Perceiving Temporal Environment for Correlation Filters in Real-Time UAV Tracking

Fei Zhang, Shiping Ma, Yule Zhang, and Zhuling Qiu

Abstract—**Discriminative correlation filter (DCF)-based methods applied for UAV object tracking have received widespread attention due to their high efficiency. However, these methods are usually troubled by the boundary effect. Besides, the violent environment variations severely confuse trackers that neglect temporal environmental changes among consecutive frames, leading to unwanted tracking drift. In this letter, we propose a novel DCF-based tracking method to promote the insensitivity of the tracker under uncertain environmental changes. Specifically, a regularization term is proposed to learn the environment residual between two adjacent frames, which can enhance the discrimination and insensitivity of the filter in fickle tracking scenarios. Further, we design an efficient strategy to acquire the environment information based on the current observation without additional computation. Exhausted experiments are conducted on two well-known UAV benchmarks, *i.e.*, UAV123_10fps and DTB70. Results verify that the proposed tracker has comparable performance with other 22 state-of-the-art trackers while running at $\sim$ 53 FPS on a low-cost CPU.**

Index Terms—**Visual tracking, unmanned aerial vehicle, discriminative correlation filter, residual learning.**

I. Introduction

As ONE of the most vital members in the visual system of the unmanned aerial vehicle (UAV), visual tracking has attracted increasing research interest [1]–[3] in recent years. Its task is to estimate the target position in subsequent videos under the condition that the ground truth in the first frame is known. Discriminative correlation filter (DCF)-based trackers [1]–[7], especially with handcrafted features [3]–[6], have been active in UAV tracking community on account of their high efficiency on a single CPU. However, UAV-specific issues, *e.g.*, fast object motion, camera viewpoint change, and limited resources, require the tracker to possess high performance as well as low energy loss.

The core idea of DCF paradigm [8]–[10] is using ridge regression to train a filter for distinguishing the target from its environment. It is well recognized that the environmental information is of vital for raising the tracking performance. To weaken the environmental interference, Mueller *et al.* [10] took the global environment into consideration and added multiple regularization terms to suppress the response of the filter to the environment. However, the tracker [10] only considers the current environment, which causes over-fitting and overlooks the temporal environmental changes. In real-world UAV tracking, the environment of the tracked target usually undergoes drastic and uncertain changes. It is a potential factor for the tracker that is sensitive to environment changes to occur tracking drift. Furthermore, the acquisition of environmental information in [10] has the following drawbacks. i) It is time-consuming to extract features of multiple environment samples, which also impedes the fast optimization of the filter. ii) These environment samples may introduce irrelevant information outside the search region as they have the same size as the training samples of the target. Later, meaningful attempts [5], [7], [11] have been made by selecting more reasonable environment samples. Nevertheless, the above mentioned problems are still not solved.

To the above concerns, this letter proposes a novel UAV tracking algorithm with **e**nvironmental **m**utation-insensitive **c**orrelation **f**ilters (EMCF). Specifically, a simple method is designed to efficiently obtain the environmental information based on the training sample of each frame. Then, the environment residual is computed and then fed into the training phase to learn a more powerful filter. Thus, the proposed tracker can maintain robustness in violent environments. Extensive and comprehensive experiments on two challenging benchmarks, *i.e.*, UAV123_10fps [12] and DTB70 [13], manifest the effectiveness of the proposed method. Our tracker has comparable accuracy with other state-of-the-art (SOTA) trackers while possessing satisfactory tracking speed on a single CPU. Fig. 1 shows the tracking pipeline of the proposed MECF.

The contributions of this work are summarized as follows:

- A novel regularization term to learn environment residual is proposed to keep the tracker insensitive to drastic environmental changes.
- We obtain the environmental information off the shelf without additional feature extraction via a simple yet efficient method.
- Experimental results demonstrate that our tracker achieves promising performance and is suitable for real-time onboard UAV platforms.

II. Methods

A. Mutation-Insensitive Correlation Filter

We select the BACF [14] tracker as our baseline. A novel environmental mutation-insensitive regularization is added to strengthen the discrimination of the tracker. The desired filter $\mathbf{h} = [\mathbf{h}^1, \mathbf{h}^2, \ldots, \mathbf{h}^{\mathrm{D}}] \in \mathbb{R}^{\mathrm{N}\times\mathrm{D}}$ can be optimized through minimizing the following objective

Manuscript received August 23, 2021; revised October 12, 2021; accepted October 12, 2021. Date of publication October 19, 2021; date of current version January 20, 2022. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Kunal Narayan Chaudhury. *(Corresponding author: Fei Zhang.)*

Fei Zhang and Yule Zhang are with the Graduate College, Air Force Engineering University, Xi'an 710038, China (e-mail: kgfzhang@163.com; yule_zhang0921@163.com).

Shiping Ma and Zhuling Qiu are with the Aeronautical Engineering College, Air Force Engineering University, Xi'an 710038, China (e-mail: maship-ing@126.com; kgdqzl@163.com).

Digital Object Identifier 10.1109/LSP.2021.3120943

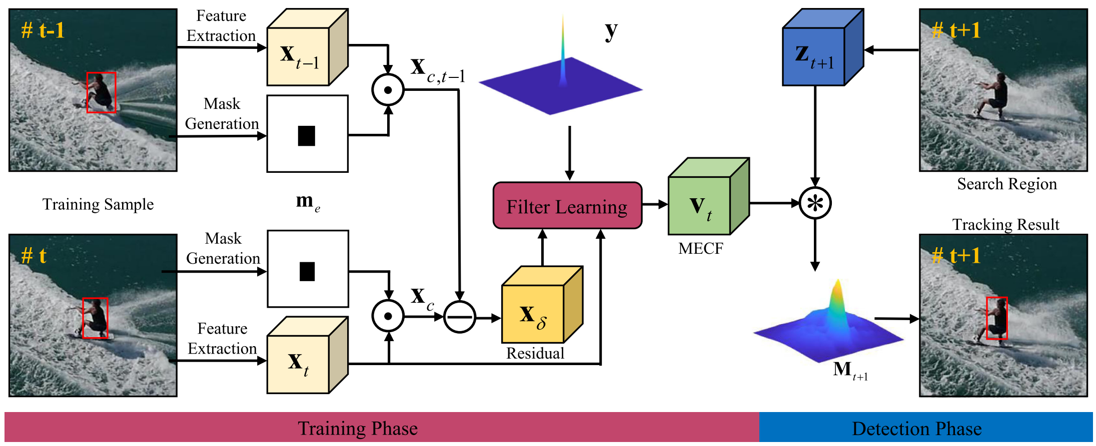

Fig. 1. Overall flowchart of the proposed MECF. In the training phase: Our tracker is trained with the environment residual information, which can be efficiently acquired through fast environment awareness. In the detection phase: the MECF tracker performs the tracking task.

function:

$$\Pi(\mathbf{h}) = \frac{1}{2}\left\|\sum_{d=1}^{\mathrm{D}} \mathbf{B}\mathbf{x}^d \circledast \mathbf{h}^d - \mathbf{y}\right\|_2^2 + \frac{\lambda}{2}\sum_{d=1}^{\mathrm{D}}\left\|\mathbf{h}^d\right\|_2^2 + \frac{\gamma}{2}\left\|\sum_{d=1}^{\mathrm{D}} \mathbf{B}(\mathbf{x}_c^d - \mathbf{x}_{c,t-1}^d) \circledast \mathbf{h}^d\right\|_2^2 , \tag{1}$$

where $\mathbf{x}^d \in \mathbb{R}^{\mathrm{N}}$ and $\mathbf{x}_c^d \in \mathbb{R}^{\mathrm{N}}$ are the feature maps of object patch and environment patch in the d-th channel, respectively. D is the total number of channels. $\mathbf{B} \in \mathbb{R}^{\mathrm{M}\times\mathrm{N}}(\mathrm{M} \ll \mathrm{N})$ represents a binary matrix to acquire the middle elements of $\mathbf{x}$. $\circledast$ denotes the correlation operator. λ is the regularization parameter. γ is the sensitivity factor, which controls the sensitivity of the filter to the environment. $t-1$ means the last frame.

Due to the temporal continuity of video sequences, it is reasonable to assume that the environment changes in short intervals are similar. Through learning the environment residual $\mathbf{x}_c - \mathbf{x}_{c,t-1}$, the proposed tracker can maintain robustness to environmental change in the next frame.

Since the correlation calculation can be efficiently processing in Fourier domain, the optimization will be carried out in Fourier form. Introducing an auxiliary variable $\mathbf{v}^d = \mathbf{B}^{\mathsf{T}}\mathbf{h}^d \in \mathbb{R}^{\mathrm{N}}$ and denoting the environment residual $\mathbf{x}_\delta = \mathbf{x}_c - \mathbf{x}_{c,t-1}$, the Fourier form of (1) can be written as:

$$\Pi(\mathbf{h}, \hat{\mathbf{v}}) = \frac{1}{2\mathrm{N}}\left\|\sum_{d=1}^{\mathrm{D}} \hat{\mathbf{x}}^d \odot \hat{\mathbf{v}}^d - \hat{\mathbf{y}}\right\|_2^2 + \frac{\lambda}{2}\sum_{d=1}^{\mathrm{D}}\left\|\mathbf{h}^d\right\|_2^2 + \frac{\gamma}{2\mathrm{N}}\left\|\sum_{d=1}^{\mathrm{D}} \hat{\mathbf{x}}_\delta^d \odot \hat{\mathbf{v}}^d\right\|_2^2 , \tag{2}$$

where $\odot$ is the Hadamard product and the superscript $\hat{}$ is the DFT operator.

Then, the Augmented Lagrangian form of (2) can be formulated as:

$$\begin{aligned}\Pi(\mathbf{h}, \hat{\mathbf{v}}, \hat{\boldsymbol{\zeta}}) = {} & \frac{1}{2\mathrm{N}}\left\|\sum_{d=1}^{\mathrm{D}} \hat{\mathbf{x}}^d \odot \hat{\mathbf{v}}^d - \hat{\mathbf{y}}\right\|_2^2 \\ & + \frac{\lambda}{2}\sum_{d=1}^{\mathrm{D}}\left\|\mathbf{h}^d\right\|_2^2 + \frac{\gamma}{2\mathrm{N}}\left\|\sum_{d=1}^{\mathrm{D}} \hat{\mathbf{x}}_\delta^d \odot \hat{\mathbf{v}}^d\right\|_2^2 \\ & + \frac{\mu}{2}\sum_{d=1}^{\mathrm{D}}\left\|\hat{\mathbf{v}}^d - \sqrt{\mathrm{N}}\mathbf{F}\mathbf{B}^{\mathsf{T}}\mathbf{h}^d\right\|_2^2 \\ & + \sum_{d=1}^{\mathrm{D}}(\hat{\mathbf{v}}^d - \sqrt{\mathrm{N}}\mathbf{F}\mathbf{B}^{\mathsf{T}}\mathbf{h}^d)^{\mathsf{T}}\hat{\boldsymbol{\zeta}}^d ,\end{aligned} \tag{3}$$

where $\boldsymbol{\zeta} = [\boldsymbol{\zeta}^1, \boldsymbol{\zeta}^2, \dots, \boldsymbol{\zeta}^{\mathrm{D}}] \in \mathbb{R}^{\mathrm{N}\times\mathrm{D}}$ and μ are the Lagrangian vector and the penalty factor, respectively.

Next, (3) can be converted into three subproblems via ADMM [15] algorithm:

$$\begin{cases} \mathbf{h}_{i+1}^d = \operatorname{argmin}_{\mathbf{h}^d} \frac{\lambda}{2}\left\|\mathbf{h}^d\right\|_2^2 \\ \quad + \frac{\mu}{2}\left\|\hat{\mathbf{v}} - \sqrt{\mathrm{N}}\mathbf{F}\mathbf{B}^{\mathsf{T}}\mathbf{h}^d\right\|_2^2 \\ \quad (\hat{\mathbf{v}}^d - \sqrt{\mathrm{N}}\mathbf{F}\mathbf{B}^{\mathsf{T}}\mathbf{h}^d)^{\mathsf{T}}\hat{\boldsymbol{\zeta}}^d \\ \hat{\mathbf{v}}_{i+1} = \operatorname{argmin}_{\mathbf{v}} \frac{1}{2\mathrm{N}}\left\|\sum_{d=1}^{\mathrm{D}} \hat{\mathbf{x}}^d \odot \hat{\mathbf{v}}^d - \hat{\mathbf{y}}\right\|_2^2 \\ \quad + \frac{\gamma}{2\mathrm{N}}\left\|\sum_{d=1}^{\mathrm{D}} \hat{\mathbf{x}}_\delta^d \odot \hat{\mathbf{v}}^d\right\|_2^2 \\ \quad + \frac{\mu}{2}\sum_{d=1}^{\mathrm{D}}\left\|\hat{\mathbf{v}}^d - \sqrt{\mathrm{N}}\mathbf{F}\mathbf{B}^{\mathsf{T}}\mathbf{h}^d\right\|_2^2 \\ \quad + \sum_{d=1}^{\mathrm{D}}(\hat{\mathbf{v}}^d - \sqrt{\mathrm{N}}\mathbf{F}\mathbf{B}^{\mathsf{T}}\mathbf{h}^d)^{\mathsf{T}}\hat{\boldsymbol{\zeta}}^d \\ \hat{\boldsymbol{\zeta}}_{i+1} = \hat{\boldsymbol{\zeta}}_i + \mu(\hat{\mathbf{v}}_{i+1} - \hat{\mathbf{h}}_{i+1}), \end{cases} \tag{4}$$

where i stands for the iterations.

Subproblem $\mathbf{h}^d$: Given $\mathbf{v}$ and $\boldsymbol{\zeta}$, the optimal $\mathbf{h}$ can be obtained as:

$$\mathbf{h}^d = \frac{\boldsymbol{\zeta}^d + \mu\mathbf{v}^d}{\lambda/\mathrm{N} + \mu} . \tag{5}$$

Subproblem $\hat{\mathbf{v}}$: If other variables $\mathbf{h}$ and $\boldsymbol{\zeta}$ are available, the solution of subproblem $\hat{\mathbf{v}}$ can be optimized by decomposing the third problem in (4) into N smaller problems:

$$\begin{aligned}\hat{\mathbf{v}}(n)^* = {} & \frac{1}{2\mathrm{N}}\left\|\hat{\mathbf{x}}(n)^{\mathsf{T}}\hat{\mathbf{v}}(n) - \hat{\mathbf{y}}(n)\right\|_2^2 + \frac{\gamma}{2\mathrm{N}}\left\|\hat{\mathbf{x}}_\delta(n)^{\mathsf{T}}\hat{\mathbf{v}}(n)\right\|_2^2 \\ & + (\hat{\mathbf{v}}(n) - \hat{\mathbf{h}}(n))^{\mathsf{T}}\hat{\boldsymbol{\zeta}}(n) + \frac{\mu}{2}\left\|\hat{\mathbf{v}}(n) - \hat{\mathbf{h}}(n)\right\|_2^2 .\end{aligned} \tag{6}$$

Algorithm 1: EMCF Tracker.

Input: Continuous video sequences and ground truth in the first frame.
Output: Esitimated target position in $t > 1$ frame.

```
for frame t = 1 to end do
    if t > 1 then
        Extract the feature z of the search region.
        Generate the response map using Eq. 10.
        Search the peak of response and return the position.
    end
    Update the target appearance model using Eq. 9.
    Compute the environment residual x_δ.
    Learn the filter v̂ with the residual information using Eq. 7.
end
```

For clarity of description, we define $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}_\delta$ as $\hat{\mathbf{x}}_0$ and $\hat{\mathbf{x}}_1$, respectively. Then, the closed-form solution of $\hat{\mathbf{v}}$ can be accelerated through Sherman-Morrison [16] fomula:

$$\hat{\mathbf{v}}(n)^* = \frac{1}{\mu \mathrm{N}}\left(\hat{\mathbf{x}}(n)\hat{\mathbf{y}}(n) - \mathrm{N}\hat{\boldsymbol{\zeta}}_f + \mu\mathrm{N}\hat{\mathbf{h}}(n)\right) - \frac{\sum_{k=0}^{1} p_k \hat{\mathbf{x}}_k(n)}{\mu\theta}$$
$$\left(\frac{1}{\mathrm{N}}\eta\hat{\mathbf{y}}(n) - \sum_{k=0}^{1}\hat{\mathbf{x}}_k(n)^\mathsf{T}\hat{\boldsymbol{\zeta}}(n) + \mu\sum_{k=0}^{1}\hat{\mathbf{x}}_k(n)^\mathsf{T}\hat{\mathbf{h}}(n)\right), \tag{7}$$

where $p_0 = 1$, $p_1 = \gamma$, $\theta = \mu\mathrm{N} + \sum_{k=0}^{1} p_k \hat{\mathbf{x}}_k(n)^\mathsf{T}\hat{\mathbf{x}}_k(n)$, and $\eta = \sum_{k=0}^{1} p_k \hat{\mathbf{x}}_k(n)^\mathsf{T}\hat{\mathbf{x}}(n)$.

B. Fast Environment Awareness

We design a simple method to acquire environmental information without bells and whistles. As shown in the training phase of Fig. 1, the environmental information can be directly obtained through the Hadamard product between the environment-aware mask $\mathbf{m}_e$ and the feature of the current training sample $\mathbf{x}$, which can be formulated as:

$$\mathbf{x}_c = \mathbf{x} \odot \mathbf{m}_e. \tag{8}$$

The process of the mask generation is described as follows: Firstly, an all-one matrix is generated with the same size as the training sample. Then, the area corresponding to the target (red rectangle) is set to 0. Finally, the generated matrix is resized to obtain an environment-aware mask $\mathbf{m}_e$ (same size as $\mathbf{x}$).

As a result, this environment awareness strategy avoids additional feature extraction of the environment patch as well as irrelevant background information.

C. Model Update

To raise the robustness, the update of the target appearance model is defined as follows:

$$\hat{\mathbf{x}}^M = (1-\beta)\hat{\mathbf{x}}^M_{t-1} + \beta\hat{\mathbf{x}}, \tag{9}$$

where $\hat{\mathbf{x}}^M$ and $\hat{\mathbf{x}}^M_{t-1}$ denote the model in the current and last frames, respectively. $\hat{\mathbf{x}}$ is the training sample of the current frame. β represents the online learning rate.

TABLE I
AVERAGE PERFORMANCE COMPARISON WITH HANDCRAFTED-BASED TRACKERS ON TWO BENCHMARKS. RED, GREEN, AND BLUE REPRESENT THE TOP THREE TRACKERS IN TERMS OF DP, AUC, AND FPS, RESPECTIVELY

Tracker	Venue	Prec.	Succ.	FPS	Tracker	Venue	Prec.	Succ.	FPS
EMCF	-	0.690	0.484	52.805	ECO_HC [19]	'17CVPR	0.648	0.461	62.163
AutoTrack [4]	'20CVPR	0.694	0.478	50.263	CSR-DCF [22]	'17CVPR	0.645	0.444	11.535
ARCF [3]	'19ICCV	0.680	0.472	23.930	KCC [20]	'18AAAI	0.485	0.333	40.730
STRCF [18]	'18CVPR	0.635	0.447	26.630	BACF [14]	'17ICCV	0.581	0.407	49.515
MCCT_H [21]	'18CVPR	0.600	0.419	58.125	staple_CA [10]	'17CVPR	0.545	0.386	56.255
SAMF_CA [10]	'17CVPR	0.528	0.355	10.170	SRDCF [23]	'15ICCV	0.543	0.393	12.310

The superscript * means GPU speed.

D. Tracking

As shown in the detection phase of Fig. 1, when a new frame is upcoming, we acquire the search region $\mathbf{z}$ with the motion-aware search method [6]. Then, the response map $\mathbf{M}_{t+1}$ for localization is produced according to the following equation:

$$\mathbf{M}_{t+1} = \mathcal{F}^{-1}\left(\sum_{d=1}^{\mathrm{D}} \hat{\mathbf{z}}^d_{t+1} \odot \hat{\mathbf{v}}^d_t\right), \tag{10}$$

where $\mathcal{F}^{-1}$ denotes the inverse DFT operator.

The detailed tracking procedure of our tracker is dispalyed in Algorithm 1.

III. EXPERIMENTS

A. Implementation

Our tracker uses Hog, CN, and Grayscale for feature representation. The learning rate and the sensitivity factor are empirically chosen as $\beta = 0.0199$ and $\gamma = 0.2$, respectively. We use two ADMM iterations to train the filter quickly. The penalty factor is updated following $\mu_{i+1} = \min(\mu_{\max}, \rho\mu_{\mathrm{i}})$, where $\mu_{max} = 10^4$, $\mu_0 = 1$, and $\rho = 10$. Our tracker runs on a PC with an i7-9750H CPU (2.60 GHz), 32 GB RAM, and a single RTX 2060 GPU. The matlab project of our tracker is available on https://github.com/FreeZhang96/EMCF.

For all benchmarks, we evaluate the performance of all trackers based on the one pass evaluation (OPE) [17]. Two metrics, *i.e.*, area under the curve (AUC) and distance precision (DP), are employed for ranking all trackers. Besides, frame per second (FPS) is for measuring the tracking speed.

B. Comparison With Handcrafted-Based Trackers

1) Quantitative Evaluation: We exhibit quantitative evaluation as well as qualitative evaluation on two well-known UAV benchmarks, *i.e.*, UAV123_10fps [12] and DTB70 [13].

Overall Analysis. We compare the proposed tracker with other 11 SOTA handcrafted-based trackers on all two well-known UAV benchmarks. These trackers include STRCF [18], ECO_HC [19], KCC [20], SAMF_CA [10], ARCF [3], AutoTrack [4], MCCT_H [21], CSR-DCF [22], SRDCF [23], BACF [14], and Staple_CA [10]. As shown in Fig. 2, the proposed tracker achieves the best AUC scores on all benchmarks. In terms of DP score, our tracker ranks first and second on UAV123_10fps [12] and DTB70 [13] benchmarks, respectively. Besides, Table I shows the average performance of all trackers on two benchmarks. The proposed tracker has comparable performance with the recent SOTA tracker, *i.e.*, AutoTrack [4] and ARCF [3], while having more faster tracking speed.

Attribute-based Analysis. Fig. 3 provides success plots under different attributes (viewpoint change, low resolution, fast camera motion, and background clutter) from two benchmarks. In terms of these attributes, the scores of EMCF exceed that of other algorithms by a large margin. In particular, our tracker surpasses the second-best 6.6% under background clutter.

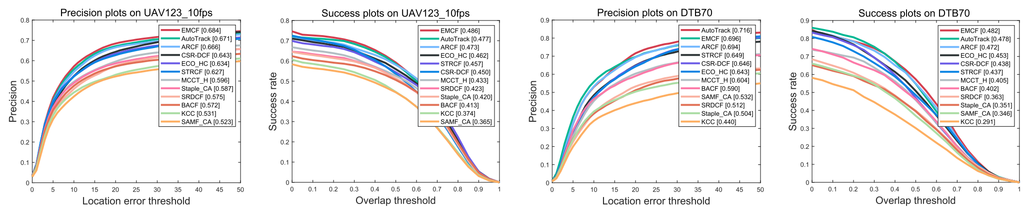

Fig. 2. Overall performance of the proposed tracker and other 11 SOTA handcrafted feature-based trackers on UAV123_10fps [12] and DTB70 [13].

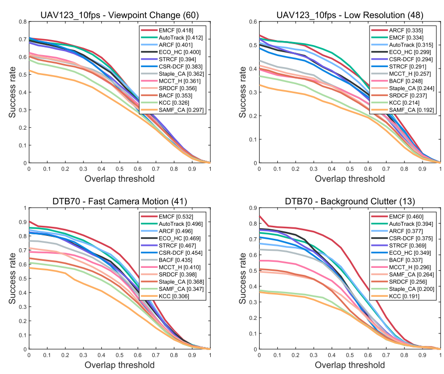

Fig. 3. Attribute-based analysis of the proposed tracker and other 11 SOTA handcrafted feature-based trackers on UAV123_10fps [12] and DTB70 [13].

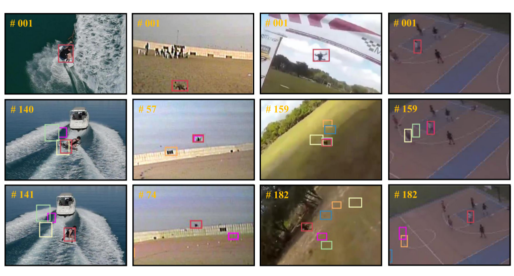

Fig. 4. Qualitative evaluation between the proposed tracker and other 5 SOTA trackers.

2) *Qualitative Evaluation:* We visualize the results of our tracker and other 5 top-ranked handcrafted-based trackers on four challenging sequences. As shown in Fig. 4, from left to right are wakeboard4 and uav3 from UAV123_10fps [12], ChasingDrones and StreetBasketball1 from DTB70 [13]. Disturbed by the rapidly changing environment, other trackers fail to track while our tracker is not sensitive to environmental change. Our tracker achieves robust tracking, which can be attributed to the use of environmental mutation-insensitive.

C. Comparison With Deep-Based Trackers

To comprehensively evaluate the proposed tracker, eleven state-of-the-art trackers with deep features, *i.e.*, DaSiamRPN [24], SiamRPN++ [25], ASRCF [26], UDT [27], MCCT [21], TADT [28], fECO [29], DSTRCF [18], UDT+ [27], fDSTRCF [29], and KAOT [7], are selected for comparison on DTB70 [13] benchmark. As is shown in Table II, while these deep-based trackers rely on a high-end GPU, our tracker runs on a CPU with a speed of 51.412 FPS. Meawhile, both AUC and DP scores of the proposed tracker rank fourth on DTB70 benchmark. The efficiency of our tracker confirms that it is sufficient for real-time UAV platforms.

TABLE II
COMPARISON WITH DEEP-BASED TRACKERS ON DTB70 [13] BENCHMARK. RED, GREEN, AND BLUE REPRESENT THE TOP THREE TRACKERS IN TERMS OF DP, AUC, AND FPS, RESPECTIVELY

Tracker	Venue	Prec.	Succ.	FPS	Tracker	Venue	Prec.	Succ.	FPS
EMCF	-	0.696	0.482	**51.412**	KAOT [7]	'21TMM	0.692	0.469	*14.045
fDSTRCF [29]	'20TIP	0.667	0.458	*14.800	fECO [29]	'20TIP	0.668	0.454	*21.085
TADT [28]	'19CVPR	0.693	0.464	*35.314	DSTRCF [18]	'18CVPR	**0.734**	**0.506**	*5.816
UDT+ [27]	'19CVPR	0.658	0.462	*40.135	UDT [27]	'19CVPR	0.602	0.422	***55.621**
MCCT [21]	'18CVPR	**0.725**	**0.484**	*8.622	ASRCF [19]	'19CVPR	0.695	0.469	*22.158
DaSiamRPN [24]	'18ECCV	0.694	0.472	*8.169	SiamRPN++ [25]	'19CVPR	**0.795**	**0.589**	***45.652**

The superscript * means GPU speed.

TABLE III
ABLATION ANALYSIS OF THE PROPOSED TRACKER

Tracker	Module			UAV123_10fps		DTB70		Average	
	MA	EL	EM	Succ.	Prec.	Succ.	Prec.	Succ.	Prec.
Baseline				0.468	0.654	0.473	0.685	0.471	0.670
Baseline+MA	✓			0.477	0.668	0.472	0.686	0.474	0.677
Baseline+MA+EL	✓	✓		0.478	0.680	0.473	0.678	0.476	0.679
EMCF	✓		✓	0.486	0.684	0.482	0.696	0.484	0.690

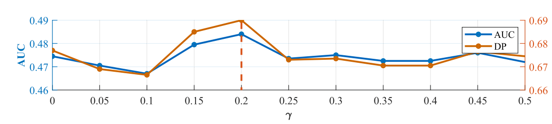

Fig. 5. Analysis of key parameter γ on UAV123_10fps [12] and DTB70 [13].

D. Ablation Study and Parameter Analysis

We conducted ablation analysis and key parameter analysis of the proposed tracker on UAV123_10fps [12] and DTB70 [13] benchmarks.

Table III shows the detailed and average AUC and DP. We take the BACF tracker using Hog, CN, and Grayscale features as the baseline. When the baseline tracker with the motion-aware (MA) and environmental mutation-insensitivity (EM) modules, average AUC and DP scores are improved to 0.484 and 0.690, exceeding the baseline tracker by 1.3% and 2.0%, respectively. Compared with environmental learning (EL) based on the current environment ($\mathbf{x}_{c,t-1} = 0$), EM module can improve the tracking performance more.

Fig. 5 provides the average AUC and DP when γ varies from 0 to 0.5 with a step of 0.05. When $\gamma = 0.2$, DP and AUC reach the maximum. Therefore, $\gamma = 0.2$ in this work is reasonable.

IV. CONCLUSION

In this letter, we propose a novel environmental mutation-insensitive tracking method based on DCF, aiming at improving the discriminative power of the tracker under drastic environmental change. By learning the environment residual, the proposed tracker can keep robust to irregular environment change. In addition, an efficient environment-aware strategy is designed to obtain the environmental information. Experimental results on two challenging benchmarks demonstrate the superiority of the proposed tracker in accuracy and running speed, which is sufficient for real-time UAV applications.

REFERENCES

[1] C. Fu, B. Li, F. Ding, F. Lin, and G. Lu, "Correlation filter for UAV-based aerial tracking: A review and experimental evaluation," *IEEE Geosci. Remote Sens. Mag.*, early access, doi: 10.1109/MGRS.2021.3072992.

[2] C. Deng, S. He, Y. Han, and B. Zhao, "Learning dynamic spatial-temporal regularization for UAV object tracking," *IEEE Signal Process. Lett.*, vol. 28, pp. 1230–1234, 2021, doi: 10.1109/LSP.2021.3086675.

[3] Z. Huang, C. Fu, Y. Li, F. Lin, and P. Lu, "Learning aberrance repressed correlation filters for real-time UAV tracking," in *Proc. IEEE Int. Conf. Comput. Vision*, 2019, pp. 2891–2900.

[4] Y. Li, C. Fu, F. Ding, Z. Huang, and G. Lu, "Autotrack: Towards high-performance visual tracking for UAV with automatic spatio-temporal regularization," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2020, pp. 11923–11932.

[5] C. Fu, W. Jiang, F. Lin, and Y. Yue, "Surrounding-aware correlation filter for UAV tracking with selective spatial regularization," *Signal Process.*, vol. 167, 2020, Art. no. 107324.

[6] C. Fu, F. Ding, Y. Li, J. Jin, and C. Feng, "Learning dynamic regression with automatic distractor repression for real-time UAV tracking," *Eng. Appl. Artif. Intell.*, vol. 98, 2021, Art. no. 104116.

[7] Y. Li, C. Fu, Z. Huang, Y. Zhang, and J. Pan, "Intermittent contextual learning for keyfilter-aware UAV object tracking using deep convolutional feature," *IEEE Trans. Multimedia*, vol. 23, pp. 810–822, 2021, doi: 10.1109/TMM.2020.2990064.

[8] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2010, pp. 2544–2550.

[9] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 583–596, Mar. 2015.

[10] M. Mueller, N. Smith, and B. Ghanem, "Context-aware correlation filter tracking," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 1396–1404.

[11] Y. Yan, X. Guo, J. Tang, C. Li, and X. Wang, "Learning spatio-temporal correlation filter for visual tracking," *Neurocomputing*, vol. 436, pp. 273–282, 2021.

[12] M. Mueller, N. Smith, and B. Ghanem, "A benchmark and simulator for UAV tracking," in *Proc. Eur. Conf. Comput. Vision*, 2016, pp. 445–461.

[13] S. Li and D.-Y. Yeung, "Visual object tracking for unmanned aerial vehicles: A benchmark and new motion models," in *Proc. AAAI Conf. Artif. Intell*, vol. 31, no. 1, 2017, pp. 4179–4186.

[14] H. Kiani Galoogahi, A. Fagg, and S. Lucey, "Learning background-aware correlation filters for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vision*, 2017, pp. 1135–1143.

[15] S. Boyd, N. Parikh, and E. Chu, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, pp. 1–122, 2011.

[16] J. Sherman and W. J. Morrison, "Adjustment of an inverse matrix corresponding to a change in one element of a given matrix," *Ann. Math. Statist.*, vol. 21, no. 1, pp. 124–127, 1950.

[17] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2013, pp. 2411–2418.

[18] F. Li, C. Tian, W. Zuo, L. Zhang, and M. H. Yang, "Learning spatial-temporal regularized correlation filters for visual tracking," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 4904–4913.

[19] M. Danelljan, G. Bhat, F. Shahbaz Khan, and M. Felsberg, "ECO: Efficient convolution operators for tracking," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 6638–6646.

[20] C. Wang, L. Zhang, L. Xie, and J. Yuan, "Kernel cross-correlator," in *Proc. AAAI Conf. Artif. Intell.*, 2017, pp. 4179–4186.

[21] N. Wang, W. Zhou, Q. Tian, R. Hong, M. Wang, and H. Li, "Multi-cue correlation filters for robust visual tracking," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 4844–4853.

[22] A. Lukezic, T. Vojir, L. ˇCehovin Zajc, J. Matas, and M. Kristan, "Discriminative correlation filter with channel and spatial reliability," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 6309–6318.

[23] M. Danelljan, G. Hager, F. Shahbaz Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vision*, 2015, pp. 4310–4318.

[24] Z. Zhu, Q. Wang, B. Li, W. Wu, J. Yan, and W. Hu, "Distractor-aware siamese networks for visual object tracking," in *Proc. Eur. Conf. Comput. Vision*, 2018, pp. 101–117.

[25] B. Li, W. Wu, Q. Wang, F. Zhang, J. Xing, and J. Yan, "Siamrpn : Evolution of siamese visual tracking with very deep networks," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, 2019, pp. 4282–4291.

[26] K. Dai, D. Wang, H. Lu, C. Sun, and J. Li, "Visual tracking via adaptive spatially-regularized correlation filters," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2019, pp. 4665–4674.

[27] N. Wang, Y. Song, C. Ma, W. Zhou, W. Liu, and H. Li, "Unsupervised deep tracking," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2019, pp. 1308–1317.

[28] X. Li, C. Ma, B. Wu, Z. He, and M.-H. Yang, "Target-aware deep tracking," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2019, pp. 1369–1378.

[29] N. Wang, W. Zhou, Y. Song, C. Ma, and H. Li, "Real-time correlation tracking via joint model compression and transfer," *IEEE Trans. Image Process.*, vol. 29, pp. 6123–6135, 2020, doi: 10.1109/TIP.2020.2989544.