [12] uses saliency map to train a sub-tracker, which complement other features during tracking process. Moreover, saliency detection is capable of providing prior knowledge in training phase and estimating possible appearance of object. Dynamic Saliency-Aware Regularized Correlation Filter (DSAR-CF) [14] adopt saliency detection into guiding the update of spatial regularization. Saliency-Aware Dual Regularized Correlation Filter (DRCF) [15] utilizes saliency map to penalize background noise in order to suppress boundary effects and background noise. Saliency-Aware Correlation Filter(SACF) [16] uses saliency detection to further enhance the discriminative capability of trackers. However, saliency detection can still hardly cope well with complex scenarios, which may mistake potential noise in foreground region as object appearance and further lead to model drift.

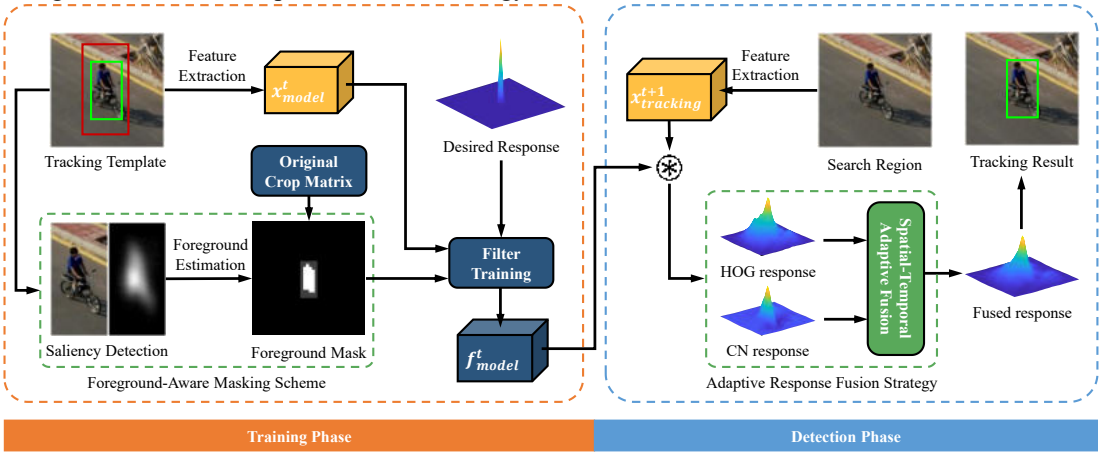


Fig. 1. Main workflow of the proposed FACF

III. Proposed Work

A. Motivation

BACF designed a two-dimensional binary cropping matrix to automatically crop the samples in the search area into multiple sub-regions of the same size as the target, reducing non-real samples caused by matrix cyclic sampling during the training process, allowing the filter to learn richer background information while reducing boundary effects. The filter is obtained by minimizing the following objective function:

$$\varepsilon(\boldsymbol{h}) = \frac{1}{2}\left\| \boldsymbol{y} - \sum_{k=1}^{K} \boldsymbol{x}_t^k \star (\boldsymbol{P}^{\mathrm{T}} \boldsymbol{h}_t^k) \right\|_2^2 + \frac{\lambda}{2} \sum_{k=1}^{K} \left\| \boldsymbol{h}_t^k \right\|_2^2 \qquad (1)$$

where $\boldsymbol{y} \in \mathbb{R}^{M \times N}$ is desired response that satisfies gaussian-shaped distribution, $\boldsymbol{x}_t^k \in \mathbb{R}^{M \times N} (k=1,2,...,K)$ denotes the K-dimensional feature sample extracted from the search region. $\star$ denotes the correlation operator. $\boldsymbol{h}_t^k \in \mathbb{R}^{M \times N} (k=1,2,...,K)$ denotes trained filter at t th frame. $\boldsymbol{P}^{\mathrm{T}} \in \mathbb{R}^{M \times N}$ is a binary matrix for cropping more complete negative samples. λ denotes regularization parameter.

Although boundary effects got suppressed to a certain extent, there is still potential noise existing in foreground region, which would lead to model drift. Moreover, the singularity of HOG feature makes BACF incapable of handling challenging scenarios. Therefore, in this paper, we propose a tracker for addressing problems mentioned above. The workflow of the proposed tracker is as shown in Fig. 1.

B. Foreground-Aware Masking Scheme

For the sake of boundary effects, BACF crops the target area to avoid the response to non-real samples in the background area. In practical tracking process, the target area still contains potential noise while the fixed two-dimensional matrix mask in BACF neglect interference factors in it. These interference factors tend to make the filter gradually degrade in complex scenarios especially in viewpoint change, causing tracking drift or even tracking failure, affecting the accuracy of the tracker. In order to solve such problem, in this paper we

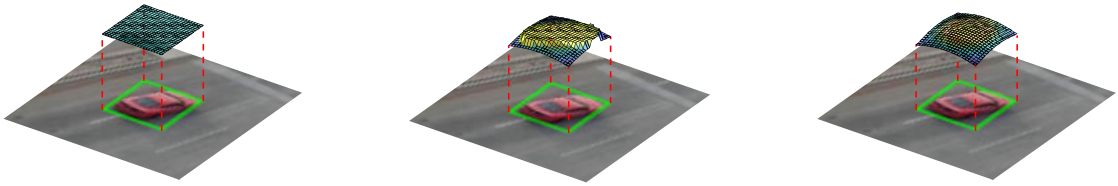
propose a foreground-aware masking scheme based on saliency detection to reduce the impact of potential noise in foreground region on tracking accuracy.

Based on the two-dimensional cropping matrix of BACF, the proposed foreground-aware mask at object region is defined as follows:

$$\boldsymbol{w}_s = \begin{cases} 1 & , w_{(x,y)} > \varphi \\ \exp\left(-\eta \cdot \left(\frac{x^2}{m^2} + \frac{y^2}{n^2} \right) \right) & , others \end{cases} \tag{2}$$

where $w_{(x,y)}$ denotes the value corresponding to the location (x,y) in the saliency map. m and n denote the width and height of the cropped sample patch. φ is the predefined threshold.

In our method, the saliency detection method is used to extract foreground information within the current search region. Pixels with value higher than φ in the saliency map is defined as 1 at the corresponding position in the foreground-aware mask while pixels lower than the threshold satisfy Gaussian distribution to weaken the interference factors of potential noise. To explain our method concretely, Fig. 2 shows the visualization of each mask separately. As shown in the figure, our foreground-aware mask effectively recognized salient foreground region and distribute weights more reasonably in region space compared with other strategies.



(a) crop matrix mask of BACF (b) foreground-aware mask (c) gaussian-shaped mask

Fig. 2. The visualization of foreground-aware mask

C. Adaptive Feature Response Fusion Strategy

The HOG feature can well describe the detailed texture information of the object, but when dealing with viewpoint change scenes, the tracking accuracy of the algorithm model will be greatly reduced. Correspondingly, The Color Names (CN) feature can effectively exploits color sematic information, yet may lead to decrease of model robustness when faced with illumination variation caused by rapid change of object color. Compared with other handicraft features, these two have superior complementarity and the combination of them can gain more accurate description of objects. Although feature response fusion strategy uses the complementarity between features to improve the robustness, utilizing fixed weights or only considering spatial information cannot effectively deal with the distortion of a certain feature response, which may lead to tracking failure.

In response to such problems, we propose an adaptive feature response fusion strategy: using Peak-to-Sidelobe Ratio (PSR) [12] to evaluate the spatial reliability of the responses, and designing a Temporal Response Smooth Factor (TRSF) to evaluate the temporal reliability:

$$\begin{cases} PSR = \dfrac{R_t^{\max} - \mu_t}{\sigma_t} \\ TRSF = \dfrac{R_t^{\max}}{\left\| R_t - R_{t-1}[\Delta\phi] \right\|_2} \end{cases} \tag{3}$$

where $R_t^{\max}$, μ_t and σ_t respectively denote the maximum value, mean value, and standard deviation of the response map of t th frame. R_t and R_{t-1} respectively represent the response map of t th and $(t-1)$ th frame. $[\Delta\phi]$ indicates the shifting operation in order to make the peaks of two response maps located at the same position to eliminate the impact of motion displacement.

The confidence of response is defined as follows:

$$\begin{cases} c_{HOG} = \rho \cdot PSR_{HOG} + (1-\rho) \cdot TRSF_{HOG} \\ c_{CN} = \rho \cdot PSR_{CN} + (1-\rho) \cdot TRSF_{CN} \end{cases} \tag{4}$$

where c_{HOG} and c_{CN} denote the reliability of the response corresponding to HOG and CN features, respectively. ρ represents the weight for balancing spatial and temporal reliabilities.

When the response map corresponding to a certain feature mutates or the overall spatial reliability decreases, its confidence will become relatively low, thereby occupying a lower weight in the fusion strategy and improving the robustness of the model during the detection phase. The final fused response map that introduces the adaptive fusion strategy is defined as follows:

$$R_t^{fuse} = c_t^{HOG} \cdot R_t^{HOG} + c_t^{CN} \cdot R_t^{CN} \tag{5}$$

where, R_t^{fuse} denotes fused response map. R_t^{HOG} and R_t^{CN} represent the response maps corresponding to the HOG and CN features respectively. c_t^{HOG} and c_t^{CN} denote the confidence factors corresponding to each response maps of t th frame.

D. Optimization

Replace the crop matrix in (1) with our foreground-aware mask, the objective function can be rewritten as follows:

$$\varepsilon(\boldsymbol{h}) = \frac{1}{2} \left\| \boldsymbol{y} - \sum_{k=1}^{K} \boldsymbol{x}_t^k \star (\boldsymbol{w}^{\mathrm{T}} \boldsymbol{h}_t^k) \right\|_2^2 + \frac{\lambda}{2} \sum_{k=1}^{K} \left\| \boldsymbol{h}_t^k \right\|_2^2 \tag{6}$$

where $\boldsymbol{w}$ is the foreground-aware mask. Assuming $\boldsymbol{g} = \boldsymbol{w}^{\mathrm{T}} \boldsymbol{h}$ as auxiliary variable, (5) can be rewritten in frequency domain as follows:

$$\varepsilon(\hat{\boldsymbol{g}}, \boldsymbol{h}) = \frac{1}{2} \left\| \hat{\boldsymbol{y}} - \sum_{K=1}^{K} \hat{\boldsymbol{x}}_t^k \hat{\boldsymbol{g}}_t^k \right\|_2^2 + \frac{\lambda}{2} \sum_{k=1}^{K} \left\| \boldsymbol{h}_t^k \right\|_2^2 \tag{7}$$

where $(\cdot)$ represents the Discrete Fourier Transform (DFT) operator. For optimization, the alternating direction method of multipliers (ADMM) is utilized to update each variables iteratively and (6) can be written as follows:

$$\begin{aligned}\varepsilon\left(\hat{\boldsymbol{g}},\boldsymbol{h},\hat{\boldsymbol{\zeta}}\right)=&\frac{1}{2T}\left\|\hat{\boldsymbol{y}}-\sum_{k=1}^{K}\hat{\boldsymbol{x}}_t^k\odot\hat{\boldsymbol{g}}_t^k\right\|_2^2+\frac{\lambda}{2}\sum_{k=1}^{K}\left\|\boldsymbol{h}_t^k\right\|_2^2\\&+\frac{\mu}{2}\sum_{k=1}^{K}\left\|\hat{\boldsymbol{g}}_t^k-\sqrt{T}F\boldsymbol{h}_t^k\right\|_2^2+\hat{\boldsymbol{\zeta}}^T\sum_{k=1}^{K}(\hat{\boldsymbol{g}}_t^k-\sqrt{T}F\boldsymbol{h}_t^k)\end{aligned}\tag{8}$$

where $\hat{\boldsymbol{\zeta}}=[\hat{\boldsymbol{\zeta}}^1,\hat{\boldsymbol{\zeta}}^2,\hat{\boldsymbol{\zeta}}^3,...,\hat{\boldsymbol{\zeta}}^K]$ denotes Lagrange multipliers and μ is the penalty coefficient. Using ADMM, Closed form solution can be found by solving each sub-problems.

1) Subproblem $\hat{\boldsymbol{g}}_k^*$

$$\begin{aligned}\hat{\boldsymbol{g}}_k^*=\arg\min_{\hat{\boldsymbol{g}}_k}\Bigg\{&\frac{1}{2T}\left\|\hat{\boldsymbol{x}}_k^{\mathrm{H}}\odot\hat{\boldsymbol{g}}_k-\hat{\boldsymbol{y}}\right\|_2^2\\&+\frac{\mu}{2}\left\|\hat{\boldsymbol{g}}_k-\sqrt{T}F\boldsymbol{h}_k\right\|_2^2+\hat{\boldsymbol{\zeta}}^{\mathrm{T}}(\hat{\boldsymbol{g}}_k-\sqrt{T}F\boldsymbol{h}_k)\Bigg\}\end{aligned}\tag{9}$$

Since burden of directly solving (8), the subproblem $\hat{\boldsymbol{g}}_k^*$ can be decomposed into N problems as follows:

$$\begin{aligned}\hat{\boldsymbol{g}}_k^*(n)=\arg\min_{\hat{\boldsymbol{g}}_k}\Bigg\{&\frac{1}{2T}\left\|\hat{\boldsymbol{x}}_k^{\mathrm{H}}(n)\odot\hat{\boldsymbol{g}}_k(n)-\hat{\boldsymbol{y}}(n)\right\|_2^2\\&+\frac{\mu}{2}\left\|\hat{\boldsymbol{g}}_k(n)-\sqrt{T}F\boldsymbol{h}_k(n)\right\|_2^2+\hat{\boldsymbol{\zeta}}^{\mathrm{T}}\left[\hat{\boldsymbol{g}}_k(n)-\sqrt{T}F\boldsymbol{h}_k(n)\right]\Bigg\}\end{aligned}\tag{10}$$

where $\hat{\boldsymbol{g}}_k(n)=\left[\hat{\boldsymbol{g}}_1(n),\hat{\boldsymbol{g}}_2(n),...,\hat{\boldsymbol{g}}_K(n)\right]^{\mathrm{T}}$ and $\hat{\boldsymbol{h}}_k(n)=\sqrt{T}F\boldsymbol{h}_k(n)$. The closed solution for $\hat{\boldsymbol{g}}_k^*(n)$ can be accelerated by Sherman-Morrison formula:

$$\begin{aligned}\hat{\boldsymbol{g}}_k^*(n)=&\frac{1}{\mu}\left(\frac{1}{T}\hat{\boldsymbol{x}}_k(n)\hat{\boldsymbol{y}}(n)+\mu\hat{\boldsymbol{h}}_k(n)-\hat{\boldsymbol{\zeta}}\right)\\&-\frac{\hat{\boldsymbol{x}}_k(n)}{\mu b}\left(\frac{1}{T}\hat{\boldsymbol{S}}(n)\hat{\boldsymbol{y}}(n)+\mu\hat{\boldsymbol{x}}_k(n)^{\mathrm{T}}\hat{\boldsymbol{h}}_k(n)-\hat{\boldsymbol{x}}_k(n)^{\mathrm{T}}\hat{\boldsymbol{\zeta}}_k(n)\right)\end{aligned}\tag{11}$$

where $\hat{\boldsymbol{S}}(n)=\hat{\boldsymbol{x}}_k(n)^{\mathrm{T}}\hat{\boldsymbol{x}}_k(n)$ and $b=\hat{\mathbf{S}}(n)+\mu T$.

2) Subproblem $\boldsymbol{h}_k^*$

$$\begin{aligned}\boldsymbol{h}_k^*=\arg\min_{\boldsymbol{h}_k}\Bigg\{&\frac{\lambda}{2}\sum_{k=1}^{K}\left\|\boldsymbol{h}_k\right\|_2^2+\frac{\mu}{2}\sum_{k=1}^{K}\left\|\hat{\boldsymbol{g}}_k-\sqrt{T}F\boldsymbol{h}_k\right\|_2^2\\&+\hat{\boldsymbol{\zeta}}^T\sum_{k=1}^{K}(\hat{\boldsymbol{g}}_k-\sqrt{T}F\boldsymbol{h}_k)\Bigg\}\end{aligned}\tag{12}$$

The variable $\boldsymbol{h}_k^*$ can be obtained as:

$$\boldsymbol{h}_k^*=\frac{\mu\hat{\boldsymbol{g}}_k+\boldsymbol{\zeta}}{\lambda/T+\mu}\tag{13}$$

3) Subproblem $\hat{\boldsymbol{\zeta}}$

The Lagrangian parameter is updated according to:

$$\hat{\boldsymbol{\zeta}}_k^{(i+1)}=\hat{\boldsymbol{\zeta}}_k^{(i)}+\mu(\hat{\boldsymbol{g}}_k^{(i+1)}-\hat{\boldsymbol{h}}_k^{(i+1)})\tag{14}$$

IV. Experiments

In this section, we evaluate the proposed FACF on three challenging UAV benchmarks (UAV123_10fps [6], DTB70 [7] and UAV112 [8]) with a total number of 305 UAV video sequences. We compare the tracker with 15 mainstream trackers and analyze the contribution of different modules to tracking performance.

A. Evaluation Metrics

For evaluation, we adopt the One Pass Evaluation (OPE), which includes two metrics (precision and success rate) for quantitative evaluation of the performance. The precision is defined by the center location error (CLE) between the ground truth and estimated bounding boxes. The percentages of frames with CLE lower than predefined threshold during whole tracking process are utilized to rank trackers. Moreover, the Intersection over Union (IoU) score between ground truth and estimated bounding box is adopted for success plot. Note that the area under the curve (AUC) on success plot is utilized to rank trackers.

B. Implementation

All experiments in this paper are conducted on a computer with an Intel(R) Core(TM) i5-8300H CPU and 16GB RAM. The main parameters of the proposed method are as follows: $\lambda=0.01$, response fusion weight coefficient ϕ is set to 0.4 and the learning rate of model apearance $\eta=0.0199$. For the update of ADMM, we set $\gamma_{initial}=1$, $\sigma=10$, $\gamma_{\max}=1000$, the number of iterations is 2.

C. Quantitative Evaluation

1) Overall Analysis

In this paper, we select 15 mainstream hand-crafted DCF-based trackers i.e., MRCF [17], DRCF [15], OMFL [12], MSCF [18], ECO_HC [19], STRCF [20], EFSCF [21], ARCF_H [22], MCCT_H [23], Staple_CA [24], SRDCF [9], Staple [10], BACF [4], fDSST [25], SAMF [26], KCF [2], and proposed FACF in this work for evaluation on three UAV tracking benchmarks and the results of the tracking performances are shown in Fig. 3. From the results we can see that our proposed FACF tracker perform excellently among the trackers.

In Fig. 3(a) and Fig. 3(d), FACF improves the baseline BACF by 10.7% in precision and 7.1% in AUC. Compared with OMFL, which uses multiple features for adaptive response fusion, FACF gains 7.9% and 6.2% improvements in precision and AUC respectively as shown in Fig. 3(b) and Fig. 3(e). Notice that DTB70 benchmark includes a lot of scenarios of viewpoint change and illumination variation. Moreover, FACF improves 2.7% and 1.7% to the performance of DRCF (saliency-based method) in precision and AUC in Fig. 3(c) and Fig. 3(f). In conclusion, FACF shows an excellent performance in both accuracy and robustness.

2) Attributes Analysis

To demonstrate the capability of the proposed FACF tracker, we select five trackers of the same type for comparison on the UAV123_10fps benchmark [6]. The dataset includes 123 UAV video sequences subsampled from the original sequences of the UAV123 dataset. The data volume is 1/3 of the original dataset (a total of 37885 frames). The motion and attitude of the tracked object change between adjacent frames is greater, making it more challenging. There are 12 challenging attributes, including aspect ratio change (ARC), background clutter (BC), camera motion (CM), fast motion (FM), full occlusion (FOC), illumination variation (IV), low resolution (LR), out of view (OV), partial occlusion (PO), similar object (SO), viewpoint change (VC). We conduct comparison of precision and success rate on mentioned attributes to verify the advancement of our trackers. As shown in Fig. 4, the proposed FACF demonstrates prominent advantages in different attributes especially under VC and IV.

3) Speed Analysis

Under conditions of limited hardware resources, it is essential for trackers to be engineered with minimized computational expense in order to fulfill the practical demands of real-world applications. Thus, to substantiate the efficiency of our proposed FACF, a comparative analysis of their speeds is conducted using the UAV123_10fps benchmark. As shown in TABLE I. , the speed of the proposed FACF stands out as the fourth-fastest among listed mainstream trackers, reaching 48.29 fps, which obviously satisfies the requirements of real-time tracking.

TABLE I. SPEED COMPARISON

Trackers	Speed (fps)	Trackers	Speed (fps)
FACF	**48.29**	STRCF	21.45
BACF	36.87	ARCF_H	34.63
SRDCF	11.51	MCCT_H	43.87
ECO_HC	53.40	OMFL	12.57
Staple	80.74	KCF	593.41
Staple_CA	46.53	DRCF	29.18
MRCF	31.11	fDSST	114.26
EFSCF	20.33	MSCF	20.99

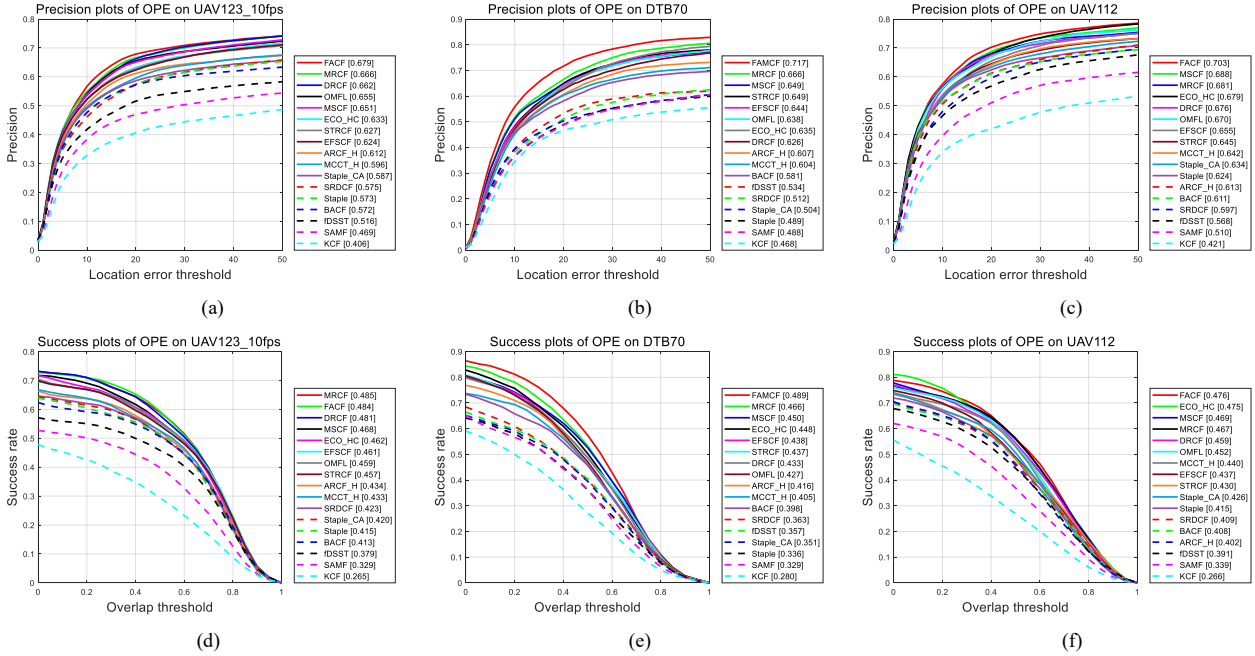


Fig. 3. Overall performance comparison of the proposed FACF tracker with other 15 mainstream trackers

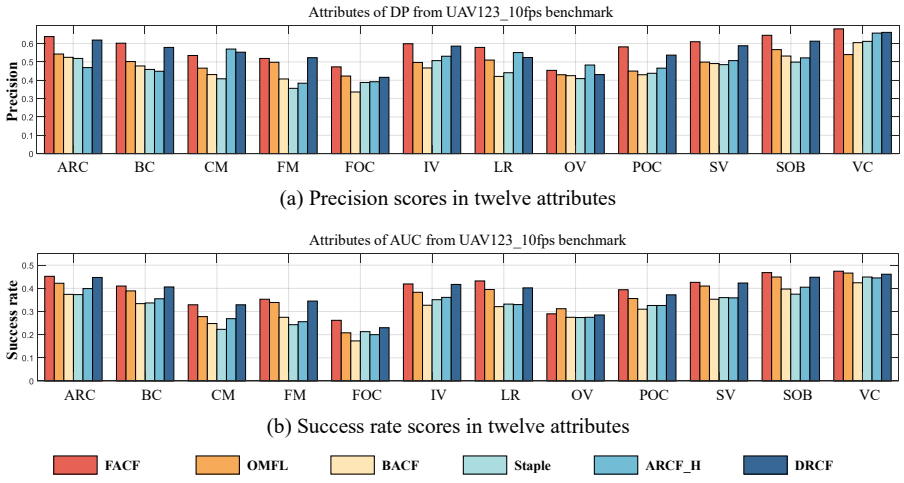


Fig. 4. Performance scores in 12 attributes on UAV123_10fps benchmark

D. Qualitative Evaluation

A series of qualitative tracking results of FACF and other five mainstream trackers of the same type, i.e., OMFL, BACF, Staple, ARCF_H, DRCF, are shown in Fig. 5. The visualization of tracking examples reveals the proposed FACF performs better than other trackers utilized for comparison in challenging scenarios especially in viewpoint change and illumination variation.

E. Ablation Analysis

For the verification of each parts of our proposed method, the performance of baseline with various modules enabled is conducted on DTB70 benchmark [7]. The results are displayed in TABLE II. With the activation of the Adaptive Feature Response Fusion Strategy (AFRFS) and the Foreground-Aware Masking Scheme (FAMS) in the baseline, the performance of the tracker improves significantly,

Compared with the baseline tracker, the overall precision and AUC of the proposed FACF exceed the baseline by 4.5% and 1.2%. Obviously both modules contribute to boosting the tracking performance.

(a) Wakeboard8

(b) RcCar4

(c) SnowBoarding4

Fig. 5. Visualization of tracking examples

TABLE II. ABALATION ANAYSIS OF THE PROPOSED FACF

Trackers	Precision	AUC	Speed (fps)
Baseline	0.672	0.477	62.30
Baseline+AFRFS	0.701	0.482	52.49
Baseline+FAMS	0.684	0.473	56.26
FACF	0.717	0.489	48.29

V. CONCLUSIONS

In this paper, we propose a foreground-aware correlation filter with adaptive feature response fusion for real-time UAV tracking. Extensive experiments on three UAV tracking benchmarks were conducted and the results demonstrate that the proposed method significantly improves performance of the baseline BACF. Moreover, the proposed tracker perform better than mainstream trackers in terms of precision, success rate and speed, which is applicable to aerial tracking with UAV platforms.

REFERENCES

[1] C. Fu, B. Li, F. Ding, F. Lin, and G. Lu, "Correlation filters for unmanned aerial vehicle-based aerial tracking: A review and experimental evaluation," IEEE Geoscience and Remote Sensing Magazine, vol. 10, no. 1, pp. 125-160, 2021.

[2] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters, " IEEE Transactions on Pattern Analysis Machine Intelligence, vol. 37, no. 3, pp. 583-596, March 2015

[3] H. Nam and B. Han, "Learning multi-domain convolutional neural networks for visual tracking," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 4293-4302.

[4] H. K. Galoogahi, A. Fagg, and S. Lucey, "Learning background-aware correlation filters for visual tracking," in 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 1144-1152.

[5] N. Dalai, B. Triggs, "Histograms of oriented gradients for human detection," in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2005, vol.1: IEEE, pp. 886-893.

[6] M. Mueller, N. Smith, and B. Ghanem, "A benchmark and simulator for uav tracking," in 14th European Conference on Computer Vision (ECCV), 2016, Proceedings, Part I 14, 2016: Springer, pp. 445-461.

[7] S. Li and D.-Y. Yeung, "Visual object tracking for unmanned aerial vehicles: A benchmark and new motion models," in Proceedings of the AAAI Conference on Artificial Intelligence, 2017, vol. 31, no. 1.

[8] C. Fu, Z. Cao, Y. Li, J. Ye, and C. Feng, "Onboard real-time aerial tracking with efficient Siamese anchor proposal network," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-13, 2021.

[9] M. Danelljan, G. Hager, F. S. Khan and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," in 2015 IEEE International Conference on Computer Vision(ICCV), pp. 4310-4318.

[10] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, and P. H. S. Torr, "Staple: complementary learners for real-time tracking," in 2016 IEEE conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1401-1409.

[11] N. Wang, W. Zhou, Q. Tian, R. Hong, M. Wang, and H. Li, "Multi-cue correlation filters for robust visual tracking," in Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 4844-4853.

[12] C. Fu, F. Lin, Y. Li, and G. Chen, "Correlation Filter-Based Visual Tracking for UAV with Online Multi-Feature Learning," Remote Sensing, vol. 11, no. 5, 2019.

[13] M. Danelljan, F. S. Khan, M. Felsberg and J. Van De Weijer, "Adaptive color attributes for real-time visual tracking," in the proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 1090-1097.

[14] W. Feng, R. Han, Q. Guo, J. Zhu, and S. Wang, "Dynamic Saliency-Aware Regularization for Correlation Filter-Based Object Tracking," IEEE Trans Image Process, vol. 28, no. 7, pp. 3232-3245, 2019.

[15] C. Fu, J. Xu, F. Lin, F. Guo, T. Liu, and Z. Zhang, "Object Saliency-Aware Dual Regularized Correlation Filter for Real-Time Aerial Tracking," IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 12, pp. 8940-8951, 2020.

[16] Y. Wang, F. Wang, C. Wang, F. Sun, and J. He, "Learning Saliency-Aware Correlation Filters for Visual Tracking," The Computer Journal, vol. 65, no. 7, pp. 1846-1859, 2021.

[17] J. Ye, C. Fu, F. Lin, F. Ding, S. An, and G. Lu, "Multi-Regularized Correlation Filter for UAV Tracking and Self-Localization," IEEE Transactions on Industrial Electronics, vol. 69, no. 6, pp. 6004-6014, 2022.

[18] G. Zheng, C. Fu, J. Ye, F. Lin, and F. Ding, "Mutation sensitive correlation filter for real-time UAV tracking with adaptive hybrid label," in 2021 IEEE International Conference on Robotics and Automation (ICRA), 2021: IEEE, pp. 503-509.

[19] M. Danelljan, G. Bhat, F. S. Khan, and M. Felsberg, "ECO: Efficient Convolution Operators for Tracking," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6931-6939.

[20] F. Li, C. Tian, W. Zuo, L. Zhang, and M.-H. Yang, "Learning spatial-temporal regularized correlation filters for visual tracking," in Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 4904-4913.

[21] J. Wen, H. Chu, Z. Lai, T. Xu, and L. Shen, "Enhanced robust spatial feature selection and correlation filter learning for UAV tracking," Neural Networks, vol. 161, pp. 39-54, 2023.

[22] Z. Huang, C. Fu, Y. Li, F. Lin, and P. Lu, "Learning Aberrance Repressed Correlation Filters for Real-Time UAV Tracking," in 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 2891-2900.

[23] N. Wang, W. Zhou, Q. Tian, R. Hong, M. Wang, and H. Li, "Multi-cue correlation filters for robust visual tracking," in Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2018, pp. 4844-4853.

[24] M. Matthias, S. Neil, G. Bernard, "Context-Aware Correlation Filter Tracking" in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1396-1404.

[25] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, "Discriminative Scale Space Tracking," IEEE Trans Pattern Anal Mach Intell, vol. 39, no. 8, pp. 1561-1575, 2017.

[26] M. Danelljan, G. Häger, F. Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," in British machine vision conference, 2014, pp.254-265.