Disruptor-Aware Interval-Based Response Inconsistency for Correlation Filters in Real-Time Aerial Tracking

Changhong Fu, *Member, IEEE*, Junjie Ye, Juntao Xu, Yujie He, and Fuling Lin, *Graduate Student Member, IEEE*

***Abstract*— Aerial object tracking approaches based on discriminative correlation filter (DCF) have attracted wide attention in the tracking community due to their impressive progress recently. Many studies introduce temporal regularization into the DCF-based framework to achieve a more robust appearance model and further enhance the tracking performance. However, existing temporal regularization approaches usually utilize the information of two consecutive frames, which are not robust enough due to limited information. Although some methods attempt to incorporate abundant training samples and generally improve the tracking performance, these improvements are at the expense of significantly increased computing consumption. Besides, most existing methods introduce historical information directly without denoising, which means that background noises are also introduced into the filter training and may degrade the tracking accuracy. To tackle the drawbacks mentioned earlier, this work proposes a novel aerial object tracking approach to exploit disruptor-aware interval-based response inconsistency, i.e., IBRI tracker. The proposed method is able to incorporate historical interval information by utilizing responses in the filter training process, thereby obtaining a robust tracking performance while maintaining the real-time speed. Moreover, to reduce the disruptions caused by similar object, partial occlusion, and other challenging scenes, a novel disruptor-aware scheme based on response bucketing is introduced to detect the disruptor and enforce a spatial penalty for the disruptive area around the tracked object. Exhausted experiments on multiple well-known challenging aerial tracking benchmarks demonstrate the accuracy and robustness of the proposed IBRI tracker against other 35 state-of-the-art trackers. With a real-time speed of ∼32 frames/s on a single CPU, the proposed approach can be applied for typical aerial platforms to achieve aerial visual object tracking efficiently.**

***Index Terms*— Aerial object tracking, discriminative correlation filter (DCF), disruptor-aware bucketing, historical frame information, interval-based response inconsistency, temporal regularization.**

I. Introduction

AS AN essential research branch of remote sensing, aerial visual object tracking has broad applications, e.g., traffic monitoring [1], [2], military surveillance [3], [4], and motion analysis [5], [6]. Although various tracking approaches have been designed for aerial tracking, it is still a challenging task due to many visual uncertainties, e.g., fast object/aerial platform motion and partial/full occlusion. Besides, the stringent requirement of tracking efficiency further aggravates the difficulty for object tracking on the typical aerial platforms.

Some long-term tracking approaches [7]–[10] have been proposed recently and achieve competitive performance. In [7], [8], and [10], novel redetection mechanisms are presented to relocate the object. Motivated by the two-stage object detector Faster-RCNN [11], Huang *et al.* [9] proposed GlobalTrack. However, most of these trackers rely on the deployment of high-performance GPUs [7], [9], [10] or lack of real-time capabilities [7]–[9], which is difficult to deploy on aerial platforms with limited computing power. Therefore, this work is dedicated to completing high-performance tracking tasks on a single CPU.

Due to the remarkable efficiency and accuracy, the discriminative correlation filter (DCF)-based tracking approaches have attracted widespread interests in recent years [13]–[16]. The DCF-based methods learn a filter frame by frame. Specifically, given the location of the object in the previous frame, the training patch centered on this position is extracted. After that, the circular correlation with dense sampling technique is applied to the patch to train the filter with circularly shifted samples. Utilizing the fast Fourier transformation (FFT), the filter and training samples are transferred from spatial domain to frequency domain. Thus, the time-consuming correlation operation can be replaced by effective element-wise matrix multiplication, which significantly reduces the computational complexity and boosts the efficiency of the DCF-based methods. However, the circular shift operation results in some unreal training samples at the extension boundaries, i.e., boundary effects. These corrupt samples contaminated by the unwarranted boundary effects used in the training step would degrade the overall performance of the tracker. Moreover, traditional DCF-based approaches are easy to drift in some challenging scenarios, e.g., similar object, fast object/aerial platform motion, and partial/full occlusion.

To tackle the defects mentioned earlier, Huang *et al.* [12] proposed the aberrance repressed correlation filters, i.e., ARCF. Drawing lessons from BACF [14], ARCF uses a binary cropping matrix to alleviate the boundary effects by expanding the search region, thereby training the filter with real negative samples. Specifically, the temporal regularization

Manuscript received May 9, 2020; revised July 18, 2020 and August 21, 2020; accepted October 7, 2020. Date of publication October 26, 2020; date of current version July 22, 2021. This work was supported by the National Natural Science Foundation of China under Grant 61806148. *(Corresponding author: Changhong Fu.)*

The authors are with the School of Mechanical Engineering, Tongji University, Shanghai 201804, China (e-mail: changhongfu@tongji.edu.cn).

Digital Object Identifier 10.1109/TGRS.2020.3030265

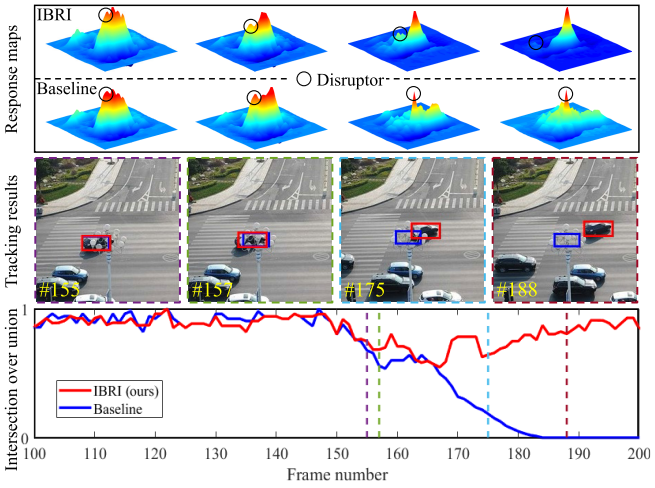

Fig. 1. Comparisons of the proposed IBRI (in red) with its baseline ARCF [12] (in blue) in terms of response maps, tracking results, and intersection over union (IoU) curves on the *S*1606 sequence from the UAVDT benchmark. The main challenge of this sequence is large occlusion. To evaluate the tracking quality of the two trackers during the challenge, their IoU curves between the estimated positions of the trackers and the ground truth during frame 100 to 200 are drawn, respectively. By means of suppressing the disruptor-aware interval-based response inconsistency, IBRI can detect and suppress the disruptors on responses caused by occlusion and precisely relocates the object when it reappears from occlusion. With only two frames of information and no measure to handle the disruptors, ARCF suffers from occlusion and is misled by the disruptors.

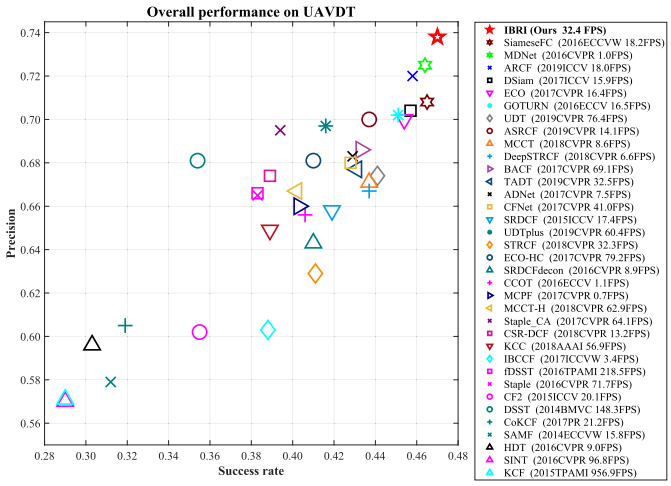

Fig. 2. Overall performance of the proposed IBRI tracker and other 35 state-of-the-art trackers on the UAVDT benchmark. The result shows that the IBRI tracker outperforms other state-of-the-art trackers in terms of success rate and precision with a favorable real-time speed. Note that the running speed in FPS is also presented in the legend of the trackers.

term in ARCF is used to restrict the difference of response maps between two consecutive frames. As a result, the model drift is avoided to some extent. Nevertheless, it only introduces the information of two frames into its temporal regularization, which is not robust enough. Since significant anomalies caused by abrupt viewpoint change, fast motion, and other factors often appear in aerial tracking, there can be drastic changes between two consecutive frames in these cases. The information with great changes is incorporated into the training phase of ARCF and may degrade the filter. In addition, the response maps it introduced may contain disruptors caused by aerial tracking-specific challenges, e.g., partial occlusion and similar object. No measures are taken to handle these interferences. The background noises in the temporal regularizer can thus mislead the learned correlation filter.

Previous studies have proposed some methods to score the quality of the response maps, e.g., APCE in LMCF [17], Kurtosis in SAT [18], and PSR in SRDCFdecon [19]. In the visual odometry application, Kitt *et al.* [20] proposed a bucketing mechanism to achieve a preferable estimation of the overall egomotion of the vehicle and reduce the drift rates. Motivated by this, an original response bucketing scheme is proposed in this work to denoise the response maps. This scheme can not only measure the quality of the response maps but also detect and suppress disruptors in the response maps precisely when the quality is poor. Furthermore, a novel DCF-based tracker utilizing the disruptor-aware interval-based response inconsistency (IBRI) is proposed. Fig. 1 shows a visualized comparison between IBRI and our baseline ARCF on a typical challenging sequence. Attributing to the novel disruptor-aware temporal regularization, the proposed method can handle object appearance variations more efficiently. Fig. 2 compares the overall tracking performance of the proposed approach against other well-known state-of-the-art trackers on the UAVDT benchmark. The competitive performance of the IBRI tracker verifies its suitableness for real-time robust aerial tracking.

The main contributions of this work can be summarized as follows.

1) An innovative temporal regularization based on historical interval response inconsistency is proposed. Integrating the responses from multiple historical frames, the temporal regularization enables the filter to be trained with interval information. Attributing to this strategy, the proposed IBRI tracker can better perceive the change of the object appearance and achieve competitive robustness in the face of object appearance variations.
2) To suppress the background disruptors, a novel disruptor-aware mechanism based on response bucketing is proposed to detect the disruptive areas on the response and generate an adaptive penalty mask to enforce the spatial penalty for the detected disruptive areas. As a result, misleading background noises can be restrained.
3) Exhausted evaluations are performed on multiple well-known challenging aerial object tracking benchmarks, i.e., UAV123@10fps [21], VisDrone2018-SOT [22], and UAVDT [23]. The experiments demonstrate that the proposed method is superior to other 35 state-of-the-art trackers in terms of accuracy and robustness.

The rest of this article is organized as follows. Section II briefly reviews the previous literature that is most relevant to this work. Details of the proposed IBRI tracker are described in Section III. Section IV shows the comprehensive results of extensive experiments and exhausted evaluations of the proposed approach. Finally, conclusions are drawn in Section V.

II. Related Work

In recent years, many profound studies in the field of aerial object tracking have made impressive achievements. The

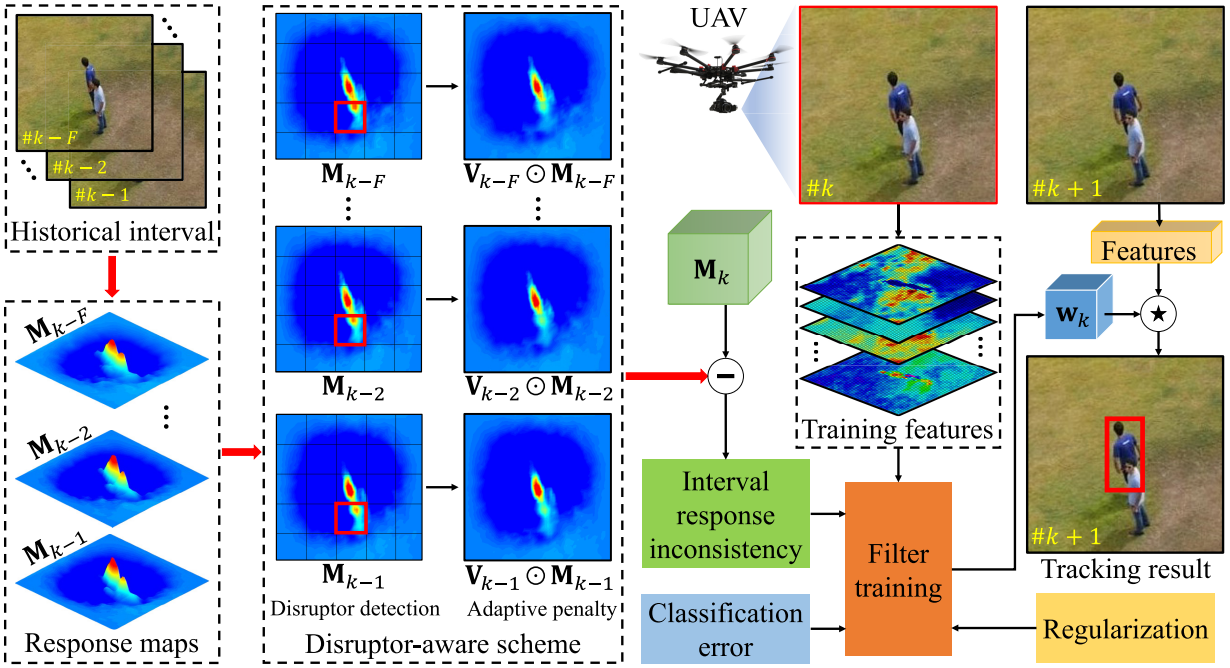

Fig. 3. Tracking procedure of the proposed IBRI tracker in the kth frame. Historical interval responses are incorporated into the filter training phase after denoising by a novel disruptor-aware scheme based on response bucketing.

following three sections revisit some representative research in three aspects that are tightly related to the proposed IBRI tracker, respectively.

A. Tracking With DCF

Due to the low computational complexity, DCF-based tracking algorithms have drawn wide attention in the field of visual tracking. Starting from the minimum output sum of squared error (MOSSE) filter proposed by Bolme *et al.* [24], the DCF-based approaches have made great progress. Henriques *et al.* [25] introduced the cyclic matrix and the kernel trick into the DCF framework and generally improved the tracking performance. Danelljan *et al.* [26] proposed the discriminative scale-space tracker (DSST), which enabled the scale estimation capability of the tracker by training another single dimension filter to detect the scale. Later, Li *et al.* [27] proposed another scale adaptive tracker (SAMF). Integrating the histogram of gradient (HOG) [28] and the color names (CN) [29] features, SAMF achieved more robust tracking performance. Mueller *et al.* [30] introduced context information into the training stage of the correlation filter to tackle background noises and further improved the performance. In addition, deep features have been introduced to visual tracking applications [31]–[35] and achieved remarkable improvement. However, these methods train correlation filters only with information in the current frame, which is not reliable and discriminative enough when object appearance drastic variations occur, e.g., partial/full occlusion and fast object/aerial platform motion.

B. Tracking With Temporal Information

Some previous research attempts to enhance the robustness and discriminability of the tracker by utilizing the information from the previous frame [12], [16], [36]–[38]. In [16] and [36], the correlation filter from the last frame is introduced into the training process to prevent the learned filter in the current frame from a sudden variation. ASRCF [37] considered temporal consistency during filter training by introducing an adaptive spatial regularizer to avoid abrupt appearance variation. Proposed by Huang *et al.* [12], ARCF incorporated temporal cue into the tracking system by utilizing the response from the previous frame. As a result, aberrance reflected in the last response map can be repressed. Since the quality of the response map directly determines the accuracy and successfulness of tracking, by regularizing the change in response maps, ARCF further improved the tracking performance. Although these methods attempt to embed the temporal information into the DCF-based framework, they usually merely consider the information from only the previous frame and cannot exploit the vital information from historical interval frames. In case of sudden changes in the object appearance, e.g., abrupt viewpoint change and fast object/aerial platform motion, information from only two frames is too limited to contribute to precise tracking. These methods are prone to drift in these scenes. Besides, background noises are also involved in the temporal information, which means that the variations of background play the same role as that of the object in these methods. As a result, the filter training may be misled by the disruptors in background information, especially in challenging scenes such as occlusion and similar object.

C. Tracking With Historical Interval Information

To obtain a more discriminative and comprehensive appearance model, some approaches attempt to integrate sufficient historical information [15], [19], [34], [39]–[41]. Wang *et al.* [40] proposed to introduce historical correlation filters to gain an accurate spatiotemporal model. In [15]

and [34], filters are trained with hundreds of historical training samples, which improves the tracking performance at the price of increased calculation complexity. To mitigate the interference caused by corrupt historical samples, Danelljan *et al.* [19] further proposed the SRDCFdecon tracker, which is able to manage the training set dynamically by means of reweighting the samples. However, as the number of training samples increases, the filter training phase can still be time-consuming, which is not suitable for real-time aerial tracking. In [41], for the purpose of speeding up the training step, a Gaussian mixture model is employed to simplify the training set. Although this strategy well balances the speed and performance, there is no measure to detect and suppress the disruptors in the historical information, which limits the further improvement of the discriminative power of the tracker. In addition, the historical information introduced in [41] are training samples, while response maps can better reflect the appearance changes.

Different from the approaches mentioned earlier, the proposed IBRI tracker integrates historical interval information more appropriately. Multiple consecutive response maps are chosen as the carrier of this information, for the reason that they can intuitively reflect the change of the object appearance and only occupy little memories compared with training samples. Furthermore, a novel disruptor-aware scheme based on response bucketing is applied to detect and restrain the misleading background noises in the responses. Attributing to the suppression of the disruptor-aware interval-based response inconsistency, the proposed IBRI tracker obtains a more favorable performance while maintaining a real-time speed.

III. Proposed Method

In this section, a new tracking pipeline exploiting disruptor-aware interval-based response inconsistency (IBRI) is proposed. Fig. 3 shows the workflow of the IBRI tracker. It is noted that the main symbols in this work are presented in Table I.

In the DCF-based tracking pipeline, when a new frame arrives, a tracker will use the learned filter to perform a correlation operation within the search region in the new frame to generate a detection response map. The peak of the response is considered to be the estimated location of the object in this new frame. Therefore, object appearance variations are first and directly reflected on response maps. Inspired by this fact, we utilize responses when designing a temporal regularization.

A. Disruptor-Aware Scheme Based on Response Bucketing

Some misleading background noises may be contained in the responses, in order to detect and suppress these disruptors, a novel disruptor-aware scheme based on response bucketing is presented.

As shown in Fig. 4, when a response map $\mathbf{M}_{k-f}$ is generated in the $(k-f)$th frame, it will be divided into $\alpha \times \alpha$ nonoverlapping rectangles by means of bucketing. By detecting disruptors in each response bucket and setting penalty coefficients, an adaptive penalty mask $\mathbf{V}_{k-f}$ will be generated. In the disruptor detection, the ratio $\theta_{(i,j)}$ between

TABLE I

Dimension and Meanings of the Main Symbols in This Work

Symbol	Meaning
$D \in \mathbb{N}$	Total number of feature channels
$N \in \mathbb{N}$	Dimension of response maps/features
$F \in \mathbb{N}$	Historical interval length
$\mathbf{M}_k \in \mathbb{R}^N$	Response map in the k-th frame
$\alpha \in \mathbb{N}$	Bucketing number $\alpha \times \alpha$
$\mathbf{V}_k \in \mathbb{R}^N$	Adaptive penalty mask for the k-th response map
$m_{max} \in \mathbb{R}$	Global maximum value of a response map
$\kappa \in \mathbb{R}$	Penalty factor in $\mathbf{V}_k$
$m_{(i,j)} \in \mathbb{R}$	Local maximum value of the (i,j)-th bucket
$\theta_{(i,j)} \in \mathbb{R}$	Ratio of $m_{(i,j)}$ to m_{max}
$\pi \in \mathbb{R}$	A threshold to judge disruptors
$\mathbf{y} \in \mathbb{R}^N$	Desired regression label
$\mathbf{x}^d \in \mathbb{R}^N$	Training patch of the d-th channel
$\mathbf{w}^d \in \mathbb{R}^N$	Correlation filter of the d-th channel
$\mathbf{P} \in \mathbb{R}^{N \times N}$	Cropping matrix
$\mathbf{F} \in \mathbb{C}^{N \times N}$	DFT matrix
$\lambda \in \mathbb{R}$	Regularization parameter
$\gamma_1 \in \mathbb{R}$	The base value of frame wheight
$[\phi_{p_f,q_f}]$	Shift operator

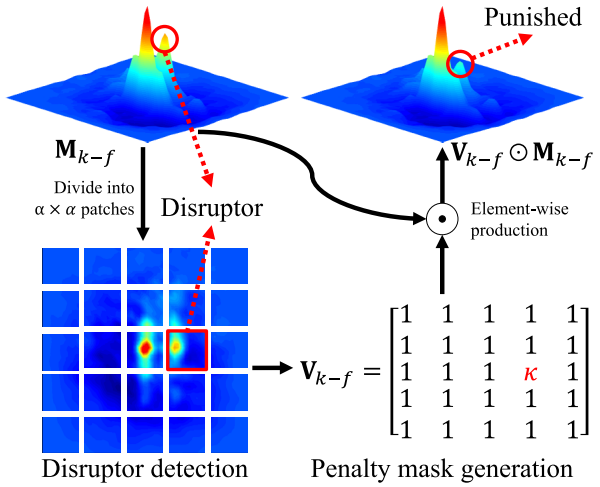

Fig. 4. Workflow of the disruptor-aware scheme based on response bucketing.

the local maximum value $m_{(i,j)}$ in the ith row and the jth column response bucket and the global maximum value m_{max} on the entire response map is calculated as

$$\theta_{(i,j)} = \frac{m_{(i,j)}}{m_{\max}}. \tag{1}$$

Thus, the penalty factor κ is set by the following:

$$\begin{cases} \kappa_{(i,j)} = \dfrac{1}{\delta\theta_{(i,j)}}, & \pi < \theta_{(i,j)} < 1 \\ \kappa_{(i,j)} = 1, & \text{others} \end{cases} \tag{2}$$

where δ is a predefined constant and π is a threshold set to judge whether this response bucket contains disruptors.

Remark 1: The response bucket where the global maximum value $m_{\max}$ located will not be penalized, and the penalty factor κ in this bucket is set to 1.

After that, the penalty mask $\mathbf{V}_{k-f}$ can be obtained as

$$\mathbf{V}_{k-f} = \begin{bmatrix} \kappa_{(1,1)} & \kappa_{(1,2)} & \cdots & \kappa_{(1,\alpha)} \\ \kappa_{(2,1)} & \kappa_{(2,2)} & \cdots & \kappa_{(2,\alpha)} \\ \vdots & \vdots & \ddots & \vdots \\ \kappa_{(\alpha,1)} & \kappa_{(\alpha,2)} & \cdots & \kappa_{(\alpha,\alpha)} \end{bmatrix}. \tag{3}$$

The response map $\mathbf{M}_{k-f}$ is then denoised as $\mathbf{V}_{k-f} \odot \mathbf{M}_{k-f}$.

Remark 2: There are no measures to denoise the interval responses in the methods mentioned in Section II, including our baseline ARCF. The proposed disruptor-aware scheme can detect and repress disruptors in the responses effectively and accurately. Due to that, the discriminability of the IBRI tracker is enhanced.

B. Interval-Based Response Inconsistency

Since aerial object tracking contains many visual challenging scenes, e.g., fast object/aerial platform motion, it is tough to maintain a robust tracking performance with the information in only two consecutive frames. Nevertheless, tracking with hundreds of historical training samples is hard to keep real-time speed and occupies sizeable memory. Consequently, this work proposes a tracking approach by suppressing the interval-based response inconsistency. As presented in Fig. 3, in the filter training process of the kth frame, response maps in the previous F frames are introduced, the learned filter is expected to perceive the variations of the object and performs an accurate estimation in the oncoming frame.

Remark 3: Incorporating historical interval responses that does not occupy much memory but repress aberrances efficiently, the improvements in robustness are obtained without sacrificing the real-time speed. The proposed IBRI pipeline meets the real-time requirement of typical aerial tracking platforms with a favorable tracking performance.

C. IBRI Framework

An original temporal regularization is presented to measure the changes of the object's response in the current frame relative to those of the multiple historical frames. Therefore, the correlation filter $\mathbf{w}^d \in \mathbb{R}^N$ of the IBRI tracker can be obtained by minimizing the objective function as follows:

$$\mathcal{E}(\mathbf{w}) = \frac{1}{2}\left\|\mathbf{y} - \sum_{d=1}^{D} \mathbf{x}_k^d \star (\mathbf{P}^\mathsf{T}\mathbf{w}_k^d)\right\|_2^2 + \frac{\lambda}{2}\sum_{d=1}^{D}\left\|\mathbf{w}_k^d\right\|_2^2 + \sum_{f=1}^{F}\frac{\gamma_f}{2}\left\|\mathbf{V}_{k-f} \odot \mathbf{M}_{k-f}[\phi_{p_f,q_f}] - \mathbf{M}_k\right\|_2^2 \quad (4)$$

where $\mathbf{x}^d \in \mathbb{R}^N$ denotes the vectoried feature extracted from the training patch. The subscript d represents the dth channel of the total D feature channels. Besides, the symbols $\star$ and $\odot$ stand for the spatial correlation and the Hadamard product, respectively. The subscript k denotes the kth frame and $\mathbf{M}_k$ is equal to $\sum_{d=1}^{D} \mathbf{x}_k^d \star (\mathbf{P}^\mathsf{T}\mathbf{w}_k^d)$. The superscript T denotes the transpose operation. $\mathbf{P} \in \mathbb{R}^{N\times N}$ is a diagonal binary matrix borrowed from BACF [14] to address the boundary effect. To ensure that the temporal regularizer mainly focus on the appearance variation of the object, the peaks of the response maps are shifted to the current peak location by $[\phi_{p_f,q_f}]$. Since $\mathbf{M}_{k-f}$ is generated in the past frame, it can be considered constant. Moreover, $\gamma_f \in (0, 1)$ is a coefficient controlling the contribution of the fth temporal regularizer.

Remark 4: In practice, we assume $\gamma_f = \gamma_1^f$ to put a higher weight on the more recent information. The coefficient γ_1 in this work is set to 0.443.

By introducing an auxiliary variable $\mathbf{g}_k = \mathbf{P}^\mathsf{T}\mathbf{w}_k$, the augmented Lagrangian form of (4) can be reformulated as

$$\begin{aligned}\mathcal{E}(\mathbf{w}_k, \mathbf{g}_k, \zeta) &= \frac{1}{2}\left\|\mathbf{y} - \sum_{d=1}^{D} \mathbf{x}_k^d \star \mathbf{g}_k^d\right\|_2^2 + \frac{\lambda}{2}\sum_{d=1}^{D}\left\|\mathbf{w}_k^d\right\|_2^2 \\ &\quad + \sum_{f=1}^{F}\frac{\gamma_f}{2}\left\|\mathbf{V}_{k-f} \odot \mathbf{M}_{k-f}^{\mathrm{s}} - \sum_{d=1}^{D} \mathbf{x}_k^d \star \mathbf{g}_k^d\right\|_2^2 \\ &\quad + \sum_{d=1}^{D}(\mathbf{g}_k^d - \mathbf{P}^\mathsf{T}\mathbf{w}_k^d)^\mathsf{T}\zeta^d + \frac{\mu}{2}\sum_{d=1}^{D}\left\|\mathbf{g}_k^d - \mathbf{P}^\mathsf{T}\mathbf{w}_k^d\right\|_2^2 \end{aligned} \quad (5)$$

where μ is the penalty factor and $\zeta = [\zeta^1, \zeta^2, \ldots, \zeta^D] \in \mathbb{R}^{N\times D}$ is the Lagrange multiplier. The signal $\mathbf{M}_{k-f}[\phi_{p_f,q_f}]$ is written as $\mathbf{M}_{k-f}^{\mathrm{s}}$ for clarity. By constraining $\mathbf{r} = (\zeta/\mu)$, (5) can be reformulated as

$$\begin{aligned}\mathcal{E}(\mathbf{w}_k, \mathbf{g}_k, \mathbf{r}) &= \frac{1}{2}\left\|\mathbf{y} - \sum_{d=1}^{D} \mathbf{x}_k^d \star \mathbf{g}_k^d\right\|_2^2 + \frac{\lambda}{2}\sum_{d=1}^{D}\left\|\mathbf{w}_k^d\right\|_2^2 \\ &\quad + \sum_{f=1}^{F}\frac{\gamma_f}{2}\left\|\mathbf{V}_{k-f} \odot \mathbf{M}_{k-f}^{\mathrm{s}} - \sum_{d=1}^{D} \mathbf{x}_k^d \star \mathbf{g}_k^d\right\|_2^2 \\ &\quad + \frac{\mu}{2}\sum_{d=1}^{D}\left\|\mathbf{g}_k^d - \mathbf{P}^\mathsf{T}\mathbf{w}_k^d + \mathbf{r}^d\right\|_2^2. \end{aligned} \quad (6)$$

The ADMM technique [42] is applied to solve (6) by solving the following three subproblems iteratively:

$$\begin{cases} \mathbf{w}_k^{(i+1)} = \arg\min\limits_{\mathbf{w}_k} \left\{ \dfrac{\lambda}{2}\displaystyle\sum_{d=1}^{D}\left\|\mathbf{w}_k^{d(i)}\right\|_2^2 + \dfrac{\mu}{2}\displaystyle\sum_{d=1}^{D}\left\|\mathbf{g}_k^{d(i)} - \mathbf{P}^\mathsf{T}\mathbf{w}_k^{d(i)} + \mathbf{r}^{d(i)}\right\|_2^2 \right\} \\ \mathbf{g}_k^{(i+1)} = \arg\min\limits_{\mathbf{g}_K} \left\{ \dfrac{1}{2}\left\|\mathbf{y} - \displaystyle\sum_{d=1}^{D} \mathbf{x}_k^d \star \mathbf{g}_k^{d(i)}\right\|_2^2 + \displaystyle\sum_{f=1}^{F}\frac{\gamma_f}{2}\left\|\mathbf{V}_{k-f} \odot \mathbf{M}_{k-f}^{\mathrm{s}} - \sum_{d=1}^{D} \mathbf{x}_k^d \star \mathbf{g}_k^{d(i)}\right\|_2^2 + \dfrac{\mu}{2}\displaystyle\sum_{d=1}^{D}\left\|\mathbf{g}_k^{d(i)} - \mathbf{P}^\mathsf{T}\mathbf{w}_k^{d(i)} + \mathbf{r}^{d(i)}\right\|_2^2 \right\} \\ \mathbf{r}^{(i+1)} = \mathbf{r}^{(i)} + \mathbf{g}_k^{(i+1)} - \mathbf{P}^\mathsf{T}\mathbf{w}_k^{(i+1)} \end{cases} \quad (7)$$

where the superscript (i) denotes the ith iteration.

1) Subproblem $\mathbf{w}$*:* The first subproblem in (7) can be solved by deriving $\mathbf{w}_k$

$$\mathbf{w}_k^{d*(i+1)} = \frac{\mu N}{\lambda + \mu N}(\mathbf{g}_k^d + \mathbf{r}^d) \quad (8)$$

where the superscript * denotes the conjugate transpose operation.

Remark 5: For interpreting clarity, the superscripts (i) on the right-hand side of (8) are omitted. Since (8) only contains element-wise operations, it is computationally efficient during ADMM iterations.

2) Subproblem $\mathbf{g}$*:* Note that the time-consuming correlation operators are still contained in the second subproblem in (7). Thus, the convolution theorem is considered to simplify the calculation. Therefore, subproblem $\mathbf{g}$ is transferred to Fourier domain first

$$\begin{aligned}\hat{\mathbf{g}}_k^{(i+1)} &= \arg\min_{\hat{\mathbf{g}}_k^d} \left\{ \frac{1}{2} \left\| \hat{\mathbf{y}} - \sum_{d=1}^{D} \hat{\mathbf{x}}_k^d \odot \hat{\mathbf{g}}_k^d \right\|_2^2 \right. \\ &\quad + \sum_{f=1}^{F} \frac{\gamma_f}{2} \left\| \widehat{\mathbf{V}_{k-f} \odot \mathbf{M}_{k-f}}^{\mathrm{s}} - \sum_{d=1}^{D} \hat{\mathbf{x}}_k^d \odot \hat{\mathbf{g}}_k^d \right\|_2^2 \\ &\quad \left. + \frac{\mu}{2} \sum_{d=1}^{D} \left\| \hat{\mathbf{g}}_k^d - \sqrt{N} \mathbf{F} \mathbf{P}^{\mathsf{T}} \mathbf{w}_k^d + \hat{\mathbf{r}}^d \right\|_2^2 \right\} \end{aligned} \tag{9}$$

where the superscript $\hat{}$ denotes the discrete Fourier transformation (DFT) operater, i.e., $\hat{\mathbf{g}} = \sqrt{N}\mathbf{F}\mathbf{g}$.

Based on the element-wise matrix multiplication $\hat{\mathbf{x}}_k^d \odot \hat{\mathbf{g}}_k^d$ in (9), each element of $\hat{\mathbf{y}}$ (i.e., $\hat{\mathbf{y}}(n), n = 1, 2, \ldots, N$) is dependent only on $\hat{\mathbf{x}}(n) = [\hat{\mathbf{x}}_k^1(n), \hat{\mathbf{x}}_k^2(n), \ldots, \hat{\mathbf{x}}_k^D(n)]^{\mathsf{T}}$ and $\hat{\mathbf{g}}(n) = [\hat{\mathbf{g}}_k^1(n), \hat{\mathbf{g}}_k^2(n), \ldots, \hat{\mathbf{g}}_k^D(n)]^{\mathsf{T}}$. Therefore, (9) can be reformed as N independent objectives as

$$\begin{aligned}\hat{\mathbf{g}}_k^{(i+1)}(n) &= \arg\min_{\hat{\mathbf{g}}_k(n)} \left\{ \frac{1}{2} \left\| \hat{\mathbf{y}}(n) - \hat{\mathbf{x}}_k^{\mathsf{T}}(n) \hat{\mathbf{g}}_k(n) \right\|_2^2 \right. \\ &\quad + \sum_{f=1}^{F} \frac{\gamma_f}{2} \left\| \widehat{\mathbf{V}_{k-f} \odot \mathbf{M}_{k-f}}^{\mathrm{s}}(n) - \hat{\mathbf{x}}_k^{\mathsf{T}}(n) \hat{\mathbf{g}}_k(n) \right\|_2^2 \\ &\quad \left. + \frac{\mu}{2} \left\| \hat{\mathbf{g}}_k(n) - \sqrt{N} \mathbf{F} \mathbf{P}^{\mathsf{T}} \mathbf{w}_k(n) + \hat{\mathbf{r}}(n) \right\|_2^2 \right\}. \end{aligned} \tag{10}$$

Therefore, the subproblem $\mathbf{g}$ can be solved by solving N smaller problems, respectively, as follows:

$$\begin{aligned}\hat{\mathbf{g}}_k^{*(i+1)}(n) &= \frac{1}{\mu} \left(\mathbf{I}_D - \frac{\hat{\mathbf{x}}_k \hat{\mathbf{x}}_k^{\mathsf{T}}}{\mu b + \hat{\mathbf{x}}_k^{\mathsf{T}} \hat{\mathbf{x}}_k} \right) \\ &\quad \times \left(\hat{\mathbf{x}}_k(n) \hat{\mathbf{y}}_k(n) + \sum_{f=1}^{F} (\gamma_f \hat{\mathbf{x}}_k(n) \widehat{\mathbf{V}_{k-f} \odot \mathbf{M}_{k-f}}^{\mathrm{s}}(n)) \right. \\ &\quad \left. - \mu \hat{\mathbf{r}}(n) + \mu \hat{\mathbf{w}}_k(n) \right) \end{aligned} \tag{11}$$

where $\mathbf{I}_D$ is an identity matrix and

$$b = \frac{1}{1 + \sum_{f=1}^{F} \gamma_f}.$$

Remark 6: A detailed derivation of this subproblem is presented in the Appendix.

3) Lagrangian Multiplier Update: In each iteration, the Lagrangian multiplier is updated according to the last formula in (7). Following [42]

$$\mu^{(i+1)} = \min(\mu_{\max}, \beta \mu^{(i)}). \tag{12}$$

D. Model Update

In order to enhance the adaptability of DCF-based trackers in the face of unpredictable object appearance variations, a typical approach is to apply an online learning strategy [25]. In IBRI, the appearance model $\mathbf{x}^{\text{model}}$ is updated with a linear interpolation method as follows:

$$\mathbf{x}_k^{\text{model}} = (1 - \eta)\mathbf{x}_{k-1}^{\text{model}} + \eta \mathbf{x}_k \tag{13}$$

where η is the learning rate of the appearance model. In the training step, $\mathbf{x}^{\text{model}}$ is applied to (11) in the ADMM iterations to learn the correlation filter $\mathbf{w}$. Since $\mathbf{w}$ is trained by the integrated appearance model, it is learned online implicitly. Therefore, $\mathbf{w}$ is updated directly by a newly learned filter frame by frame.

Algorithm 1 IBRI Tracker

Input: A video sequence with K frames.
Position ($\mathbf{p}_1$) and size ($\mathbf{s}_1$) of the tracked object in the first frame I_1.

Output: Estimated position ($\mathbf{p}_k$) and size ($\mathbf{s}_k$) of the object in all upcoming frames.

```
Construct the Gaussian label function y.
Initialize the historical interval responses M_{k-F}, ..., M_{k-1} with a zero matrix.
for frame number k = 1 to end do
    if k = 1 then
        Crop and extract feature x_1^l from I_1 with p_1 and s_1.
        Initialize the appearance model x_l^model = x_1^l.
        Learn a correlation filter w_1 (Section III-C).
        Generate a response M_1 = x_1^l ⋆ w_1.
    else
        Crop and extract feature x_k^t from I_k with p_{k-1} and s_{k-1}.
        Generate the response map M_k = x_k^t ⋆ w_{k-1}.
        Detect disruptors in M_k and generate a penalty mask V_k (Section III-A).
        Estimate p_k and s_k according to M_k.
        Shift M_{k-f}, ..., M_{k-1} to make their peaks conincide with M_k.
        Crop and extract feature x_k^l from I_k with p_k and s_k.
        Update the appearance model using (13).
        Learn a correlation filter w_k and discard the old one. (Section III-C).
    end
end
```

E. Tracking Pipeline

The tracking pipeline of the proposed IBRI tracker can be summarized in Algorithm 1. When receiving a new frame, IBRI will crop a searching patch centered on the object location of the previous frame and extract features from the patch. After that, a response map will be generated by performing the circular correlation operation between the features and the filter trained in the previous frame. The new position of the object is considered as the location of the maximum response. Utilizing the scale estimation strategy in [43], the size of the object is updated. According to the new location of the object, a training patch is copped. Using the features extracted from the training patch, the correlation filter is retrained following Section III-C.

IV. EXPERIMENTS

In this section, the proposed tracker is extensively evaluated with other 35 state-of-the-art trackers on three well-known challenging aerial tracking benchmarks, i.e., UAVDT [23], UAV123@10fps [21], and VisDrone2018-SOT [22]. First, the proposed IBRI tracker is comprehensively verified for its effectiveness and superiority with other trackers based on handcrafted features. Second, ablation study and analysis of key parameters are conducted to evaluate the components in IBRI further. Third, the IBRI tracker is investigated in comparison with tracking approaches based on deep learning. Finally, the limitations of the proposed tracker are discussed.

A. Implementation Details

The proposed IBRI tracker is implemented in MATLAB R2018a. The same computer generates all the experimental results with an Intel i7-8700K CPU (3.7 GHz) and an NVIDIA 2080 GPU. Following the basic settings in ARCF [12], a combination of HOG [28], CN [44], and gray-scale features is employed for object representation in IBRI. The main parameters of the proposed IBRI tracker are presented in Table II. All parameters remain fixed for all image sequences in the following experiments. The source code is publicly available at https://github.com/vision4robotics/IBRI-tracker.

Remark 7: All the compared state-of-the-art approaches are evaluated with the open-source codes provided by the authors and remain default.

B. Evaluation Criteria

Based on one-pass evaluation (OPE), the experiments involve two metrics, i.e., precision and success rate [45]. These two metrics are employed to assess the precision and success rate for objective evaluation. Specifically, the tracking precision is defined by the Euclidean distance between the center of the ground truth and estimated bounding boxes, i.e., the center location error (CLE). As a result, the percentage of frames with a CLE below a given threshold (commonly set to 20 pixels) is employed to rank the trackers in the precision plot (PP). In terms of success rate, the IoU between the ground truth and estimated bounding boxes is reported in the success plot (SP), and this work uses the area under the curve (AUC) of SP to rank all the trackers. In addition, the operational speed is measured through frames per second (FPS).

C. Comparison With Trackers Based on Handcrafted Features

The proposed IBRI tracker is compared with other 15 state-of-the-art trackers using handcrafted features, i.e., KCF [25], DSST [43], SAMF [27], SRDCFdecon [19], Staple [46], BACF [14], CSR-DCF [47], ECO-HC [41], fDSST [26], SRDCF [15], Staple_CA [30], KCC [48], MCCT-H [49], STRCF [16], and ARCF [12]. The experiment is conducted on three different benchmarks.

1) Overall Performance: The proposed IBRI tracker outperforms other compared trackers on all three benchmarks.

TABLE II
MAIN PARAMETERS OF THE PROPOSED IBRI TRACKER

Paramter	Value	Paramter	Value
Number of ADMM iterations	5	Penalty factor μ	0.844
Regularization parameter λ	9.1	Learning rate η	0.019
Frame weight base γ_1	0.443	Step length β	10
Bucketing number $\alpha \times \alpha$	11×11	Interval length F	3
Disruptor threshold π	0.47	Disruptor factor δ	3

a) UAVDT: Consisting of 50 challenging sequences, the UAVDT benchmark is mainly focusing on aerial vehicle tracking in various challenging scenarios, e.g., high density, small object, camera motion, and occlusion. As shown in Fig. 5(a), the IBRI tracker achieves the best precision score (0.738), exceeding the second- and third-best tracker ARCF (0.720) and Staple_CA (0.695) by 2.5% and 6.2%. In terms of success rate, IBRI obtains the best score (0.470), followed by ARCF (0.458) and BACF (0.433). This excellent performance verifies that IBRI is suitable for complex aerial vehicle tracking scenarios.

b) UAV123@10fps: According to [21], the sequences in the UAV123@10fps benchmark are downsampled from the ones that are recorded in 30 FPS. Consequently, the appearance model of the tracked object can change more drastically between successive frames and more challenges are brought in. As presented in Fig. 5(b), the proposed tracker has achieved favorable performance. In terms of precision, IBRI has achieved the best performance with a precision score of 0.673 and has surpassed the second-best ARCF (0.666) with a margin of 1.1%. Similarly, in terms of success rate, IBRI (0.475) has also improved over ARCF (0.473) and ranks in the first place. By considering multiple historical frames, the proposed IBRI tracker has proved itself to be competent in handling large appearance changes.

c) VisDrone2018-SOT: The VisDrone2018-SOT benchmark is a challenge visual object tracking contest particularly designed for aerial object tracking. Compared with other generic tracking contest such as [50], the VisDrone2018-SOT contest evaluates the trackers on 35 image sequences and especially focuses on aerial tracking challenges, such as camera motion and viewpoint change. As shown in Fig. 5(c), the proposed IBRI tracker consistently ranks the first place in both precision score (0.817) and success score (0.593). Compared with other state-of-the-art trackers involved in the evaluation, the prominent performance of IBRI is more consistent in the used benchmarks, verifying its favorable robustness.

2) Attribute-Based Evaluation: To analyze the performance of different visual uncertainties in more detail, the UAVDT benchmark annotates the sequences into nine different attributes, i.e., background clutter (BC), camera motion (CM), illumination variations (IV), large occlusion (LO), long-term tracking (LTT), object blur (OB), object motion (OM), scale variations (SV), and small object (SO). In this section, a quantitative analysis of these attributes is performed.

Attribute-based performance evaluation of the IBRI tracker and other 15 state-of-the-art trackers using handcrafted features is shown in Table III. Specifically, the IBRI tracker ranks the first place in most of the attributes, making up seven in

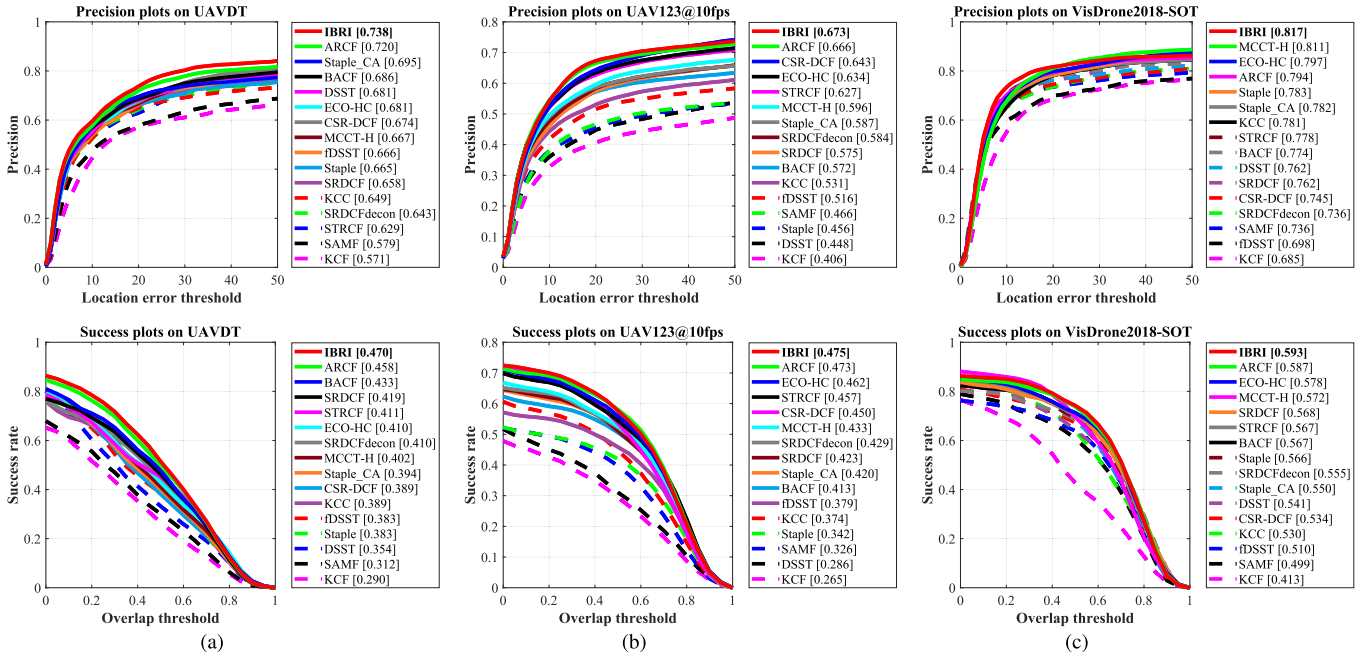

Fig. 5. PPs and SPs of the IBRI tracker and other 15 handcrafted feature-based trackers on three different benchmarks. The proposed IBRI tracker outperforms other state-of-the-art trackers in all challenging aerial tracking benchmarks. (a) Results on UAVDT. (b) Results on UAV123@10fps. (c) Results on VisDrone2018-SOT.

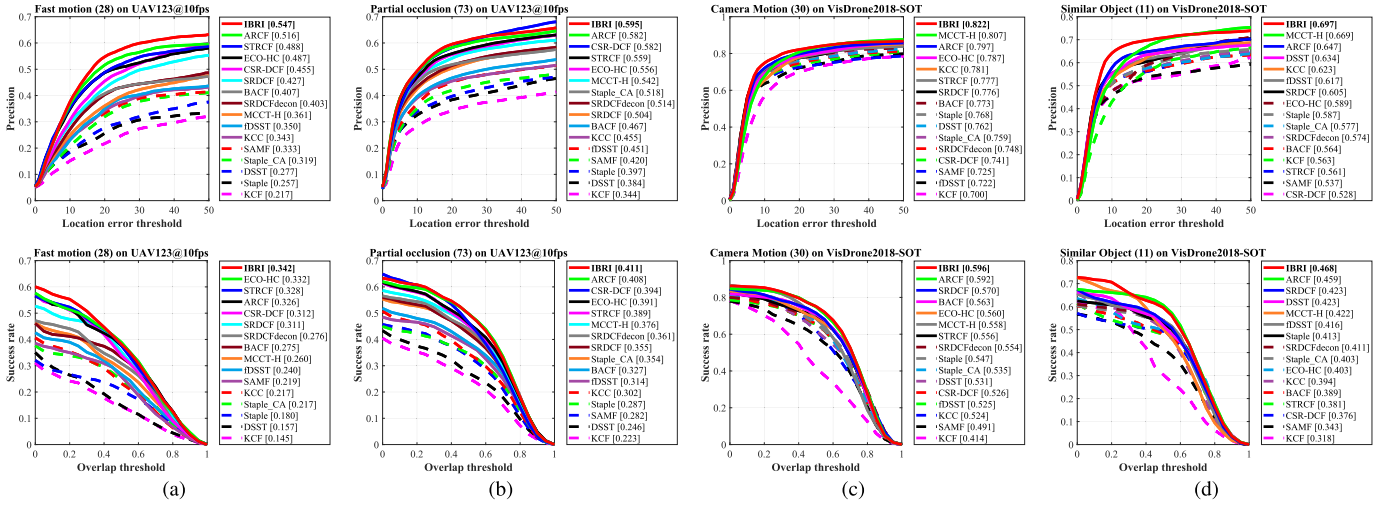

Fig. 6. Some attribute-based comparisons of the proposed IBRI tracker against other 15 state-of-the-art trackers on UAV123@10fps [Fig. 6(a) and (b)] and VisDrone2018-SOT [Fig. 6(c) and (d)]. IBRI shows excellent performance on the aerial tracking-specific attributes.

precision and six in success rate. Especially under the visually challenging factors unique to aerial tracking, IBRI achieves a remarkable improvement compared with the baseline tracker ARCF, i.e., more than 5% in BC, 7% in LO, and 4% in OM. Besides, there is merely a small gap between IBRI and the best performance in IV and OB.

Moreover, to further corroborate the robustness of the IBRI tracker on the aerial tracking-specific attributes (e.g., similar object, partial occlusion, and fast object/aerial platform motion), some comparisons on related attributes from UAV123@10fps and VisDrone2018-SOT are presented in Fig. 6. Specifically, on fast motion [Fig. 6(a)] and partial occlusion [Fig. 6(b)] from UAV123@10fps, IBRI outperforms other trackers. In both terms of camera motion [Fig. 6(c)] and similar object [Fig. 6(d)] from VisDrone2018-SOT, IBRI also shows attractive performance.

This competitive performance of the IBRI tracker is attributed to the introduction of the disruptor-aware scheme and the exploitation of the interval-based response inconsistency, thereby enhancing the discriminative capability against large appearance variation.

Remark 8: Long-term tracking is very common in the practical application of UAV. To further evaluate the performance of the proposed tracker in the face of long-term tracking, the sequences from UAV20L [21] and the sequences from UAVDT and VisDrone2018-SOT with the lengths longer than 1600 frames are combined as a long-term tracking benchmark, i.e., UAV-LTT. The UAV-LTT benchmark consists of 33 long-term sequences, with an average of 2638 frames per sequence. The evaluation results on the UAV-LTT benchmark are shown in Fig. 7. Note that the top five trackers in terms of overall performance, i.e., IBRI, BACF, ARCF, STRCF,

TABLE III

ATTRIBUTE-BASED EVALUATION OF THE IBRI TRACKER AND OTHER 15 STATE-OF-THE-ART TRACKERS USING HANDCRAFTED FEATURES ON THE UAVDT BENCHMARK. THE BEST THREE PERFORMANCES ARE, RESPECTIVELY, HIGHLIGHTED WITH RED, GREEN, AND BLUE COLOR. THE PROPOSED IBRI TRACKER RANKS NO.1 IN MOST OF THE ATTRIBUTES

	BC		CM		IV		LO		LTT		OB		OM		SV		SO	
Tracker	Prec.	Succ.	Prec.	Succ.	Prec.	Succ.	Prec.	Succ.	Prec.	Succ.	Prec.	Succ.	Prec.	Succ.	Prec.	Succ.	Prec.	Succ.
KCF [25]	0.458	0.235	0.534	0.267	0.657	0.312	0.345	0.229	0.675	0.312	0.653	0.298	0.455	0.244	0.491	0.254	0.581	0.251
DSST [43]	0.574	0.304	0.640	0.329	0.729	0.379	0.472	0.299	0.894	0.408	**0.714**	0.357	0.572	0.288	0.578	0.296	**0.824**	0.360
BACF [14]	0.599	**0.367**	0.614	**0.387**	0.739	**0.460**	0.488	0.340	0.886	**0.581**	0.698	**0.443**	0.604	**0.371**	0.604	**0.408**	0.770	**0.428**
SAMF [27]	0.503	0.268	0.568	0.283	0.650	0.315	0.374	0.256	0.675	0.340	0.590	0.297	0.469	0.257	0.447	0.264	0.634	0.290
Staple_CA [30]	0.589	0.326	**0.655**	0.349	**0.769**	0.428	0.493	0.324	**0.945**	0.539	0.713	0.405	**0.618**	0.344	**0.621**	0.366	0.796	0.379
SRDCF [15]	0.567	0.356	0.618	0.385	0.711	0.443	0.451	0.333	0.825	0.524	0.671	0.406	0.574	0.361	0.588	0.399	0.726	0.416
SRDCFdecon [19]	0.533	0.339	0.588	0.374	0.690	0.429	0.433	0.321	0.812	0.515	0.649	0.395	0.560	0.351	0.565	0.389	0.716	0.410
MCCT-H [49]	0.571	0.343	0.622	0.367	0.703	0.415	0.482	0.348	**0.925**	0.565	0.667	0.390	0.561	0.342	0.594	0.383	0.796	0.389
CSR-DCF [47]	0.599	0.345	0.613	0.347	0.697	0.398	**0.516**	**0.352**	0.874	0.487	0.666	0.373	0.583	0.342	0.599	0.366	0.760	0.355
STRCF [16]	0.517	0.340	0.552	0.365	0.659	0.421	0.420	0.319	0.849	0.525	0.641	0.406	0.517	0.341	0.543	0.388	0.749	0.421
ECO-HC [41]	**0.607**	0.364	0.647	0.379	0.723	0.434	0.504	0.347	**0.924**	0.573	0.669	0.391	0.596	0.358	0.607	0.389	0.767	0.375
fDSST [26]	0.558	0.315	0.650	0.363	0.708	0.395	0.483	0.332	0.863	0.518	0.660	0.369	0.552	0.312	0.552	0.337	0.790	0.380
KCC [48]	0.560	0.332	0.600	0.351	0.744	0.431	0.447	0.304	0.716	0.480	0.686	0.398	0.548	0.322	0.541	0.340	0.732	0.390
Staple [46]	0.552	0.313	0.614	0.335	0.728	0.413	0.441	0.308	0.891	0.517	0.709	0.401	0.567	0.330	0.601	0.358	0.777	0.370
ARCF [12]	**0.632**	**0.394**	**0.682**	**0.430**	**0.760**	**0.457**	**0.547**	**0.387**	0.882	**0.577**	**0.716**	**0.444**	**0.635**	**0.399**	**0.641**	**0.438**	**0.804**	**0.468**
IBRI	**0.664**	**0.416**	**0.688**	**0.441**	**0.761**	**0.459**	**0.590**	**0.417**	0.886	**0.582**	**0.717**	**0.443**	**0.661**	**0.418**	**0.670**	**0.457**	**0.808**	**0.470**

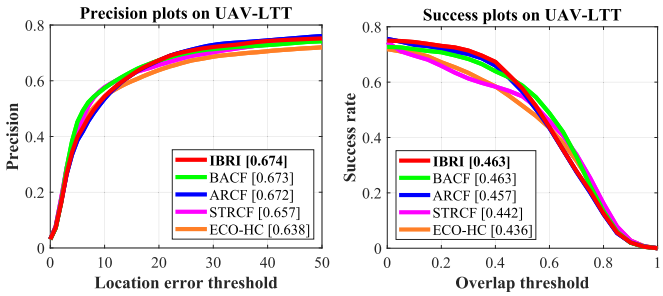

Fig. 7. Evaluation results on the UAV-LTT benchmark. The proposed IBRI tracker shows competitive performance in long-term tracking.

and ECO-HC, are involved in the evaluation. As shown in Fig. 7, IBRI ranks the first place in both terms of precision and success rate. This competitive performance demonstrates the effectiveness of the proposed IBRI tracker in long-term tracking.

D. Qualitative Evaluation

For more intuitive evaluation, some qualitative comparisons are presented with the proposed tracker and other top seven state-of-the-art trackers involved, i.e., ARCF, ECO-HC, STRCF, CSR-DCF, MCCT-H, Staple_CA, and BACF. Fig. 8 shows some screenshots of the tracking results on six challenge sequences, i.e., *S*1302, *S*1101 from UAVDT, person7_2, truck2 from UAV123@10fps, and uav0000093_00000_*s* and *uav*0000011_00000_*s* from VisDrone2018-SOT. The main challenges of these sequences include fast motion, occlusion, and similar object. Attributing to the suppression of disruptor-aware interval response inconsistency, IBRI achieves better tracking performance compared with other trackers.

E. Ablation Study

To validate the effectiveness of the proposed tracking approach, the tracking performance of the IBRI tracker with different modules and the baseline ARCF tracker on UAVDT is shown in Fig. 9. Denoting the interval-based response inconsistency learning as IRI and the disruptor-aware scheme as DA, ARCF can be regarded as a special case of IBRI when DA is disabled and the interval length F in IRI is set to 1.

As presented in Fig. 9, with the activation of IRI, the performance of IBRI (w/o DA) has surpassed the ARCF tracker, hence validating the effectiveness of constant inconsistency learning. Furthermore, by incorporating the DA, IBRI with both novel components achieves the best performance, with an improvement of 2.50% in precision (0.738) and 2.62% in success rate (0.470) compared with baseline.

F. Analysis of Key Parameter

To investigate the effect of the core parameters on the overall performance of IBRI, analysis of key parameters on the UAVDT benchmark is performed in this section. The bucketing number $\alpha \times \alpha$ and the number of interval length F are set to different numerical values for further verification.

1) Bucketing Number $\alpha \times \alpha$: α is set from 5 to 17 for the trial. In most cases, the peak of the response map is at the center. Thus, α is set to singular values to avoid damaging the area where the object is located. The results of precision and success rate are reported in Fig. 10. As α increases, the success rate reaches the highest point (0.470) at $\alpha = 11$. After that, the success rate fluctuates slightly until $\alpha = 17$. Likewise, the precision follows a similar path as the success rate and achieves the best score (0.738) at $\alpha = 11$. Compared to the IBRI tracker without the disruptor-aware scheme, the success rate and precision obtain a gain of 1.7% and 1.2%, respectively. The results show that when α is set in a certain range, the bucketing-based disruptor-aware scheme can detect and suppress the disruptors effectively. As a result, the discriminative power of IBRI is enhanced, thus leading to the improvement of overall performance. Therefore, α is set to 11 in this work.

2) Interval Length F: F is set from 1 to 10 for the trial, with a step size of 1. The results for precision and success rate are shown in Fig. 11. As F increases, the precision and success rate both reach the highest point at $F = 3$. After that, they gradually decline and become flat. Thus, a conclusion can be drawn that if F is set to an appropriate range, the utilization of

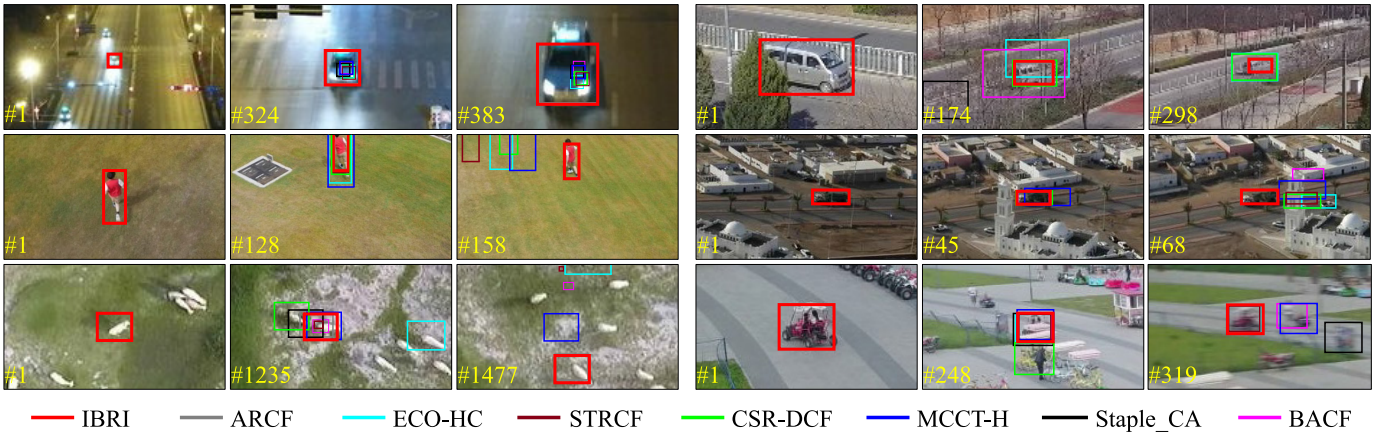

Fig. 8. Qualitative performance comparisons of the proposed IBRI tracker against other state-of-the-art trackers on challenging sequences (from left to right and top to bottom are *S*1302, *S*1101 from UAVDT, person7_2, truck2 from UAV123@10fps, and uav0000093_00000_*s* and uav0000011_00000_*s* from VisDrone2018-SOT). More aerial tracking examples are presented at https://youtu.be/5RFpQuDo6rc. The IBRI tracker shows superior aerial tracking performance.

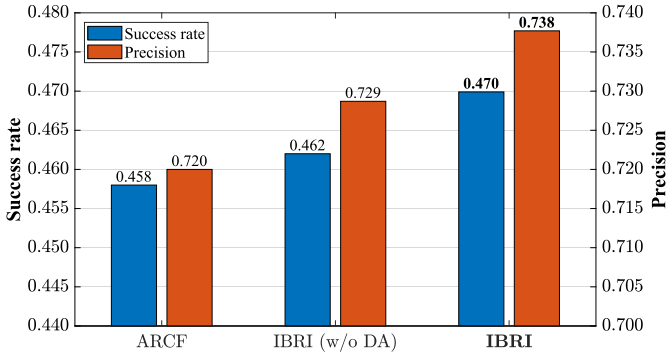

Fig. 9. Ablation studies of the IBRI tracker on the UAVDT benchmark. The overall results demonstrate the effectiveness of each module in the presented IBRI tracker.

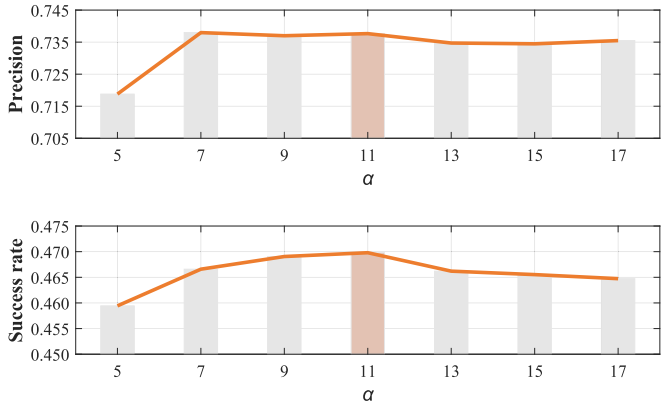

Fig. 10. Precision and success rate of the proposed method under different values of bucketing number $\alpha \times \alpha$ on the UAVDT benchmark. At $\alpha = 11$, both the precision and success rate plots reach the highest points.

the interval-based response inconsistency can indeed improve the tracking performance. Therefore, F is set to 3 in this work.

G. Comparison With Deep-Based Trackers

To further evaluate the tracking performance and effectiveness, the IBRI tracker is compared with other 20 recent state-of-the-art deep-based trackers on the UAVDT benchmark, including DeepSTRCF [16], ECO [41], CCOT [34],

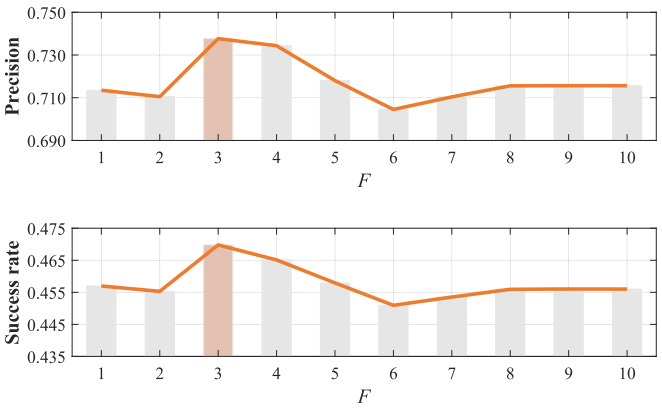

Fig. 11. Precision and success rate of the proposed method under different values of interval length F on the UAVDT benchmark. At $F = 3$, both the precision and success rate plots reach the highest points.

CFNet [33], ADNet [51], HDT [52], SINT [53], MCPF [54], UDT [55], UDT+ [55], IBCCF [56], ASRCF [37], TADT [57], CF2 [35], MCCT [49], CoKCF [58], DSiam [59], GOTURN [60], MDNet [31], and SiameseFC [61].

As presented in Table IV, IBRI tracker with handcrafted features outperforms other deep-based trackers in terms of precision and success rate while maintaining an attractive real-time speed. Specifically, IBRI achieves the best precision score (0.738) by improving 1.8% and 4.2% of the second-best tracker MDNet (0.725) and the third-best tracker SiameseFC (0.708), respectively. Besides, the highest success rate (0.470) also belongs to IBRI, followed by SiameseFC (0.465) and MDNet (0.464). As for tracking speed, SINT ranks the first place with 96.8 fps, followed by UDT (76.4) and UDT+ (60.4). IBRI obtains a real-time speed of 32.4 FPS.

Remark 9: It is worth to mention that running these deep-based trackers requires a high-performance GPU, which is not suitable for aerial tracking platforms with limited computing power. Instead, our IBRI tracker can run on a single CPU with a real-time speed of 32.4 FPS. With the competitive tracking performance, we believe that the IBRI tracker is suitable for aerial tracking.

TABLE IV

PERFORMANCE COMPARISON OF IBRI AND OTHER 20 STATE-OF-THE-ART DEEP-BASED TRACKERS ON THE UAVDT BENCHMARK. THE BEST THREE PERFORMANCES ARE, RESPECTIVELY, SHOWED IN RED, GREEN, AND BLUE COLOR. WITH A REAL-TIME SPEED, IBRI IS SUPERIOR TO OTHER DEEP-BASED TRACKERS IN PRECISION AND SUCCESS RATE

Tracker	Prec.	Succ.	FPS	Venue	GPU Usage
CF2	0.602	0.355	20.1	15'ICCV	✓
MDNet	**0.725**	**0.464**	1.0	16'CVPR	✓
HDT	0.596	0.303	9.0	16'CVPR	✓
SINT	0.570	0.290	**96.8**	16'CVPR	✓
GOTURN	0.702	0.451	16.5	16'ECCV	✓
CCOT	0.656	0.406	1.1	16'ECCV	✓
SiameseFC	**0.708**	**0.465**	18.2	16'ECCVW	✓
ECO	0.700	0.454	16.4	17'CVPR	✓
ADNet	0.683	0.429	7.5	17'CVPR	✓
CFNet	0.680	0.428	41.0	17'CVPR	✓
MCPF	0.660	0.403	0.7	17'CVPR	✓
DSiam	0.704	0.457	15.9	17'ICCV	✓
IBCCF	0.603	0.388	3.4	17'ICCVW	✓
CoKCF	0.605	0.319	21.2	17'PR	✓
MCCT	0.671	0.437	8.6	18'CVPR	✓
DeepSTRCF	0.667	0.437	6.6	18'CVPR	✓
UDT	0.674	0.441	**76.4**	19'CVPR	✓
ASRCF	0.700	0.437	14.1	19'CVPR	✓
TADT	0.677	0.431	32.5	19'CVPR	✓
UDT+	0.697	0.416	**60.4**	19'CVPR	✓
IBRI	**0.738**	**0.470**	32.4	Ours	✗

Fig. 12. Failure cases of the proposed IBRI tracker. The first to third rows are the tracking performances on *S*0801 from UAVDT, uav0000324_00069_*s* from VisDrone2018-SOT, and person8_1 from UAV123@10fps, respectively.

H. Limitations

Some failure cases of the proposed IBRI tracker from different benchmarks are presented in Fig. 12. In the first and second rows, the tracked object is completely obscured for a long while, during which the training patches without the appearance information of the object is used for filter learning. Thus, the interval response inconsistency provides limited contribution and is hard to guide accurate tracking in these scenes with long-term full occlusion. In the third row, due to the large deformation and rapid rotation of the object under background clutter, response maps in this period are chaotic. It is hard to distinguish and suppress the disruptors. Therefore, the IBRI tracker does not track well in these cases. Hence, we will focus on developing methods to incorporate useful historical information more self-adaptively in future work.

V. CONCLUSION

In this work, a novel DCF-based tracking approach exploiting disruptor-aware interval-based response inconsistency, i.e., IBRI, is proposed to carry out aerial tracking tasks. By adaptively denoising the response maps of multiple historical frames and constantly integrating them into the training process, IBRI can effectively utilize the historical information to maintain robust tracking performance. Comprehensive experiments are conducted on three challenging aerial tracking benchmarks. Extensive experimental results demonstrate that the proposed approach generally outperforms the aerial tracking performance in terms of accuracy and robustness compared with other 35 state-of-the-art trackers. Especially in the face of common challenges of aerial tracking, e.g., fast object/aerial platform motion, full/partial occlusion, and similar object, our tracker shows competitive robustness. With low computation requirements, it is compelling that the IBRI is suitable for real-time aerial tracking tasks. We strongly believe that the proposed IBRI tracker can greatly contribute to aerial tracking in terms of accuracy, robustness, and efficiency.

APPENDIX
DERIVATION OF THE SUBPROBLEM $\mathbf{g}$

This appendix presents the detailed derivation process of the subproblem $\mathbf{g}$ from (10) to (11).

Taking the partial derivatives of $\hat{\mathbf{g}}_k(n)$, (10) can be rewritten as

$$\begin{aligned}&\hat{\mathbf{g}}_k(n)^{*(i+1)}\\&= b\left(\hat{\mathbf{x}}_k(n)\hat{\mathbf{x}}_k^{\mathsf{T}}(n)+\mu b\mathbf{I}_D\right)^{-1}\Big(\hat{\mathbf{x}}_k(n)\hat{\mathbf{y}}_k(n)\\&\quad+\sum_{f=1}^{F}(\gamma_f\hat{\mathbf{x}}_k(n)\overset{\mathrm{s}}{\widehat{\mathbf{V}_{k-f}\odot\mathbf{M}_{k-f}}}(n))-\mu\hat{\mathbf{r}}(n)+\mu\hat{\mathbf{w}}_k(n)\Big)\end{aligned}\tag{14}$$

where $b=(1/(1+\sum_{f=1}^{F}\gamma_f))$.

Since (14) contains matrix inversion, it will be time-consuming to solve directly without simplification. To reduce the computation load, the Sherman–Morrison formula as shown in (15) is adopted here

$$(\mathbf{A}+\mathbf{u}\mathbf{v}^{\mathsf{T}})^{-1}=\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{u}(\mathbf{I}_D+\mathbf{v}^{\mathsf{T}}\mathbf{A}^{-1}\mathbf{u})^{-1}\mathbf{v}^{\mathsf{T}}\mathbf{A}^{-1}.\tag{15}$$

In this case, $\mathbf{A}=(\mu/(1+\sum_{f=1}^{F}\gamma_f))\mathbf{I}_D$, and $\mathbf{u}=\mathbf{v}=\hat{\mathbf{x}}_k(n)$. Therefore, (14) can be further simplified as

$$\begin{aligned}&\hat{\mathbf{g}}_k^{*(i+1)}(n)\\&=\frac{1}{\mu}\left(\mathbf{I}_D-\frac{\hat{\mathbf{x}}_k\hat{\mathbf{x}}_k^{\mathsf{T}}}{\mu b+\hat{\mathbf{x}}_k^{\mathsf{T}}\hat{\mathbf{x}}_k}\right)\Big(\hat{\mathbf{x}}_k(n)\hat{\mathbf{y}}_k(n)\\&\quad+\sum_{f=1}^{F}(\gamma_f\hat{\mathbf{x}}_k(n)\overset{\mathrm{s}}{\widehat{\mathbf{V}_{k-f}\odot\mathbf{M}_{k-f}}}(n))-\mu\hat{\mathbf{r}}(n)+\mu\hat{\mathbf{w}}_k(n)\Big).\end{aligned}\tag{16}$$

Note that (16) only consists of element operations and is efficient in ADMM iterations. ■

REFERENCES

[1] S. Xuan, S. Li, M. Han, X. Wan, and G.-S. Xia, "Object tracking in satellite videos by improved correlation filters with motion estimations," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 2, pp. 1074–1086, Feb. 2020.

[2] J. Caro-Gutierrez, M. E. Bravo-Zanoguera, and F. F. González-Navarro, "Methodology for automatic collection of vehicle traffic data by object tracking," in *Proc. Adv. Comput. Intell.*, 2017, pp. 482–493.

[3] J. Shao, B. Du, C. Wu, and L. Zhang, "Tracking objects from satellite videos: A velocity feature based correlation filter," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7860–7871, Oct. 2019.

[4] B. Du, Y. Sun, S. Cai, C. Wu, and Q. Du, "Object tracking in satellite videos by fusing the kernel correlation filter and the three-frame-difference algorithm," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 168–172, Feb. 2018.

[5] M. Thomas, C. Kambhamettu, and C. A. Geiger, "Motion tracking of discontinuous sea ice," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 5064–5079, Dec. 2011.

[6] Z. He, S. Yi, Y.-M. Cheung, X. You, and Y. Yan Tang, "Robust object tracking via key patch sparse representation," *IEEE Trans. Cybern.*, vol. 47, no. 2, pp. 354–364, Feb. 2017.

[7] K. Dai, Y. Zhang, D. Wang, J. Li, H. Lu, and X. Yang, "High-performance long-term tracking with meta-updater," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 6298–6307.

[8] A. Lukežič, L. V. C. Zajc, T. Vojíř, J. Matas, and M. Kristan, "Fucolot – a fully-correlational long-term tracker," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, 2018, pp. 595–611.

[9] L. Huang, X. Zhao, and K. Huang, "Globaltrack: A simple and strong baseline for long-term tracking," in *Proc. Assoc. Adv. Artif. Intell. (AAAI)*, 2020, pp. 1–8.

[10] B. Yan, H. Zhao, D. Wang, H. Lu, and X. Yang, "Skimming-Perusal' tracking: A framework for real-time and robust long-term tracking," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2385–2393.

[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2015, pp. 91–99.

[12] Z. Huang, C. Fu, Y. Li, F. Lin, and P. Lu, "Learning aberrance repressed correlation filters for real-time UAV tracking," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2891–2900.

[13] C. Fu, F. Lin, Y. Li, and G. Chen, "Correlation filter-based visual tracking for UAV with online multi-feature learning," *Remote Sens.*, vol. 11, no. 5, p. 549, Mar. 2019.

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

[15] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 4310–4318.

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

[17] M. Wang, Y. Liu, and Z. Huang, "Large margin object tracking with circulant feature maps," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4021–4029.

[18] Y. Han, C. Deng, B. Zhao, and D. Tao, "State-aware anti-drift object tracking," *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 4075–4086, Aug. 2019.

[19] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, "Adaptive decontamination of the training set: A unified formulation for discriminative visual tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1430–1438.

[20] B. Kitt, A. Geiger, and H. Lategahn, "Visual odometry based on stereo image sequences with RANSAC-based outlier rejection scheme," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2010, pp. 486–492.

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

[22] L. Wen *et al.*, "Visdrone-sot2018: The vision meets drone single-object tracking challenge results," in *Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops*, 2019, pp. 469–495.

[23] D. Du *et al.*, "The unmanned aerial vehicle benchmark: Object detection and tracking," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 370–386.

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

[25] 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.

[26] 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, Aug. 2017.

[27] Z. Liu, Z. Lian, and Y. Li, "A novel adaptive kernel correlation filter tracker with multiple feature integration," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 254–265.

[28] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2008, pp. 886–893.

[29] J. van de Weijer, C. Schmid, J. Verbeek, and D. Larlus, "Learning color names for real-world applications," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1512–1523, Jul. 2009.

[30] M. Mueller, N. Smith, and B. Ghanem, "Context-aware correlation filter tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1396–1404.

[31] H. Nam and B. Han, "Learning multi-domain convolutional neural networks for visual tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4293–4302.

[32] C. Sun, D. Wang, H. Lu, and M.-H. Yang, "Learning spatial-aware regressions for visual tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8962–8970.

[33] J. Valmadre, L. Bertinetto, J. Henriques, A. Vedaldi, and P. H. S. Torr, "End-to-end representation learning for correlation filter based tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2805–2813.

[34] M. Danelljan, A. Robinson, F. S. Khan, and M. Felsberg, "Beyond correlation filters: Learning continuous convolution operators for visual tracking," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 472–488.

[35] C. Ma, J.-B. Huang, X. Yang, and M.-H. Yang, "Hierarchical convolutional features for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3074–3082.

[36] Y. Han, C. Deng, B. Zhao, and B. Zhao, "Spatial-temporal context-aware tracking," *IEEE Signal Process. Lett.*, vol. 26, no. 3, pp. 500–504, Mar. 2019.

[37] K. Dai, D. Wang, H. Lu, C. Sun, and J. Li, "Visual tracking via adaptive spatially-regularized correlation filters," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4670–4679.

[38] T. Xu, Z.-H. Feng, X.-J. Wu, and J. Kittler, "Joint group feature selection and discriminative filter learning for robust visual object tracking," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 7950–7960.

[39] J. Zhang, S. Ma, and S. Sclaroff, "MEEM: Robust tracking via multiple experts using entropy minimization," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 188–203.

[40] W. Wang, K. Zhang, M. Lv, and J. Wang, "Hierarchical spatiotemporal context-aware correlation filters for visual tracking," *IEEE Trans. Cybern.*, early access, Jan. 30, 2020, doi: 10.1109/TCYB.2020.2964757.

[41] M. Danelljan, G. Bhat, F. S. Khan, and M. Felsberg, "ECO: Efficient convolution operators for tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6638–6646.

[42] S. Boyd, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.

[43] M. Danelljan, G. Häger, F. Shahbaz Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," in *Proc. Brit. Mach. Vis. Conf.*, 2014, pp. 1–11.

[44] M. Danelljan, F. S. Khan, M. Felsberg, and J. V. D. Weijer, "Adaptive color attributes for real-time visual tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1090–1097.

[45] Y. Wu, J. Lim, and M. H. Yang, "Object tracking benchmark," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1834–1848, Sep. 2015.

[46] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, and P. H. S. Torr, "Staple: Complementary learners for real-time tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1401–1409.

[47] A. Lukezic, T. Vojir, L. C. Zajc, J. Matas, and M. Kristan, "Discriminative correlation filter with channel and spatial reliability," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6309–6318.

[48] C. Wang, L. Zhang, L. Xie, and J. Yuan, "Kernel cross-correlator," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, 2018, pp. 4179–4186.
[49] N. Wang, W. Zhou, Q. Tian, R. Hong, M. Wang, and H. Li, "Multi-cue correlation filters for robust visual tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4844–4853.
[50] M. Kristan *et al.*, "The sixth visual object tracking vot2018 challenge results," in *Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops*, Sep. 2018, pp. 1–10.
[51] S. Yun, J. Choi, Y. Yoo, K. Yun, and J. Y. Choi, "Action-decision networks for visual tracking with deep reinforcement learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2711–2720.
[52] Y. Qi *et al.*, "Hedged deep tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4303–4311.
[53] R. Tao, E. Gavves, and A. W. M. Smeulders, "Siamese instance search for tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1420–1429.
[54] T. Zhang, C. Xu, and M.-H. Yang, "Multi-task correlation particle filter for robust object tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4335–4343.
[55] N. Wang, Y. Song, C. Ma, W. Zhou, W. Liu, and H. Li, "Unsupervised deep tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1308–1317.
[56] F. Li, Y. Yao, P. Li, D. Zhang, W. Zuo, and M.-H. Yang, "Integrating boundary and center correlation filters for visual tracking with aspect ratio variation," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2017, pp. 2001–2009.
[57] X. Li, C. Ma, B. Wu, Z. He, and M.-H. Yang, "Target-aware deep tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1369–1378.
[58] L. Zhang and P. N. Suganthan, "Robust visual tracking via co-trained kernelized correlation filters," *Pattern Recognit.*, vol. 69, pp. 82–93, Sep. 2017.
[59] Q. Guo, W. Feng, C. Zhou, R. Huang, L. Wan, and S. Wang, "Learning dynamic siamese network for visual object tracking," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1763–1771.
[60] D. Held, S. Thrun, and S. Savarese, "Learning to track at 100 FPS with deep regression networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 749–765.
[61] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. S. Torr, "Fully-convolutional siamese networks for object tracking," in *Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops*, 2016, pp. 850–865.

Changhong Fu (Member, IEEE) received the Ph.D. degree in robotics and automation from the Computer Vision and Aerial Robotics (CVAR) Laboratory, Technical University of Madrid, Madrid, Spain.

He held two research positions with Arizona State University, Tempe, AZ, USA, and Nanyang Technological University (NTU), Singapore. He was with NTU as a Post-Doctoral Research Fellow. He has worked on two international, two national, and six industrial projects related to the vision for UAV. He is an Assistant Professor with the School of Mechanical Engineering, Tongji University, Shanghai, China, and leading six projects related to the vision for unmanned systems (US). He has authored or coauthored more than 50 journal articles and conference papers [including the IEEE TRANSACTIONS ON MULTIMEDIA (TMM), IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS), IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (TCSVT), IEEE/ASME TRANSACTIONS ON MECHATRONICS (TMECH), IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS (TIE), CVPR, ICCV, IROS and ICRA] related to the intelligent vision and control for UAV. His research areas are intelligent vision and control for US in complex environment.

Junjie Ye received the B.Eng. degree in mechanical engineering from Tongji University, Shanghai, China, where he is pursuing the M.Sc. degree in mechanical engineering.

His research interests include visual object tracking, deep learning, and robotics.

Juntao Xu received the B.Eng. degree in mechanical engineering from Tongji University, Shanghai, China.

His research interests involve visual object tracking and computer vision.

Yujie He received the B.Eng. degree in mechanical engineering from Tongji University, Shanghai, China.

His research interests include robotics, visual object tracking, and place recognition.

Fuling Lin (Graduate Student Member, IEEE) received the B.Eng. degree in mechanical engineering from Tongji University, Shanghai, China, where he is pursuing the M.Sc. degree in mechanical engineering.

His research interests include robotics, visual object tracking, and computer vision.