Object Saliency-Aware Dual Regularized Correlation Filter for Real-Time Aerial Tracking

Changhong Fu, *Member, IEEE*, Juntao Xu, Fuling Lin, Fuyu Guo, Tingcong Liu, and Zhijun Zhang, *Senior Member, IEEE*

***Abstract*— Spatial regularization has been proved as an effective method for alleviating the boundary effect and boosting the performance of a discriminative correlation filter (DCF) in aerial visual object tracking. However, existing spatial regularization methods usually treat the regularizer as a supplementary term apart from the main regression and neglect to regularize the filter involved in the correlation operation. To address the aforementioned issue, this article introduces a novel object saliency-aware dual regularized correlation filter, i.e., DRCF. Specifically, the proposed DRCF tracker suggests a dual regularization strategy to directly regularize the filter involved with the correlation operation inside the core of the filter generating ridge regression. This allows the DRCF tracker to suppress the boundary effect and consequently enhance the performance of the tracker. Furthermore, an efficient method based on a saliency detection algorithm is employed to generate the dual regularizers dynamically and provide the regularizers with online adjusting ability. This enables the generated dynamic regularizers to automatically discern the object from the background and actively regularize the filter to accentuate the object during its unpredictable appearance changes. By the merits of the dual regularization strategy and the saliency-aware dynamical regularizers, the proposed DRCF tracker performs favorably in terms of suppressing the boundary effect, penalizing the irrelevant background noise coefficients and boosting the overall performance of the tracker. Exhaustive evaluations on 193 challenging video sequences from multiple well-known challenging aerial object tracking benchmarks validate the accuracy and robustness of the proposed DRCF tracker against 27 other state-of-the-art methods. Meanwhile, the proposed tracker can perform real-time aerial tracking applications on a single CPU with sufficient speed of 38.4 frames/s.**

Manuscript received January 16, 2020; revised March 23, 2020; accepted April 23, 2020. Date of publication May 14, 2020; date of current version November 24, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61806148 and in part by the State Key Laboratory of Mechanical Transmissions, Chongqing University, under Grant SKLMT-KFKT-201802. *(Corresponding author: Changhong Fu.)*

Changhong Fu, Juntao Xu, and Fuling Lin are with the School of Mechanical Engineering, Tongji University, Shanghai 201804, China (e-mail: changhongfu@tongji.edu.cn).

Fuyu Guo is with the School of Mechanical Engineering, Chongqing University, Chongqing 400044, China (e-mail: guofuyu@cqu.edu.cn).

Tingcong Liu is with the College of Liberal Arts and Sciences, University of Illinois at Urbana–Champaign, Champaign, IL 61820 USA (e-mail: tl17@illinois.edu).

Zhijun Zhang is with the School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China (e-mail: auzjzhang@scut.edu.cn).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2020.2992301

***Index Terms*— Aerial visual object tracking, discriminative correlation filter (DCF), dual regularization strategy, saliency-based dynamical regularizer.**

I. Introduction

AERIAL visual object tracking is a fast developing aspect of remote sensing and has attracted a great deal of research interests [1]–[3]. By unleashing the great maneuverability and the flexibility of the aerial tracking platforms, visual object tracking can enable them to carry out various tasks, such as geolocation enhancement [4], forest fire analysis [5], and moving object tracking [6], to name but a few. However, the successful integration of visual tracking onboard these platforms is still a demanding issue in practice due to many visual uncertainties and special application limitations. Especially for the widely used unmanned aerial vehicle (UAV) platforms, the inadequate visual sampling resolution, the fast motion of both UAV and the tracked object, the limited onboard computational resources, and the strict requirement of real-time performance are all among the many challenges that a vision-based aerial tracking method needs to resolve.

In recent years, discriminative correlation filter (DCF)-based tracking methods have gained enormous popularity and shown impressive computational efficiency in addition to adequate tracking accuracy [7]. Based on the convolution theorem, DCF-based tracking algorithms can transfer the computation consuming correlation operation from the spatial domain to the frequency domain. By substituting the operation with an efficient elementwise multiplication, the computational complexity can be substantially reduced as a result. Nevertheless, the circular correlation operation that DCF-based methods rely on incurs periodic extension of the training image samples [8] and would further result in sample discontinuity at the extension boundaries, i.e., the boundary effect. This unwanted sample discontinuity feeds the filter with corrupted image samples in the training procedure and further restricts the search area when the filter tries to locate the object. Consequently, the discriminative power of the filter is hampered and its overall tracking performance is less than satisfactory.

In this regard, Danelljan *et al*. [9] proposed the spatial regularized DCF (SRDCF) and modulated the standard DCF formulation with a spatial regularization window to mitigate

the boundary effect. Spatial regularization allows the filter to focus more on the information from the center, where the object is usually located. Meanwhile, the search region can also be enlarged so that more surrounding information is included. In this way, spatial regularization can generally improve the tracking performance of a tracker. However, this approach still does not completely resolve the boundary effect for several reasons. First, the adopted spatial regularization term serves merely as a supplement to the main correlation regression, where the boundary effect really emerges, and thus the issue the approach is designed to treat has only been inadequately addressed. Second, the aforementioned regularization window is static and thus tends to be less effective when the tracked object undergoes fast motion and/or deformation during tracking. Third, the Gauss–Seidel solving method in SRDCF also lacks the computational speed that a normal aerial tracking application requires.

In order to tackle the aforementioned problems, this article presents a novel object saliency-aware dual regularized correlation filter (DRCF) tracker. This proposed tracker is capable of alleviating the boundary effect more effectively with an original dual regularization strategy. In addition, its saliency-aware regularizers can accentuate the object more precisely. The tracker is further optimized with the adoption of an efficient alternating direction method of multipliers (ADMM) [10] to fit the real-time performance requirement in aerial tracking applications. Details of the main contributions in this article can be listed as follows.

1) An original dual regularization strategy is introduced to further resolve the deleterious boundary effect in DCF-based visual object tracking. This strategy enables the the implementation of a novel dual regularizer inside the main learning regression that directly regulates the filter involved with the correlation operation. Consequently, the boundary effect is more evidently suppressed and the discriminative power of the filter is enhanced.
2) The object saliency-aware regularizers in DRCF are dynamically constructed according to the changing appearance of the object based on an effective saliency detection method. By enabling the regularizers to recognize the region corresponding to the object, the proposed DRCF tracker can perform more targeted filter coefficient penalization and suppress irrelevant background noise during the object deformation.
3) The ADMM method is adopted to dissect the proposed dual regularized filter training optimization problem, from which an efficient closed-form iterative solution with multiple feature channels is further derived. The computational efficiency of this formulation promises the DRCF a favorable performance speed of 38.4 frames per second (FPS) on a single CPU for real-time aerial tracking applications.
4) A considerable number of experiments have been undertaken on multiple challenging aerial tracking benchmarks to evaluate the performance of the proposed DRCF tracker. As shown in Fig. 1, experimental evaluations demonstrate that the DRCF tracker performs competitively compared with other leading-edge trackers in terms of accuracy, robustness, and efficiency.

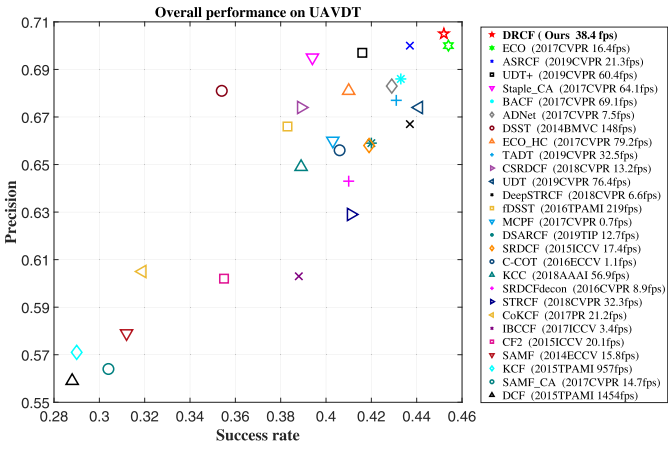

Fig. 1. Overall precision and success rate of the state-of-the-art and the proposed DRCF tracker on the UAVDT benchmark [11]. The running speed of the trackers in FPS is also presented to evaluate their efficiency.

The remainder of this article is organized as follows. Section II revisits the related previous DCF-based trackers. Section III elaborates on the main components and the tracking procedure of the proposed DRCF tracker. Specifically, by integrating a novel dual regulation strategy and an efficient saliency detection method, the main objective of the proposed DRCF tracker is formulated and solved. Section IV presents the qualitative and quantitative experimental evaluations of the DRCF tracker compared with other state-of-art tracking approaches. In addition, an ablative study is conducted to evaluate the effectiveness and contribution of each component in the DRCF tracker. Finally, conclusions are drawn in Section V.

II. Related Works

Discriminative object tracking, otherwise known as the tracking by detection method, has found numerous applications in aerial visual tracking and is essentially an online machine learning problem [12]. Given only the object location in the first frame of an aerial image sequence, a discriminative vision-based aerial tracker needs to train a classifier online to discern the tracked object from the background and estimate its trajectory in the rest of the sequence [13]. A great deal of significant works have recently been done in this field. Sections II-A–II-C, respectively, cover three main aspects and go over the relevant state-of-the-art tracking approaches that motivate the proposed DRCF tracker.

A. Tracking With DCF

First proposed as the minimum output sum of the squared error method [7], the DCF-based tracking approaches use the correlation filter as the discriminative classifier and have gained popularity due to their unprecedented accuracy and impressive running speed. Later with the unveiling of the circular structure in DCFs [8], [14], their overall tracking

performances have been successively improved through the introduction of dense sampling technique and kernelized methods [8], [15], [16]. Meanwhile, the incorporation of multichannel features, such as color name (CN) feature [17] and histogram of oriented gradient (HOG) feature [18], has further improved the well-roundedness of the object appearance model. In addition, the emergence of more advanced online model updating strategies [19]–[21] has enabled DCF-based trackers to identify and remember previous reliable appearance patterns, increasing their tracking accuracy and robustness. Moreover, the development of online scale estimation approaches [22]–[24] has empowered DCF trackers to handle the scale variance of the object, which can frequently occur in many tracking scenarios. More recently, convolution neural network (CNN)-based features have been integrated into DCF frameworks [25]–[27] to further enrich the feature channels and achieve a more comprehensive object appearance representation. Although the influx of the modifications listed above has brought significant progress to the state-of-the-art DCF trackers, the accumulation of improvement also incurs the reduction in the tracking speed. Even though some methods like continuous convolution [28] and efficient convolution with dimension reduction [29] have lately attempted to mitigate the extra computational burdens, designing a tracker with both competitive performance and favorable running speed is still a problem worth studying.

B. DCF Tracking With Spatial Regularization

As discussed in Section I, the circulant-shifted samples in DCF-based trackers always suffer from periodic repetitions on boundary positions, which further lead to sample contamination and performance degradation. To relieve this unwanted boundary effect, Danelljan *et al.* [9] proposed the aforementioned SRDCF to penalize the filter coefficients according to their spatial distance to the center of the sample. The farther the coefficient is from the center, the greater the penalization will be. By this means, the filter will absorb more information from the center region of the sample, where the target usually locates. Consequently, the effects of the boundary are suppressed. To better identify the region suitable for tracking, Fu *et al.* [13] introduced the protective region to circumscribe the object. Lukežič *et al.* [30] proposed a spatial reliability estimation to enable the regularizer to actively adapt to possible object appearance change. As more sophisticated modifications to the spatial regularization, Dai *et al.* [31] has attempted to optimize the regularizer with a twofold ADMM method. Zhang *et al.* [32] has demonstrated a nonlocal regularized tracker. Generally speaking, spatial regularization methods have profoundly improved upon the conventional DCF tracking framework with their performance enhancements and are widely recognized as an essential part of subsequent state-of-the-art trackers [28], [29]. However, these regularized trackers usually treat the regularization term as a supplementary component apart from the main regression. For this reason, the boundary effect on the correlation operation itself is left unaddressed.

C. DCF Tracking With Saliency-Awareness

As an approach capable of detecting the noteworthy region in an image and handling the irregular shape of the object, saliency-based methods are readily applied in many DCF tracking-by-detection approaches to provide performance enhancements. Proposed in [12], object saliency-awareness can help the DCF tracker discern the object from the background as an additional feature providing more comprehensive information. In [33], a saliency map was taken as a generative prior and fused with the discriminative filter response to increase the accuracy in object location estimation. Moreover, saliency detection can be exploited as a candidate provider and indicates the possible occurrence location of the object to alleviate model drift problems [34]. Recently, the saliency proposal has been integrated into [35] to provide information about the object and accordingly guide the online updating of the spatial regularization map. Our DRCF differs from [35] and utilizes the saliency information in another way to efficiently construct the dual regularizers. A stronger penalization on the irrelevant background regions is hence achieved.

III. Proposed Tracking Approach

In this section, a novel dual regularized correlation filter tracker with dynamic saliency-based regularizers (DRCF) is proposed. Successively, the solution of the tracker is developed via the ADMM method with high efficiency. The main workflow of DRCF is illustrated in Fig. 2.

A. Overall Objective

Denoting the predefined Gaussian shape response $\mathbf{y}$, a sample set across D channels $\mathbf{x} = \{\mathbf{x}_d\}_{d=1:D}$ and corresponding discriminative filters $\mathbf{w} = \{\mathbf{w}_d\}_{d=1:D}$, where $\mathbf{x}_d$, $\mathbf{w}_d$, and $\mathbf{y} \in \mathbb{R}^{M\times N}$, the proposed DRCF is obtained by minimizing the following objective:

$$\arg\min_{\mathbf{w}} \frac{1}{2}\left\|\sum_{d=1}^{D}\mathbf{x}_d \star (\mathbf{s}_1 \odot \mathbf{w}_d) - \mathbf{y}\right\|_F^2 + \frac{1}{2}\sum_{d=1}^{D}\|\mathbf{s}_2 \odot \mathbf{w}_d\|_F^2, \quad (1)$$

where $\star$ stands for the circular correlation operation, $\odot$ denotes elementwise multiplication, and $\mathbf{s}_1, \mathbf{s}_2 \in \mathbb{R}^{M\times N}$ are the dynamic dual regularizers.

Remark 1: In DRCF, we develop $\mathbf{s}_1$ and $\mathbf{s}_2$ to be the dynamical spatial regularizers based on an efficient saliency detection technique. Compared with its predecessor SRDCF, where $\mathbf{s}_1 = \mathbb{1}$ (matrix of ones) and $\mathbf{s}_2$ is a predefined static regularizer, DRCF allows the regression optimization to dynamically focus more on the information from the object region and suppressing the boundary effect directly in the correlation operation. As shown in Fig. 2, the filter in DRCF is less affected by the background noise and can achieve a better response when detecting the object.

By introducing the constraint $\mathbf{g} = \mathbf{s}_1 \odot \mathbf{w}$, the Lagrangian multiplier $\mathbf{r}$, and the corresponding step size parameter $\rho \in \mathbb{R}$, (1) takes the augmented Lagrangian

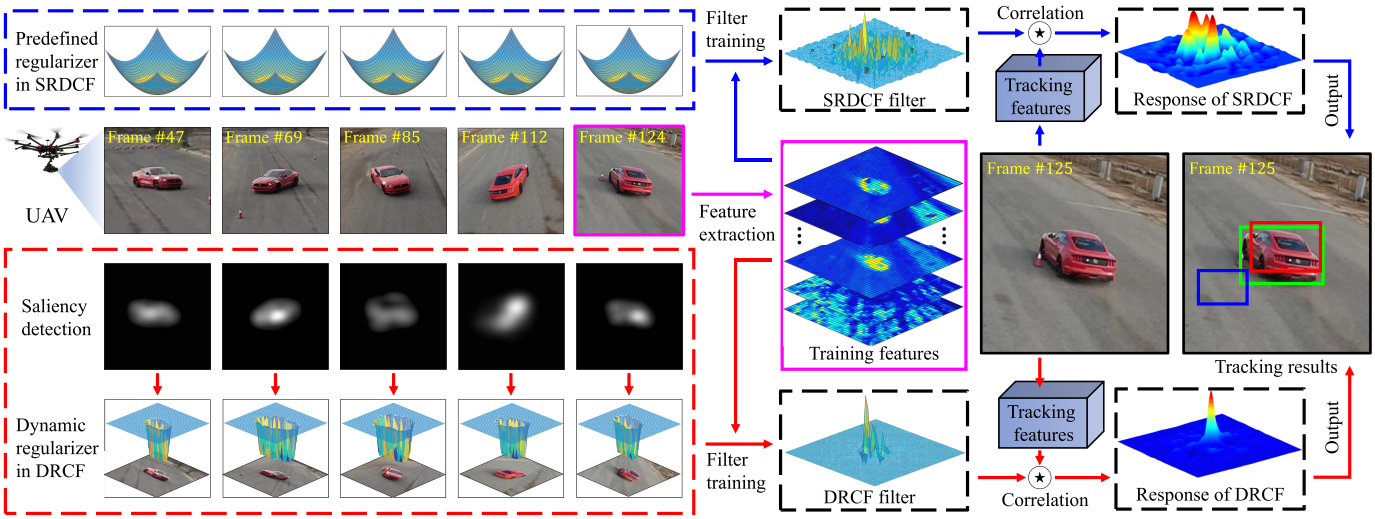

Fig. 2. Comparison between the tracking pipeline of the baseline SRDCF tracker and the proposed DRCF tracker. The baseline tracker follows the procedure denoted by the blue arrow to learn the filter while the proposed tracker follows the red one. In the tracking results, the green, red, and blue bounding box are, respectively, from the ground truth, the proposed tracker, and the baseline tracker. Notice that the proposed saliency-based regularizer in DRCF is capable of accentuating the appearance of the object while suppressing irrelevant background noise in the filter. This has notably enhanced the discriminative power of the filter and prevented the DRCF tracker from losing the object.

form of:

$$\mathcal{L}(\mathbf{w},\mathbf{g},\mathbf{r}) = \frac{1}{2}\left\|\sum_{d=1}^{D}\mathbf{x}_d \star \mathbf{g}_d - \mathbf{y}\right\|_F^2 + \frac{1}{2}\sum_{d=1}^{D}\|\mathbf{s}_2 \odot \mathbf{w}_d\|_F^2 + \sum_{d=1}^{D}\mathrm{trace}\left[(\mathbf{g}_d - \mathbf{s}_1 \odot \mathbf{w}_d)^\top \mathbf{r}_d\right] + \frac{\rho}{2}\sum_{d=1}^{D}\|\mathbf{g}_d - \mathbf{s}_1 \odot \mathbf{w}_d\|_F^2. \quad (2)$$

By substituting $\mathbf{h} = 1/\rho\mathbf{r}$, (2) can be reformulated as:

$$\mathcal{L}(\mathbf{w},\mathbf{g},\mathbf{r}) = \frac{1}{2}\left\|\sum_{d=1}^{D}\mathbf{x}_d \star \mathbf{g}_d - \mathbf{y}\right\|_F^2 + \frac{1}{2}\sum_{d=1}^{D}\|\mathbf{s}_2 \odot \mathbf{w}_d\|_F^2 + \frac{\rho}{2}\sum_{d=1}^{D}\|\mathbf{g}_d - \mathbf{s}_1 \odot \mathbf{w}_d + \mathbf{h}_d\|_F^2 - \frac{\rho}{2}\sum_{d=1}^{D}\|\mathbf{h}_d\|_F^2. \quad (3)$$

By adopting the ADMM method, (3) can be dissected and solved by iteratively solving the following three subproblems:

$$\begin{cases} \mathbf{g}^{k+1} = \arg\min_{\mathbf{g}} \left\{ \frac{1}{2}\left\|\sum_{d=1}^{D}\mathbf{x}_d \star \mathbf{g}_d - \mathbf{y}\right\|_F^2 + \frac{\rho}{2}\left\|\mathbf{g} - \mathbf{s}_1 \odot \mathbf{w}^k + \mathbf{h}^k\right\|_F^2 \right\} \\ \mathbf{w}^{k+1} = \arg\min_{\mathbf{w}} \left\{ \frac{1}{2}\|\mathbf{s}_2 \odot \mathbf{w}\|_F^2 + \frac{\rho}{2}\left\|\mathbf{g}^{k+1} - \mathbf{s}_1 \odot \mathbf{w} + \mathbf{h}^k\right\|_F^2 \right\} \\ \mathbf{h}^{k+1} = \mathbf{h}^k + \mathbf{g}^{k+1} - \mathbf{s}_1 \odot \mathbf{w}^{k+1} \end{cases}, \quad (4)$$

where the superscript k denotes the number of iterations.

As shown in the following derivation, each of the subproblems in (4) has an analytic solution.

1) Subproblem $\mathbf{g}$*:* Using Parseval's theorem and the convolution theorem, the first subproblem in (4) can be transferred into the Fourier domain:

$$\hat{\mathbf{g}}^{k+1} = \arg\min_{\hat{\mathbf{g}}} \left\{ \frac{1}{2}\left\|\sum_{d=1}^{D}\hat{\mathbf{x}}_d^* \odot \hat{\mathbf{g}}_d - \hat{\mathbf{y}}\right\|_F^2 + \frac{\rho}{2}\left\|\hat{\mathbf{g}} - \widehat{(\mathbf{s}_1 \odot \mathbf{w})} + \hat{\mathbf{h}}\right\|_F^2 \right\}. \quad (5)$$

Remark 2: Superscript k is omitted for a clearer interpretation, $*$ stands for complex conjugate, and $\hat{}$ denotes the discrete Fourier transform (DFT) of a matrix, $\hat{\mathbf{a}} = F \cdot \mathbf{a}$.

Based on the elementwise dot product $\hat{\mathbf{x}}_d^* \odot \hat{\mathbf{g}}_d$ in (5), the ith row and the jth column element of the label $\hat{\mathbf{y}}$, i.e., $\hat{\mathbf{y}}(i,j)$, only depend on the corresponding $\hat{\mathbf{g}}_d(i,j)$ and the $\hat{\mathbf{x}}_d^*(i,j)$ across all D channels. Denoting that $\mathbf{e}_{ij}(\hat{\mathbf{x}}) \in \mathbb{C}^{D\times 1}$ concatenating the element of $\hat{\mathbf{x}}_d(i,j)$ across all D channels, (5) can be split into $M \times N$ independent subproblems, such that each of them takes the form of:

$$\arg\min_{\mathbf{e}_{ij}(\hat{\mathbf{g}})} \left\{ \frac{1}{2}\left\|\mathbf{e}_{ij}(\hat{\mathbf{x}})^H \mathbf{e}_{ij}(\hat{\mathbf{g}}) - \hat{\mathbf{y}}_{ij}\right\|_F^2 + \frac{\rho}{2}\left\|\mathbf{e}_{ij}(\hat{\mathbf{g}}) - \mathbf{e}_{ij}\left(\widehat{\mathbf{s}_1 \odot \mathbf{w}}\right) + \mathbf{e}_{ij}(\hat{\mathbf{h}})\right\|_F^2 \right\}. \quad (6)$$

Setting the derivative of $\mathbf{e}_{ij}(\hat{\mathbf{g}})$ to zero and denoting the identity matrix as $\mathbf{I}$, the solution of (6) is obtained as:

$$\begin{cases} \mathbf{e}_{ij}(\hat{\mathbf{g}}) = \frac{1}{\rho}\left[\mathbf{I} - \frac{\mathbf{e}_{ij}(\hat{\mathbf{x}})^* \mathbf{e}_{ij}(\hat{\mathbf{x}})^H}{\rho + \mathbf{e}_{ij}(\hat{\mathbf{x}})^H \mathbf{e}_{ij}(\hat{\mathbf{x}})^*}\right]\mathbf{p} \\ \mathbf{p} = \rho\left[\mathbf{e}_{ij}\left(\widehat{\mathbf{s}_1 \odot \mathbf{w}}\right) - \mathbf{e}_{ij}(\hat{\mathbf{h}})\right] + \mathbf{e}_{ij}(\hat{\mathbf{x}})^* \hat{\mathbf{y}}_{ij}. \end{cases} \quad (7)$$

Remark 3: Calculating (7) only takes vector multiply–add operations and is thus time efficient in the iterations.

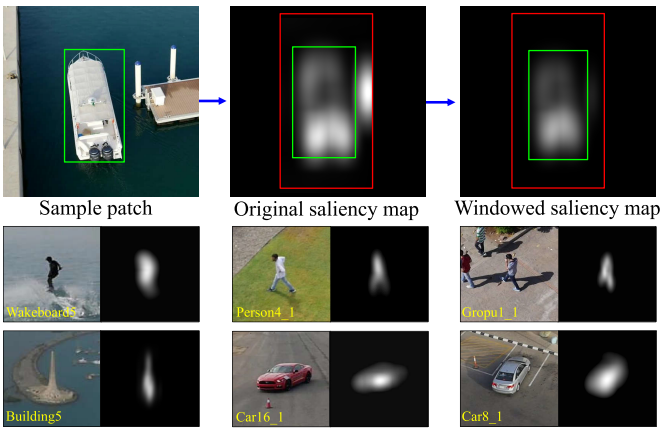

Fig. 3. Visualization of the object saliency awareness in the DRCF tracker. The first row shows the procedure of extracting the saliency awareness. Note that the misleading saliency is suppressed after windowing and the accentuation ability of the saliency map is thus enhanced. Consequently, the remapped regularizer can regulate the correct region responding to the object. Rows two and three show six more challenge saliency detection results.

A detailed derivation using the Sherman Morrison formula [36] is presented in Appendix V.

2) Subproblem $\mathbf{w}$*:* The second subproblem in (4) can be directly solved in the time domain by setting the derivative of $\mathbf{w}$ to zero, such that:

$$\frac{\partial}{\partial \mathbf{w}}\left\{\frac{1}{2}\|\mathbf{s}_2 \odot \mathbf{w}\|_F^2 + \frac{\rho}{2}\|\mathbf{g} - \mathbf{s}_1 \odot \mathbf{w} + \mathbf{h}\|_F^2\right\} = 0. \tag{8}$$

Solving (8) leads to the solution of $\mathbf{w}$ in each iteration, such that

$$\mathbf{w} = \frac{\rho \cdot \mathbf{s}_1 \odot (\mathbf{g} + \mathbf{h})}{\rho(\mathbf{s}_1 \odot \mathbf{s}_1) + (\mathbf{s}_2 \odot \mathbf{s}_2)}\ . \tag{9}$$

Remark 4: The aforementioned division is performed efficiently in an elementwise manner.

3) Iteration Update: In each iteration, $\mathbf{h}$ is updated using the last line of (4). The step size parameter ρ is also updated with the scale β following the scheme from [36]:

$$\rho^{k+1} = \min\left(\rho_{\max}, \beta\rho^k\right)\ . \tag{10}$$

B. Saliency-Based Dynamical Regularization

As mention previously, the proposed DRCF tracker attempts to generate the regularizers $\mathbf{s}_1$ and $\mathbf{s}_2$ that can dynamically accentuate the irregular shape of the object. While having the same purpose to select the object and penalize the background, two regularizers are generated using the same procedure. Starting from the estimated object bounding box, i.e., the green box in Fig. 3, a concentric region λ times larger than the bounding box is first cropped as the saliency detection area, i.e., the red box in Fig. 3. Next, saliency detection is performed in the cropped region with an existing algorithm [37] to obtain the original saliency map efficiently. After multiplying by a cos window, the original map is then resized to the corresponding size in the appearance model $\mathbf{x}$ used in filter training, and its coefficients are remapped to the regulation weights by a threshold ϵ. Finally, the remapped regulation map $\mathbf{s}_f$ is intersected with another predefined regularizer $\mathbf{s}_r$ penalizing all of the region outside of the object bounding box to obtain the final regularizer: $\mathbf{s}_i = \mathbf{s}_f \cap \mathbf{s}_r$, where $i = 1, 2$.

By intersection, the background inside the object bounding box is penalized while the saliency outside of the box is abandoned to avoid overfitting and misleading the tracker.

Remark 5: Compared with the static regularizer adopted by Danelljan *et al.* [9] and the dynamic iterative extension by Dai *et al.* [31], we have introduced a dynamic method for an efficient regularizer generation free from a heavy iterative computation load. Apart from the conciseness, the experiments described in Section IV also verify its effectiveness in highlighting the object corresponding region and boosting the performance of the tracker.

C. Tracking Pipeline

Summarized in Algorithm 1, the proposed tracker follows a simple tracking pipeline.

Algorithm 1 DRCF Tracker

Input: The image frame t, the object position $\mathbf{p}_{t-1}$, and the scale $\mathbf{l}_{t-1}$ on the previous frame $t-1$, the appearance model $\mathbf{x}_{t-1}$, the regularizer $\mathbf{s}_1$, $\mathbf{s}_2$, and the filter $\mathbf{w}_{t-1}$.

Output: The predicted position $\mathbf{p}_t$ and the scale $\mathbf{l}_t$ at frame t.

for $t = 2$ *to end* **do**
- Crop the image patch on the frame t centered around the object location $\mathbf{p}_{t-1}$ and extract features $\mathbf{x}$.
- Search for the highest value in $\dot{\mathbf{y}}$ in (13) as the object position estimation $\mathbf{p}_t$.
- Use the scale-space correlation filter from [23] to estimate the scale $\mathbf{l}_t$.
- Generate the new regularizers $\mathbf{s}_1$, $\mathbf{s}_2$ and abandon the old ones (Section III-B).
- Extract features $\mathbf{x}_t$ at $\mathbf{p}_t$ and $\mathbf{l}_t$ and update the appearance model to $\mathbf{x}_t$ using (11).
- Solve the new filter using the appearance model $\mathbf{x}_t$ and the regularizers $\mathbf{s}_1$, $\mathbf{s}_2$ (Section III-A).
- Update the filter to $\mathbf{w}_t$ using (12).

end

Training: In a frame with the appearance model updated by features extracted from the estimated object location, the DRCF tracker uses the method introduced in Section III-B to obtain the saliency-based regularizer and follows the optimization procedure in Section III-A to obtain the dual regularized filter.

Detection: As each new frame arrives, the new object position is assigned to the location of maximum response in (13) and the new scale of the object is estimated by another single scale-space correlation filter from [23].

Model Update: The appearance model $\mathbf{x}_t$ and the filter $\mathbf{w}_t$ are updated using a linear interpolation method proposed in [8] with a learning rate η to balance the memory and the

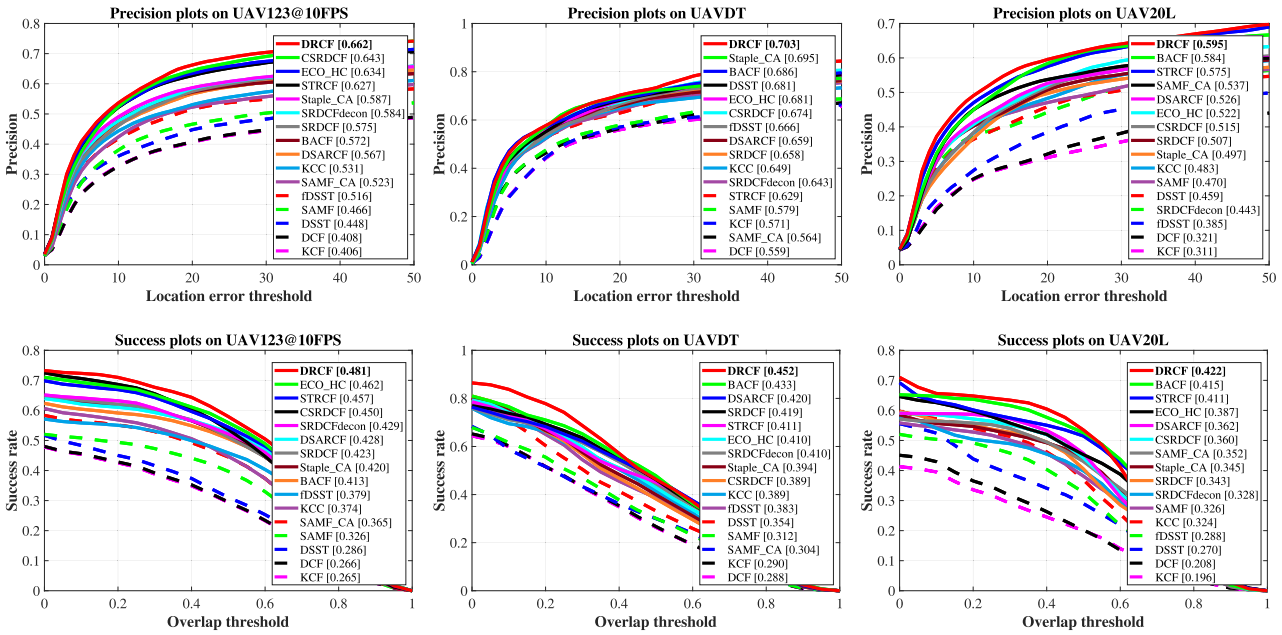

Fig. 4. Tracking results of DRCF and other 15 trackers based on handcrafted features. (From left to right) First, second, and third columns show the results on the UAV123@10FPS, the UAVDT, and the UAV20L benchmarks, respectively. First row: PP. Second row: SP. For all benchmarks, the threshold of 20 pixels in precision and the AUC of the success rate are given in the brackets to rank the trackers. It can be seen that DRCF ranks first place in the three benchmarks in terms of both precision and success rate.

adaptiveness of the tracker:

$$\mathbf{x}_t = (1-\eta)\mathbf{x}_{t-1} + \eta\mathbf{x}_\mathbf{t} \tag{11}$$

$$\mathbf{w}_t = (1-\eta)\mathbf{w}_{t-1} + \eta\mathbf{w}_\mathbf{t}. \tag{12}$$

Remark 6: Notice that the saliency-based regularizers are not memorized but newly computed at each frame to focus on the current object appearance and avoid overfitting. In addition, the detection response $\hat{\mathbf{y}}$ is computed efficiently in the Fourier domain as:

$$\hat{\mathbf{y}} = F^{-1}\left\{\sum_{d=1}^{D}\hat{\mathbf{x}}_d \odot \hat{\mathbf{w}}_d\right\}. \tag{13}$$

IV. Experiments

This section presents a comprehensive experimental evaluation of the proposed DRCF tracker. Experiments were conducted on all 193 challenging aerial image sequences (including 134 639 frames) from three well-known standard benchmarks specifically designed for aerial tracking, i.e., UAVDT [11], UAV123@10FPS [38], and UAV20L [38]. The sequences in the benchmarks were captured on UAV platforms.

A. Implementation Details

Following the basic settings in SRDCF [9], the search region of the proposed DRCF tracker is set to be a square with a size of $(5\mathbf{XY})^{1/2}$, where $\mathbf{X}$ and $\mathbf{Y}$ are respectively the height and width of the object. Next, the gray-scale, HOG [18] and CN [17] features, i.e., handcrafted features, are used to represent the object. The cell size of these features is 4×4 pixels. For the saliency-based regularizer construction, the scale $\lambda = 2$ and the remap threshold $\epsilon = 0.1$. For ADMM iterations, the initial value of $\mathbf{g}$, $\mathbf{w}$, and $\mathbf{h}$ are all null matrices. The number of iterations, the initial step size ρ, its maximum $\rho_{\max}$, and the update scale β are 2, 1, 0.1, and 10, respectively. The learning rate used for the model update is set to be $\eta = 0.0193$.

Remark 7: For a fair comparison, all the parameters of the proposed tracker are left unchanged in all experiments. Trackers in the experiments are implemented in MATLAB 2018a on the same PC with an i7-8700k processor (3.70 GHZ), 32 GB RAM, and an NVIDIA RTX 2080 GPU. For more specific details, our tracker is publicly available at: https://github.com/vision4robotics/DRCF-Tracker.

B. Evaluation Criteria

Consistent with the standard evaluation criteria in the benchmarks [11], [38] and adopting the one-pass evaluation protocol, the tracking performance of the trackers is evaluated by two measures quantitatively, i.e., precision and success rate. In each frame, the tracking precision is determined by the center location error (CLE) between the tracking result $\mathbf{R}_T$ and the manually annotated ground truth bounding box $\mathbf{R}_G$. On the other hand, tracking success rate is measured by their overlap size and is defined as the intersection over union (IoU), such that:

$$\text{IoU} = \frac{|\mathbf{R}_T \cap \mathbf{R}_G|}{|\mathbf{R}_T \cup \mathbf{R}_G|}. \tag{14}$$

In this case, the precision plot (PP) and the success plot (SP) show the percentage of frames in the test set whose CLE or IoU is within a maximum allowed threshold. As commonly

TABLE I

TRACKING SUCCESS RATE OF THE TRACKERS USING HANDCRAFTED FEATURES AT DIFFERENT ATTRIBUTES ON THE UAVDT BENCHMARK. THE TOP THREE RANKING TRACKERS IN EACH ATTRIBUTE ARE MARKED IN RED, GREEN, AND BLUE. THE RUNNING SPEED OF EACH TRACKER IN FPS IS ALSO PRESENTED AND MARKED WITH THE SAME COLORING POLICY. AMONG THE NINE ATTRIBUTES, THE PROPOSED DRCF TRACKER RANKS FIRST IN EIGHT ATTRIBUTES, WHILE ACHIEVING THE SECOND-BEST IN THE REMAINING IV CATEGORY. THE CONSISTENT OUTSTANDING PERFORMANCE OF THE DRCF TRACKER IN THESE DIFFERENT TRACKING CONDITIONS VERIFIES ITS ROBUSTNESS

Trackers	Venue	FPS	BC	CM	IV	LO	LTT	OB	OM	SV	SO	Overall
DSARCF [35]	19' TIP	12.7	0.341	**0.386**	0.430	0.325	0.500	0.405	0.350	0.377	**0.444**	**0.420**
CSRDCF [39]	18' CVPR	13.2	0.345	0.347	0.398	**0.352**	0.487	0.373	0.342	0.366	0.355	0.389
STRCF [39]	18' CVPR	32.3	0.340	0.365	0.421	0.319	0.525	**0.406**	0.341	0.388	0.421	0.411
KCC [16]	18' AAAI	56.9	0.332	0.351	0.431	0.304	0.480	0.398	0.322	0.340	0.390	0.389
ECO_HC [29]	17' CVPR	79.2	**0.364**	0.379	0.434	**0.347**	**0.573**	0.391	0.358	0.389	0.375	0.410
BACF [36]	17' CVPR	69.1	**0.367**	**0.387**	**0.460**	0.340	**0.581**	**0.443**	**0.371**	**0.408**	**0.428**	**0.433**
SAMF_CA [40]	17' CVPR	14.7	0.269	0.276	0.301	0.259	0.393	0.271	0.229	0.255	0.296	0.304
Staple_CA [40]	17' CVPR	64.1	0.326	0.349	0.428	0.324	0.539	0.405	0.344	0.366	0.379	0.394
fDSST [42]	16' TPAMI	**219**	0.315	0.363	0.395	0.332	0.518	0.369	0.312	0.337	0.380	0.383
SRDCFdecon [41]	16' CVPR	8.9	0.339	0.374	0.429	0.321	0.515	0.395	0.351	0.389	0.410	0.410
SRDCF [9]	15' ICCV	17.4	0.356	0.385	**0.443**	0.333	0.524	**0.406**	**0.361**	**0.399**	0.416	0.419
KCF [8]	15' TPAMI	**957**	0.235	0.267	0.312	0.229	0.312	0.298	0.244	0.254	0.251	0.290
DCF [8]	15' TPAMI	**1454**	0.236	0.261	0.308	0.232	0.289	0.292	0.243	0.249	0.252	0.288
SAMF [22]	14' ECCV	15.8	0.268	0.283	0.315	0.256	0.340	0.297	0.257	0.264	0.290	0.312
DSST [23]	14' BMVC	148	0.304	0.329	0.379	0.299	0.408	0.357	0.288	0.296	0.360	0.354
DRCF	Ours	38.4	**0.407**	**0.427**	**0.448**	**0.403**	**0.587**	**0.448**	**0.410**	**0.409**	**0.462**	**0.452**

used in tracker evaluations, the CLE at 20 pixels on PP and the area-under-the-cure (AUC) on SP are used to rank the trackers in precision and success rate, respectively.

C. Overall Performance Evaluation

1) Quantitative and Qualitative Evaluation: The proposed DRCF is analyzed together with other 15 state-of-the-art trackers using handcrafted features, i.e., DSARCF [35], ECO-HC [29], CSRDCF [39], STRCF [39], Staple_CA [40], SRDCF [9], SRDCFdecon [41], BACF [36], KCC [16], SAMF_CA [40], fDSST [42], SAMF [22], DSST [23], KCF [8], and DCF [8]. The experiments are thoroughly conducted on three aerial object tracking benchmarks, i.e., UAVDT [11], UAV123@10FPS [38], and UAV20L [38], with all their sequences for comprehensive evaluations.

As shown in Fig. 4, the proposed DRCF tracker has outperformed other compared trackers based on handcrafted features for all three benchmarks. More specifically, on the UAVDT benchmark, DRCF achieves top one precision (0.703) and has an advantage of 1.2% and 2.5% over the second-best and the third-best trackers, i.e., Staple_CA (0.695) and BACF (0.686), respectively. Meanwhile, with a success rate of 0.452, DRCF also achieves an advantage of 4.4% and 7.6% over the second-best tracker BACF (0.433) and the third-best tracker DSARCF (0.420), respectively. On the UAV123@10FPS benchmark, DRCF (0.662, 0.481) also achieves the best performance consistently, followed by the second-best tracker CSRDCF (0.643, 0.450) in precision and the second-best tracker ECO-HC (0.634, 0.462) in success rate, respectively. On the UAV20L benchmark, DRCF (0.595, 0.442) also outperforms the compared trackers, followed by the second-best tracker BACF (0.584, 0.415), and the third-best STRCF (0.575, 0.411). For more intuitive evaluations, Fig. 5 shows some tracking examples on these benchmarks.

Remark 8: According to the establishers of the benchmark, the UAV123@10FPS benchmark is temporally downsampled from the videos recorded in 30–10 FPS and thus poses more challenges on fast motion sequences. The UAVDT benchmark considered the performance impact caused by different flying altitude and purposed small object (SO) sequences recorded in high-altitude. The UAV20L benchmark features long-term tracking (LTT) sequences, as the camera is usually required to follow the object for a relatively long period in aerial tracking conditions. These benchmarks represent the real-world aerial object tracking scenarios and validate the prominent performance of DRCF compared to the other trackers.

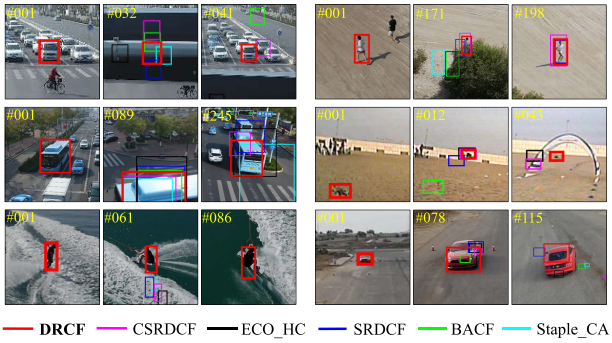

Fig. 5. Visualization of the tracking results from the top-ranking trackers and the proposed DRCF. Trackers are compared on S0601 and S0602 sequences from the UAVDT benchmark, on wakeboard2, group2_2, and UAV3 sequences from the UAV123@10FPS benchmark, as well as on car16 sequence from the UAV20L benchmark. More examples are presented online at: https://youtu.be/XEl7LcjauA8.

2) Attribute-Based Comparison: To denote the tracking challenges in the image sequences for more detailed analysis, the UAVDT benchmark has categorized them into nine different attributes: background clutter (BC), CM, illumination variation (IV), large occlusion (LO), object blur (OB), object motion (OM), scale variations (SVs), SO, and LTT. Experimental results of the attribute-based performance comparisons are presented in Table I and Fig. 6.

Generally, the proposed DRCF tracker performs very competitively among the challenging attributes compared with

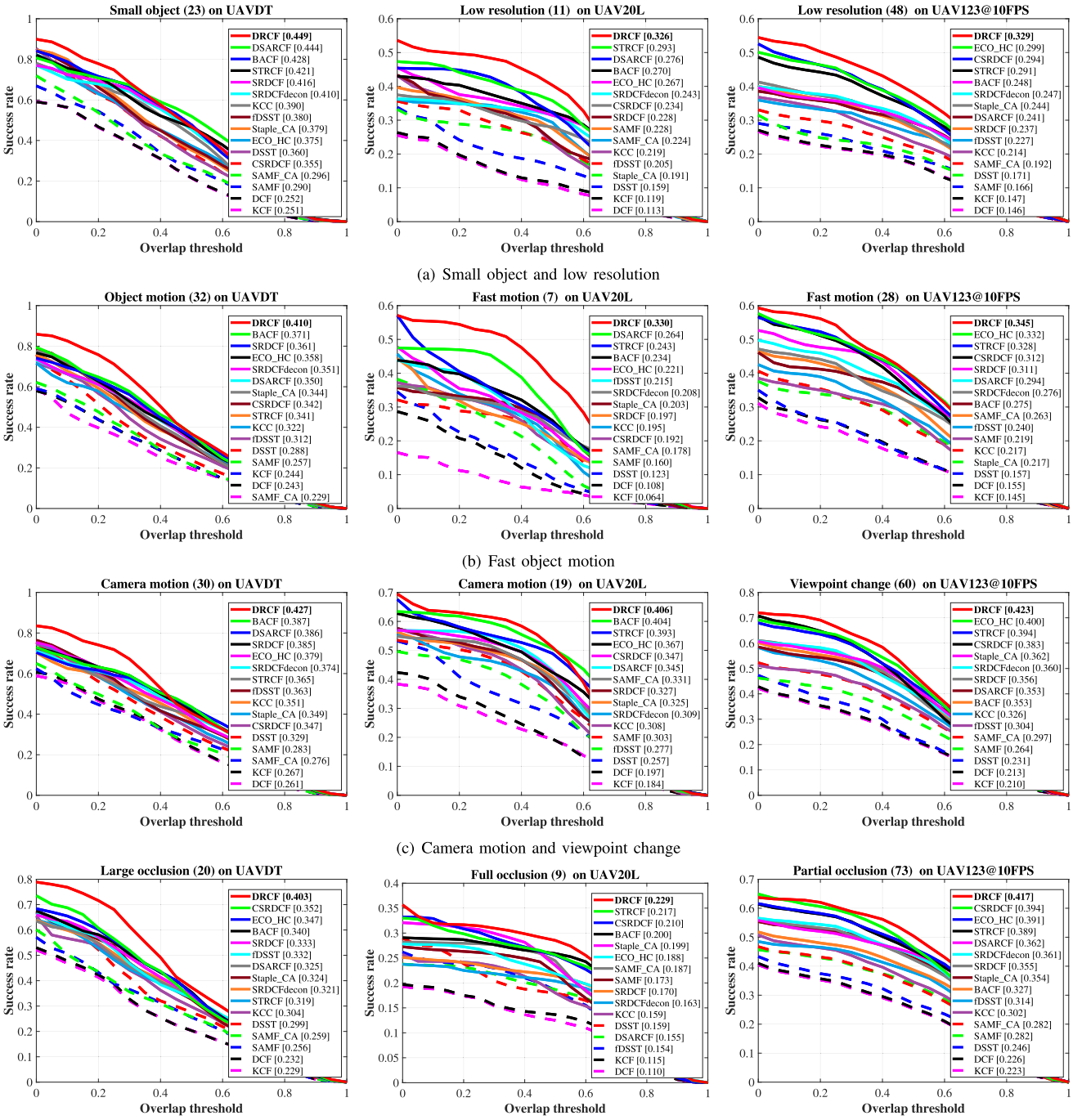

Fig. 6. Attribute-based comparison of DRCF and 15 trackers based on handcrafted features. Similar attributes from the UAVDT, UAV20L, and UAV123@10FPS benchmarks are categorized in the same row.(a) SO and low resolution. (b) Fast object motion. (c) CM and viewpoint change.(d) Occlusion. The precision at the threshold of 20 pixels and the AUC of success rate are given in brackets to rank the trackers. The competitive performance is consistent in these attributes across the benchmarks.

other state-of-the-art trackers. Among the nine attributes in the UAVDT benchmark, DRCF ranks first among the eight attributes, i.e., BC, CM, LO, LTT, OB, OM, SV, and SO. As shown in Table I, DRCF performs 10.9%, 10.3%, 14.5%, 1.0%, 1.1%, 10.5%, 0.2%, and 4.1% better in these attributes compared with the second-best tracker in the success rate. In addition to the UAVDT benchmark, results on similar attributes from UAV123@10FPS and the UAV20L benchmarks are also presented in Fig. 6 to verify the proposed DRCF tracker more comprehensively. As shown in Fig. 6, the performance of DRCF in these attributes is consistently competitive across multiple benchmarks. The credit is given to the dual saliency-based dynamical regularizers, which help the tracker focus on the object during its fast motion and dynamically suppress the irrelevant background in the case of occlusion.

Remark 9: The attribute-based comparison verifies the robustness and well-roundness of the DRCF tracker under

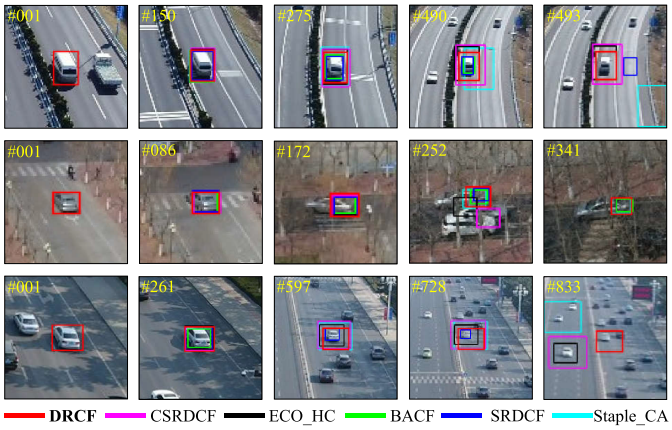

Fig. 7. Visualization of tracking results on SO sequences from the UAVDT benchmark. From top to bottom are the *S*0401, the *S*1001, and the *S*1602 SO sequences. The object size in these sequences can be less than 10×10 pixels, and the size ratio between the object and the image is smaller than $1/10\,000$. Experimental results verify the competitive performance of DRCF in SO and low-resolution tracking condition compared to other state-of-the-art trackers.

different tracking challenges. Especially, as shown in Fig. 6, DRCF demonstrates the best tracking performance in SO and low-resolution tracking sequences. In this manner, DRCF has been proved to be competent in geoscience-related aerial visual tracking, where the object may take up only a tiny portion of the image depending on its distance from the tracking platform. This promising performance is visualized in Fig. 7 with the tracking results on sequences with low-resolution conditions and SO attributes.

D. Ablation Study

In this section, an ablative study is conducted to evaluate the effectiveness of the modules in the proposed DRCF tracker more specifically. As the baseline, SRDCF [9] is considered as a special case of DRCF with $\mathbf{s}_1 = \mathbb{1}$ (matrix of ones) and $\mathbf{s}_2$ being a predefined static regularizer. Then, the dual regularization strategy is called up and its contribution is assessed in SRDCF+dr with $\mathbf{s}_1$ and $\mathbf{s}_2$, both being predefined static regularizers. Separately, the saliency module is activated in SRDCF+sa with $\mathbf{s}_1 = \mathbb{1}$ and $\mathbf{s}_2$ being a saliency-based dynamical regularizer to evaluate the impact of the saliency awareness. Finally, the benefits of these two modules are combined in the complete DRCF to achieve maximum performance.

In Table II, by comparing SRDCF+dr with the baseline, the effectiveness of the proposed dual regulation strategy results in a tracking success rate enhancement of 3.7%, 8.7%, and 8.5% on the UAVDT, UAV123@10FPS, and UAV20L benchmarks, respectively. It can be thus concluded that the presented dual regularization strategy can improve the overall tracking performance by directly regulating the filter involved with the correlation operation in the main regression. On the other hand, compared with the baseline, the saliency awareness represented by SRDCF+sa results in an improvement in tracking success rate of 4.5%, 11.3%, and 16.6% on the UAVDT, UAV123@10FPS, and UAV20L benchmarks, respectively. This indicates that the proposed saliency-aware regularizer can identify the object more precisely and thus can enhance the discriminative power as previously visualized in Fig. 2. Finally, by combining the two modules, the complete DRCF can achieve the highest performance.

TABLE II

RESULTS OF ABLATION STUDIES OF THE PROPOSED DRCF TRACKER ON THE UAVDT, UAV123@10FPS AND UAV20L BENCHMARKS. THE EFFECTIVENESS OF THE TWO PROPOSED MODULES IS VERIFIED BY THE PERFORMANCE CHANGE RELATIVE TO THE BASELINE SRDCF TRACKER. THE COMPLETE DRCF TRACKER COMBINES THEIR BENEFITS TO MAXIMIZE THE PERFORMANCE

Tracker	Regularizer		Success rate		
	$\mathbf{s}_1$	$\mathbf{s}_2$	UAVDT	UAV123@10FPS	UAV20L
SRDCF	$\mathbb{1}$	static	0.419	0.423	0.343
SRDCF+dr	static	static	0.435	0.460	0.372
SRDCF+sa	$\mathbb{1}$	dynamic	0.438	0.471	0.400
DRCF	dynamic	dynamic	0.452	0.481	0.422

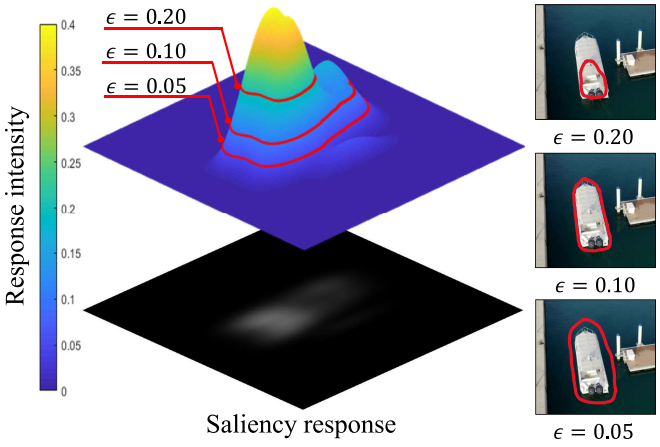

Fig. 8. Visualization of the regularized region selected by different remap threshold ϵ. (Left) Saliency detection response and its intensity. (Right) Region outlined in red is the region selected by the corresponding threshold ϵ. Note that the results are drawn on the image patches for a clear demonstration.

E. Key Parameter Analysis

As shown in Fig. 8, the accuracy of the proposed dynamical saliency-aware regularizers is controlled by the remap threshold ϵ. Based on the windowed saliency detection response map in Section III-B, the region whose response intensity is greater than ϵ would be considered as the object and remapped in the regularizers. Thus, when ϵ is too small, the regularizers after remapping would circumscribe the object outside of its contour and bring in the irrelevant background, as shown in Fig. 8 when $\epsilon = 0.05$. However, when too large, the regularized region would shrink inside the boundary of the object and miss the outer part as well, as shown in Fig. 8 when $\epsilon = 0.20$.

In Fig. 9, the impact of ϵ is quantitatively analyzed with the overall tracking performance on the UAVDT benchmark. From Fig. 9, the overall tracking performance of the tracker gradually increases as the remap threshold ϵ begins to rise from a small value. At this stage, the regularizers converge to the object from the outside and the background noise would be more correctly suppressed. After reaching the peak

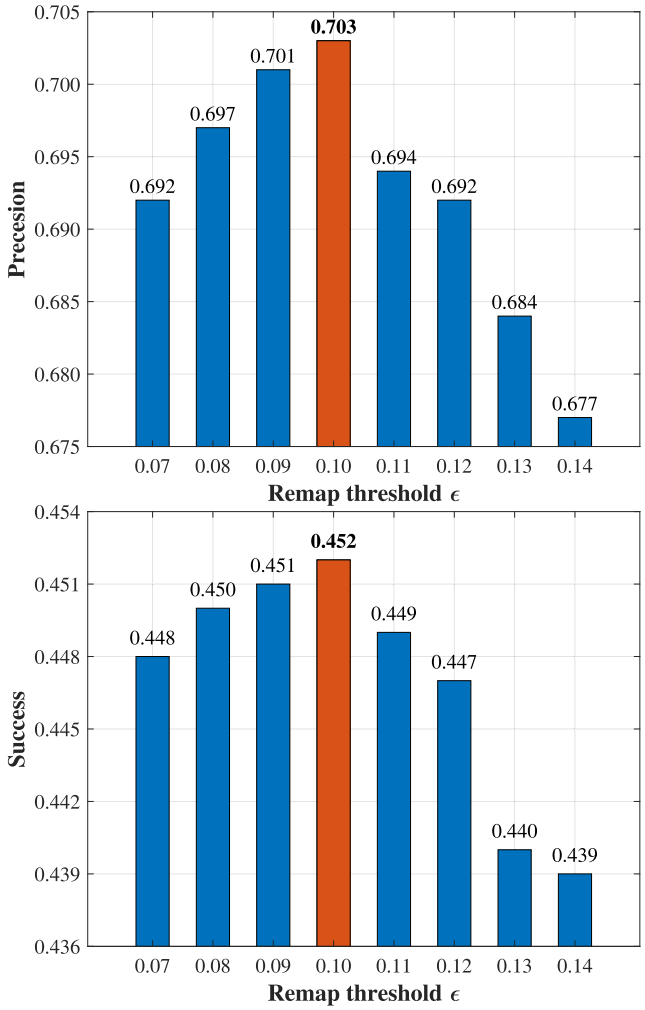

Fig. 9. Parameter analysis of the remap threshold ϵ with overall quantitative results from all the sequences in the UAVDT benchmark. With other parameters remaining fixed, the tracking precision (Top) and success rate (Bottom) are plotted in bar charts. As shown, the tracking performance generally increases as ϵ increases, and then, after reaching the peak performance at $\epsilon = 0.10$, both the tracking precision and the success rate decline as ϵ further grows.

performance at $\epsilon = 0.10$, the regularizers begin to contract inside the object and the performance tends to decrease as a result. Thus, $\epsilon = 0.10$ is considered to be the most suitable parameter to crop the region of the tracked object. For the best performance, this configuration is set fixed for all benchmark evaluations.

F. Comparison With Deep-Based Trackers

For a more comprehensive evaluation, the proposed DRCF is additionally compared with another 12 state-of-the-art deep-based trackers on the UAVDT benchmark. These trackers can be categorized in two groups, eight DCF-based trackers using convolutional features (ASRCF [31], CoKCF [15], IBCCF [24], CF2 [26], C-COT [28], ECO [29], DeepSTRCF [39], and MCPF [43]) and four end-to-end trackers based on neural networks (TADT [44], UDT+ [45], UDT [45], and ADNet [46]).

With the results presented in Table III, the tracking precision of the DRCF tracker (0.703) is 0.43% better than the second-best tracker ECO (0.700). On the other hand, the ECO tracker has the highest tracking success rate of 0.454 and the DRCF (0.452) ranks second with a performance difference of 0.44%. The close performance margin between the purposed DRCF tracker and the state-of-the-art trackers verifies its competitiveness. In addition, Table III presents the tracking speed of the trackers in terms of FPS and evaluates the tracking efficiency. In the top three ranking trackers according to tracking efficiency, UDT runs with the best tracking speed (76 FPS on GPU), followed by UDT+ (60.4 FPS on GPU) and the proposed DRCF tracker (38 FPS on CPU).

TABLE III

PERFORMANCE COMPARISON WITH 12 STATE-OF-THE-ART DEEP-BASED TRACKERS ON THE UAVDT BENCHMARK. THE BEST THREE PERFORMANCES ARE RESPECTIVELY HIGHLIGHTED IN RED, GREEN, AND BLUE. NOTE THAT THE DEEPSTRCF TRACKER IS DENOTED BY *D.STRCF* IN THE TABLE FOR A CONCISE REPRESENTATION

Tracker	Venue	Pres.	Succ.	FPS	GPU
ASRCF [31]	19' CVPR	**0.700**	**0.437**	21.3	✓
TADT [44]	19' CVPR	0.677	0.431	32.5	✓
UDT [45]	19' CVPR	0.674	0.441	**76.4**	✓
UDT+ [45]	19' CVPR	**0.697**	0.416	**60.4**	✓
D.STRCF [39]	18' CVPR	0.667	**0.437**	6.6	✓
ECO [29]	17' CVPR	**0.700**	**0.454**	16.4	✓
IBCCF [24]	17' ICCV	0.603	0.388	3.4	✓
MCPF [43]	17' CVPR	0.660	0.403	0.7	✓
CoKCF [15]	17' PR	0.605	0.319	21.2	✓
ADNet [46]	17' CVPR	0.683	0.429	7.5	✓
CCOT [28]	16' ECCV	0.656	0.406	1.1	✓
CF2 [26]	15' ECCV	0.602	0.355	20.1	✓
DRCF	Ours	**0.703**	**0.452**	**38.4**	✗

Remark 10: It is worth pointing out that most aerial tracking platforms are not currently equipped with a GPU due to cost, weight, and power consumption considerations. In this regard, the proposed DRCF tracker relies only on a single CPU and is thus more suitable for real-world aerial tracking applications with both competitive performance and adequate efficiency.

G. Limitations and Future Work

Although the presented DRCF performs competitively against the 27 aforementioned state-of-the-art trackers in general, its performance is still not perfect for all tracking scenarios. In the cases of IV and SV in Table I, the proposed DRCF tracker did not exceed the other competitors with considerable margins.

When illumination varies greatly, the contour of the tracked object is no longer salient against the background and the light source would instead become dominant in the image. In this condition, the saliency-based regularizers in DRCF could be misled to penalize the incorrect region and the tracking performance is hence not optimal. For SV, the quickly changing scale of the object can also affect the accuracy of the saliency detection. Although optimized by the experiment in Section IV-E, the remap threshold ϵ is essentially fixed and can hence be less ideal for fast SVs in SV sequences. On this subject, future work would focus on developing more

self-adaptive methods to generate the regularizers and increase the tracker's power in sensing the object more accurately.

V. Conclusion

In this article, a novel DRCF tracker with saliency-based dynamic regularizers has been proposed for real-time aerial tracking and remote sensing applications. By adopting a dual regulation strategy specifically designed to alleviate the boundary effect in the regression correlation operation, the proposed DRCF outperforms its predecessor SRDCF in both tracking accuracy and robustness. Furthermore, based on the saliency detection method, the presented regularizers are capable of accentuating the object and suppressing background noise, which boosts the performance of the tracker again from competitive to satisfying. Finally, the proposed tracker has proven itself among state-of-the-art trackers in thorough and challenging experiments on multiple aerial tracking benchmarks. We strongly believe that our contribution can enhance the tracking accuracy, efficiency, and robustness in aerial visual object tracking.

Appendix
Derivation of the Subproblem $\mathbf{g}$

This section provides the complete solution process of the subproblem $\mathbf{e}_{ij}(\hat{\mathbf{g}})$ from (6) to (7).

Taking the derivative of $\mathbf{e}_{ij}(\hat{\mathbf{g}})$ to zero and denoting $\mathbf{I}$ for identity matrix, (6) becomes

$$\begin{cases} \mathbf{q} \cdot \mathbf{e}_{ij}(\hat{\mathbf{g}}) = \rho \left[\mathbf{e}_{ij}\left(\widehat{\mathbf{s}_1 \odot \mathbf{w}} \right) - \mathbf{e}_{ij}\left(\hat{\mathbf{h}} \right) \right] + \mathbf{e}_{ij}(\hat{\mathbf{x}})^* \hat{\mathbf{y}}_{ij} \\ \mathbf{q} = \mathbf{e}_{ij}(\hat{\mathbf{x}})^* \mathbf{e}_{ij}(\hat{\mathbf{x}})^H + \rho \mathbf{I}. \end{cases} \tag{15}$$

Solving $\mathbf{e}_{ij}(\hat{\mathbf{g}})$ from (15) requires computing the matrix inverse of $\mathbf{q}$. In order to reduce the computation burden, the Sherman Morrison Formula [36] is employed stating that

$$\left(\mathbf{u}\mathbf{v}^{\mathrm{H}} + \mathbf{A}\right)^{-1} = \mathbf{A}^{-1} - \left(1 + \mathbf{v}^{\mathrm{H}}\mathbf{A}^{-1}\mathbf{u}\right)^{-1}\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^{\mathrm{H}}\mathbf{A}^{-1}. \tag{16}$$

By substituting $\mathbf{A} = \rho\mathbf{I}$ and $\mathbf{u} = \mathbf{v} = \mathbf{e}_{ij}(\hat{\mathbf{x}})$ in our case, a closed-form solution of each subproblem $\mathbf{e}_{ij}(\hat{\mathbf{g}})$ can be obtained as stated in (7)

$$\begin{cases} \mathbf{e}_{ij}(\hat{\mathbf{g}}) = \dfrac{1}{\rho}\left[\mathbf{I} - \dfrac{\mathbf{e}_{ij}(\hat{\mathbf{x}})^* \mathbf{e}_{ij}(\hat{\mathbf{x}})^H}{\rho + \mathbf{e}_{ij}(\hat{\mathbf{x}})^H \mathbf{e}_{ij}(\hat{\mathbf{x}})^*} \right] \mathbf{p} \\ \mathbf{p} = \rho \left[\mathbf{e}_{ij}\left(\widehat{\mathbf{s}_1 \odot \mathbf{w}} \right) - \mathbf{e}_{ij}\left(\hat{\mathbf{h}} \right) \right] + \mathbf{e}_{ij}(\hat{\mathbf{x}})^* \hat{\mathbf{y}}_{ij}. \end{cases}$$

Note that the abovementioned calculation only takes vector multiply–add operations and can thus be efficiently calculated in the iterations.

References

[1] J. Shao, B. Du, C. Wu, and L. Zhang, "Can we track targets from space? A hybrid kernel correlation filter tracker for satellite video," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8719–8731, Nov. 2019.

[2] 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.

[3] C. Fu, A. Carrio, M. A. Olivares-Mendez, R. Suarez-Fernandez, and P. Campoy, "Robust real-time vision-based aircraft tracking from unmanned aerial vehicles," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2014, pp. 5441–5446.

[4] D. Gibbins, P. Roberts, and L. Swierkowski, "A video Geo-location and image enhancement tool for small unmanned air vehicles (UAVs)," in *Proc. Intell. Sensors, Sensor Netw. Inf. Process. Conf.*, 2004, pp. 469–473.

[5] C. Yuan, Z. Liu, and Y. Zhang, "UAV-based forest fire detection and tracking using image processing techniques," in *Proc. Int. Conf. Unmanned Aircr. Syst. (ICUAS)*, Jun. 2015, pp. 639–643.

[6] G. R. Rodríguez-Canosa, S. Thomas, J. del Cerro, A. Barrientos, and B. MacDonald, "A real-time method to detect and track moving objects (DATMO) from unmanned aerial vehicles (UAVs) using a single camera," *Remote Sens.*, vol. 4, no. 4, pp. 1090–1111, Apr. 2012.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

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

[12] 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.

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

[14] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2012, pp. 702–715.

[15] L. Zhang and P. N. Suganthan, "Robust visual tracking via co-trained Kernelized correlation filters," *Pattern Recognit.*, vol. 69, pp. 82–93, Sep. 2017.

[16] C. Wang, L. Zhang, L. Xie, and J. Yuan, "Kernel cross-correlator," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 4179–4186.

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

[18] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2005, pp. 886–893.

[19] C. Fu, Z. Huang, Y. Li, R. Duan, and P. Lu, "Boundary effect-aware visual tracking for UAV with online enhanced background learning and multi-frame consensus verification," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Nov. 2019, pp. 1–8.

[20] P. Zhang *et al.*, "Robust visual tracking using multiframe multifeature joint modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 12, pp. 3673–3686, Dec. 2019.

[21] J. Li, Z. Hong, and B. Zhao, "Robust visual tracking by exploiting the historical tracker snapshots," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop (ICCVW)*, Dec. 2015, pp. 41–49.

[22] Y. Li and J. Zhu, "A scale adaptive kernel correlation filter tracker with feature integration," in *Proc. Eur. Conf. Comput. Vis. Workshops (ECCVW)*, 2014, pp. 254–265.

[23] 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.

[24] 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.

[25] 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.

[26] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, "Convolutional features for correlation filter based visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop (ICCVW)*, Dec. 2015, pp. 58–66.

[27] P. Li, D. Wang, L. Wang, and H. Lu, "Deep visual tracking: Review and experimental comparison," *Pattern Recognit.*, vol. 76, pp. 323–338, Apr. 2018.

[28] M. Danelljan, A. Robinson, F. Shahbaz 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.
[29] 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.
[30] 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. 4847–4856.
[31] 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.
[32] K. Zhang, X. Li, H. Song, Q. Liu, and W. Lian, "Visual tracking using spatio-temporally nonlocally regularized correlation filter," *Pattern Recognit.*, vol. 83, pp. 185–195, Nov. 2018.
[33] D. Zhao, L. Xiao, H. Fu, T. Wu, X. Xu, and B. Dai, "Augmenting cascaded correlation filters with spatial–temporal saliency for visual tracking," *Inf. Sci.*, vol. 470, pp. 78–93, Jan. 2019.
[34] G. Zhu, J. Wang, Y. Wu, X. Zhang, and H. Lu, "MC-HOG correlation tracking with saliency proposal," in *Proc. AAAI Conf. Artif. Intell.*, 2016, pp. 1–7.
[35] W. Feng, R. Han, Q. Guo, J. Zhu, and S. Wang, "Dynamic saliency-aware regularization for correlation filter-based object tracking," *IEEE Trans. Image Process.*, vol. 28, no. 7, pp. 3232–3245, Jul. 2019.
[36] 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.
[37] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
[38] 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.
[39] 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.
[40] 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.
[41] 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.
[42] 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.
[43] T. Zhang, C. Xu, and M.-H. Yang, "Multitask correlation particle filter for robust object tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4335–4343.
[44] 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.
[45] 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. 1–10.
[46] 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.

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, in 2015.

During his Ph.D., he held two research positions at Arizona State University, Tempe, AZ, USA, and Nanyang Technological University (NTU), Singapore. After receiving his Ph.D., he worked at NTU as a Post-Doctoral Research Fellow. He is currently 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 worked on two international, two national, and four industrial projects related to the vision for UAV. In addition, he has published more than 50 journal and conference papers (including the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON MECHATRONICS, the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, the Conference on Computer Vision and Pattern Recognition, the International Conference on Computer Vision, the International Conference on Intelligent Robots and Systems, and ICRA) related to the intelligent vision and control for UAV. His research areas are intelligent vision and control for US in a complex environment.

Juntao Xu is pursuing the B.Eng. degree with the School of Mechanical Engineering, Tongji University, Shanghai, China.

His research interests include computer vision and machine learning.

Fuling Lin received the B.Eng. degree from Tongji University, Shanghai, China, in 2019, where he is pursuing the M.S. degree with the School of Mechanical Engineering.

His research interests include visual tracking and computer vision.

Fuyu Guo received the B.Eng. degree in mechanical engineering from Northeastern University, Shenyang, China. He is pursuing the Ph.D. degree with the School of Mechanical Engineering, Chongqing University, Chongqing, China.

During his Ph.D., he was a Visiting Student with the Control Robotics Intelligence (CRI) Group, School of Mechanical and Aerospace Engineering (MAE), Nanyang Technological University (NTU), Singapore. His research interests include robotic manipulation and intelligent perception.

Tingcong Liu is pursuing the B.Sc. degree in mathematics with the College of Liberal Arts and Sciences, University of Illinois at Urbana–Champaign, Champaign, IL, USA.

His research interests include machine learning and data mining.

Zhijun Zhang (Senior Member, IEEE) received the Ph.D. degree in communication and information systems from Sun Yat-sen University, Guangzhou, China, in 2012.

From 2013 to 2015, he was a Post-Doctoral Research Fellow with the Institute for Media Innovation, Nanyang Technological University, Singapore. Since 2020, he has been a Full Professor with the School of Automation Science and Engineering, South China University of Technology, Guangzhou, where he is with the Human-Robot Intelligence Laboratory, Center for Brain Computer Interfaces and Brain Information Processing. His research interests include neural networks, automatic control, humanoid robots, and human–robot interaction.