

# Learning Consistency Pursued Correlation Filters for Real-Time UAV Tracking

Changhong Fu<sup>1,\*</sup>, Xiaoxiao Yang<sup>2</sup>, Fan Li<sup>1</sup>, Juntao Xu<sup>1</sup>, Changjing Liu<sup>1</sup>, and Peng Lu<sup>3</sup>

**Abstract**—Correlation filter (CF)-based methods have demonstrated exceptional performance in visual object tracking for unmanned aerial vehicle (UAV) applications, but suffer from the undesirable boundary effect. To solve this issue, spatially regularized correlation filters (SRDCF) proposes the spatial regularization to penalize filter coefficients, thereby significantly improving the tracking performance. However, the temporal information hidden in the response maps is not considered in SRDCF, which limits the discriminative power and the robustness for accurate tracking. This work proposes a novel approach with dynamic consistency pursued correlation filters, *i.e.*, the CPCF tracker. Specifically, through a correlation operation between adjacent response maps, a practical consistency map is generated to represent the consistency level across frames. By minimizing the difference between the practical and the scheduled ideal consistency map, the consistency level is constrained to maintain temporal smoothness, and rich temporal information contained in response maps is introduced. Besides, a dynamic constraint strategy is proposed to further improve the adaptability of the proposed tracker in complex situations. Comprehensive experiments are conducted on three challenging UAV benchmarks, *i.e.*, UAV123@10FPS, UAVDT, and DTB70. Based on the experimental results, the proposed tracker favorably surpasses the other 25 state-of-the-art trackers with real-time running speed ( $\sim 43$ FPS) on a single CPU.

## I. INTRODUCTION

Nowadays, due to the unmatched mobility and portability, unmanned aerial vehicle (UAV) has aroused widespread attention for various applications, such as path planning [1], autonomous landing [2], obstacle avoidance [3], and aerial cinematography [4]. As the basis of the above applications, developing a real-time, robust and accurate tracking method is imperative. However, due to many challenges introduced by unmanned airborne flight, such as aggressive UAV motion and viewpoint change, visual tracking in UAV applications is still a tough task. Besides, the nature of UAV also presents great challenges for visual tracking, *e.g.*, mechanical vibration, limited computing power, and battery capacity.

Correlation filter (CF)-based approaches [6]–[9] have been extensively applied to tackle with the aforementioned problems in UAV tracking, due to the high computational efficiency and satisfactory tracking performance. By using the property of a circular matrix, the CF-based methods can

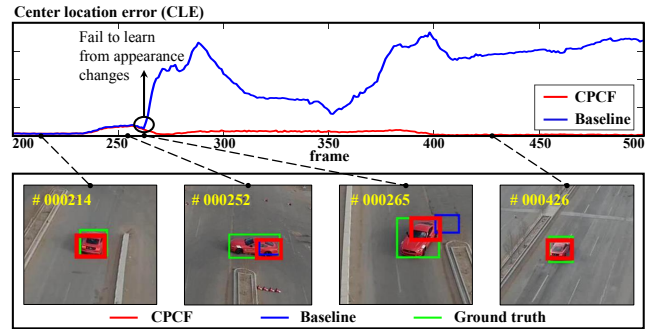


Fig. 1. Center location error (CLE) comparison between the presented CPCF tracker and the baseline, *i.e.*, SRDCF [5]. CLE denotes the Euclidean distance between the estimated center location of the target and the ground truth. As shown in the sequence *car16\_2* from UAV123@10FPS, when the viewpoint change occurs, CPCF shows great adaptability for appearance changes, while the baseline fails to track the object robustly.

learn correlation filters efficiently in the frequency domain. The number of negative samples also increases significantly without producing a heavy computational burden. Nonetheless, due to the property of cyclic shift operation, inaccurate negative samples are introduced by the undesired boundary effect, which substantially reduces the discriminative power of the learned model. To tackle this problem, spatially regularized correlation filters (SRDCF) [5] proposes the spatial regularization to penalize the filter coefficients in the background. In this way, the boundary effect is mitigated and a larger set of negative samples are introduced, which significantly improves the tracking performance. However, during a practical tracking process, the continuity between image frames implies a strong time sequence correlation worth exploring. Essentially, the time sequence correlation is caused by the continuity of the object location in the image, which is finally reflected as the continuity of response maps in the time domain. Therefore, the introduction of response maps has been a crucial issue for exploiting temporal information efficiently. Notwithstanding, SRDCF focuses on improving spatial solutions without concerning the key temporal information in response maps.

To thoroughly explore temporal information in response maps without losing computation efficiency, this work proposes to pursue dynamic consistency across frames. Specifically, with the correlation operation between response maps, the consistency map is produced to evaluate the consistency level for consecutive two frames. Furthermore, an ideal consistency map with the highest consistency level is designed as the correlation result between two ideal responses. By minimizing the difference between the ideal and the practical consistency map, the consistency is forced to maintain a high level, and thus the rich temporal information is injected

\*Corresponding author

<sup>1</sup>Changhong Fu, Fan Li, Juntao Xu and Changjing Liu are with the School of Mechanical Engineering, Tongji University, 201804 Shanghai, China. changhongfu@tongji.edu.cn

<sup>2</sup>Xiaoxiao Yang is with the School of Electronics and Information Engineering, Tongji University, 201804 Shanghai, China.

<sup>3</sup>Peng Lu is with the Adaptive Robotic Controls Lab (Ar-cLab), Hong Kong Polytechnic University (PolyU), Hong Kong, China peng.lu@polyu.edu.hk

efficiently. Moreover, considering the rate of appearance changes is distinct in various tracking periods, a dynamic constraint is introduced to avoid the mismatch between the fixed ideal consistency map and the rate of appearance changes. Concretely, depending on the quality of response maps, the ideal consistency map is dynamically adjusted to meet the requirement of consistency level and further enhance the adaptiveness in different UAV tracking scenarios. As shown in Fig. 1, in case of viewpoint change, the CPCF tracker is well adapted to the fast appearance changes, while the baseline fails to track the target robustly.

Therefore, a novel approach with dynamic consistency pursued correlation filters is proposed, *i.e.*, the CPCF tracker. The main contributions of this work are as follows:

- A novel method to pursue consistency across frames is proposed. In this way, rich temporal information in response maps is exploited thoroughly to boost the accuracy and robustness in the UAV tracking process.
- A dynamic constraint strategy is introduced to set up an adaptive restriction on the consistency level. Based on the quality of the previous response map, the dynamic constraint can adaptively adjust a suitable consistency level and further increases the flexibility to cope with object appearance changes in UAV tracking.
- The CPCF tracker is evaluated exhaustively on three challenging UAV benchmarks. It is compared with 25 state-of-the-art trackers including both hand-crafted and deep trackers. Experiments verify that the CPCF tracker favorably surpasses other trackers in terms of both accuracy and robustness with satisfactory speed for real-time tasks on a single CPU.

## II. RELATED WORKS

### A. Tracking with correlation filters

CF-based approaches have been widely applied in visual object tracking tasks since the proposal of the minimum output sum of the squared error (MOSSE) filter [10]. J. F. Henriques *et al.* [6] extend MOSSE by exploiting the kernel trick and multi-channel features to improve the CF-based method. Besides, the CF-based framework is further developed by multi-resolution scale [11] and part-based analysis [12], [13]. For feature extraction methods, the hand-crafted features including histogram of oriented gradient (HOG) [14] and color names (CN) [15] have been widely used in the tracking process. Moreover, to attain a more comprehensive object appearance representation, some recent works [5], [16] have combined deep features into the CF-based framework. Nonetheless, the heavy computational load brought by deep features deprives it of the ability to be applied in real-time UAV tracking tasks. Consequently, it is still an open problem to design a tracker with both outstanding performance and satisfactory running speed.

### B. Tracking with spatial information

To improve both the tracking accuracy and robustness, recent methods utilizing spatial information have been proposed [5], [7], [17], [18]. By integrating the spatial regular-

ization, SRDCF [5] can penalize the background representing filter coefficients and learn the filter on a significantly larger set of negative training samples. Background-aware correlation filter (BACF) [7] directly multiplies the filter with a binary matrix to expend the search regions. In this way, BACF can utilize not only the target but also the real background information for training. In the CSR-DCF tracker [19], the filter is equipped with spatial reliability maps to improve the tracking of non-rectangular targets and suppresses the boundary effects. However, the improvement brought by spatial information alone is not enough comprehensive. In addition to the spatial information, the effective introduction of both spatial and temporal information has attracted increasing attention among the CF-based tracking community.

### C. Tracking with temporal information

Considering the strong time sequence correlation between the video frames, some trackers exploit the temporal information to further improve the tracking performance [20]–[22]. SRDCFdecon [20] reweights its historical training samples to reduce the problem caused by sample corruption. However, depending on the size of the training set, the tracker may need to store and process a great number of historical samples and thereby sacrificing its tracking efficiency. STRCF [22] proposes a temporal regularization to penalize the variation of filter coefficients in an element-wise manner and ensures the temporal smoothness. However, the rigid element-wise constraint may fail the filter in learning critical appearance changes and limit the adaptiveness of the tracker. Thus, the presented CPCF tracker considers the temporal information by evaluating the consistency level between response maps as a whole. Consequently, the CPCF tracker can maintain the temporal smoothness flexibly and enhance the robustness of tracking.

## III. PROPOSED TRACKING APPROACH

In this work, the proposed CPCF tracker focuses on pursuing the dynamic consistency across frames. Therefore, this section first introduces the consistency evaluation method, then introduces the design process of the dynamic consistency constraint. Finally, the overall objective of CPCF is given. Its main workflow can be seen in Fig. 2.

### A. Consistency evaluation

To evaluate the consistency across frames, this work proposes to study the similarity between the detection response  $\mathbf{R}_k$  and target response  $\mathbf{R}'_k$ . Specifically, the responses  $\mathbf{R}_k$  and  $\mathbf{R}'_k$  are obtained as:

$$\begin{cases} \mathbf{R}_k = \sum_{d=1}^D \mathbf{w}_{k-1}^d \star \mathbf{z}_k^d \\ \mathbf{R}'_k = \sum_{d=1}^D \mathbf{w}_k^d \star \mathbf{x}_k^d \end{cases}, \quad (1)$$

where the subscripts  $(k-1)$  and  $k$  denote the  $(k-1)$ -th and  $k$ -th frame, respectively. The superscript  $d$  denotes the  $d$ -th channel.  $\mathbf{w}_{k-1}^d$  and  $\mathbf{w}_k^d$  denote the correlation filters.  $\mathbf{z}_k^d$  and  $\mathbf{x}_k^d$  denote the detection samples and training samples. The operator  $\star$  denotes the cyclic correlation operation.

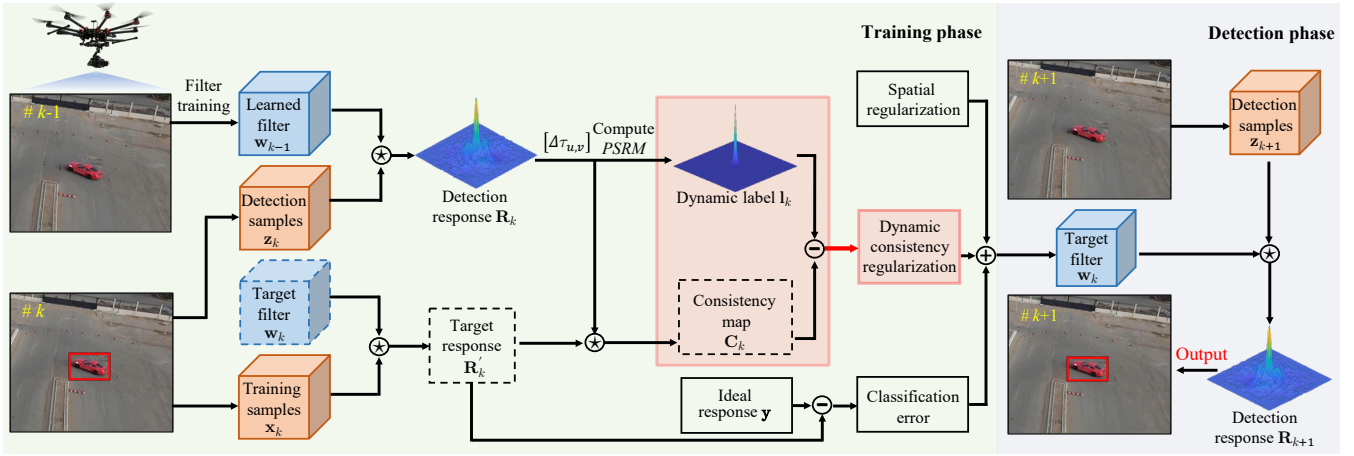


Fig. 2. The main workflow of the proposed CPCF tracker. In the training phase, to evaluate the consistency level between frames, the consistency map  $\mathbf{C}_k$  is generated by responses  $\mathbf{R}_k$  and  $\mathbf{R}'_k$  via the correlation operation. Moreover, based on the  $PSRM$  scores calculated from  $\mathbf{R}_k$ , the label  $\mathbf{I}_k$  is dynamically adjusted to set a self-adaptive constraint on  $\mathbf{C}_k$ . Combining the spatial regularization, classification error and dynamic consistency regularization, the filter training objective is solved through an ADMM optimization to obtain filter  $\mathbf{w}_k$ . Note that in the training phase of the  $k$ -th frame, the to be obtained the target filter  $\mathbf{w}_k$ , the target response  $\mathbf{R}_k$  and the consistency map  $\mathbf{C}_k$  are represented by dashed lines.

Considering the potent time sequence correlation between the detection response  $\mathbf{R}_k$  and the target response  $\mathbf{R}'_k$ , bountiful temporal information is hidden in the consistency between frames. Therefore, to exploit the temporal information, the cyclic correlation operation is adopted to evaluate the consistency between responses by the following formula:

$$\mathbf{C}_k = \mathbf{R}'_k \star \mathbf{R}_k[\Delta\tau_{u,v}], \quad (2)$$

where  $\mathbf{C}_k \in \mathbb{R}^N$  denotes the consistency map. The subscripts  $u$  and  $v$  indicate the difference between the peak positions of detection response  $\mathbf{R}_k$  and the center of the image patch. The operator  $[\Delta\tau_{u,v}]$  shifts the peak of  $\mathbf{R}_k$  to the center of the response map in the two-dimensional space.

### B. Constraint on consistency

In traditional CF-based methods, the responses  $\mathbf{R}'_k$  and  $\mathbf{R}_k$  are both forced to be equal to the ideal response  $\mathbf{y}$  in the ideal situation. Accordingly, considering the generating process of  $\mathbf{C}_k$  in Eq. (2), an ideal consistency map should be produced by two ideal responses  $\mathbf{y}$ . Thus, the ideal consistency map can be served as the consistency constraint label for the practical consistency map  $\mathbf{C}_k$ , and a fixed constraint label  $\mathbf{I}_f$  is designed as follows:

$$\mathbf{I}_f = \mathbf{y} \star \mathbf{y}, \quad (3)$$

where  $\mathbf{y}$  denotes the ideal response. Besides, since the label  $\mathbf{I}_f$  is produced by two identical ideal responses  $\mathbf{y}$ , which means the highest strength for consistency constraint is applied to  $\mathbf{C}_k$  as:

$$\|\mathbf{I}_f - \mathbf{C}_k\|_2^2. \quad (4)$$

By minimizing the squared error between  $\mathbf{I}_f$  and  $\mathbf{C}_k$  in Eq. (4), the responses  $\mathbf{R}_k$  and  $\mathbf{R}'_k$  are forced to pursue a high consistency level, and thus abundant temporal information is injected efficiently.

Moreover, for the practical tracking process, since the rate of the object appearance changing varies in different tracking scenarios, the label  $\mathbf{I}_f$  should adapt to the rate of appearance

changes instead of enforcing a fixed consistency constraint. Thus, based on the fixed constraint label in Eq. (3), the dynamic constraint label is further proposed as follows:

$$\mathbf{I}_k = h_k \mathbf{I}_f, \quad (5)$$

where the  $h_k$  indicates the dynamic regulatory factor to adjust the constraint strength for consistency. The design of  $h_k$  is described in detail in Section III-E.

### C. Overall objective

The overall objective of the presented CPCF tracker is to minimize the following loss function:

$$\begin{aligned} \mathcal{E}(\mathbf{w}_k) = & \frac{1}{2} \|\mathbf{y} - \sum_{d=1}^D \mathbf{w}_k^d \star \mathbf{x}_k^d\|_2^2 + \frac{1}{2} \sum_{d=1}^D \|\mathbf{s}_k^d \mathbf{w}_k^d\|_2^2 \\ & + \frac{\gamma}{2} \sum_{d=1}^D \|\mathbf{I}_k - (\sum_{d=1}^D \mathbf{w}_k^d \star \mathbf{x}_k^d) \star \mathbf{R}_k[\Delta\tau_{u,v}]\|_2^2 \end{aligned} \quad (6)$$

The presented loss function contains three terms, *i.e.*, the first classification error term, the second spatial regularization term and the last dynamic consistency regularization term. For the first term,  $\mathbf{x}_k^d$  and  $\mathbf{w}_k^d$  denote the  $d$ -th channel of training sample and correlation filter, respectively. For the second term, the spatial weight function  $\mathbf{s}_k^d$  is introduced to mitigate the boundary effect. For the third term,  $\gamma$  denotes the consistency penalty. The detection response  $\mathbf{R}_k$  can be expressed as  $\mathbf{R}_k = \sum_{d=1}^D \mathbf{w}_{k-1}^d \star \mathbf{z}_k^d$ , and can be considered as a constant in the training process.

### D. Optimization operations

The following optimization is derived in one-dimensional case, and can be easily extended to two-dimensional case. To convert Eq. (6) to frequency domain conveniently, the equation is firstly expressed as matrix form as follows:

$$\begin{aligned} \mathcal{E}(\mathbf{w}_k) = & \frac{1}{2} \|\mathbf{y} - \mathbf{X}_k \mathbf{w}_k\|_2^2 + \frac{1}{2} \|\mathbf{S}_k \mathbf{w}_k\|_2^2 \\ & + \frac{\gamma}{2} \|\mathbf{I}_k - \mathbf{R}^\top \mathbf{X}_k \mathbf{w}_k\|_2^2 \end{aligned}, \quad (7)$$

where  $\mathbf{y} = [\mathbf{y}(1) \cdots \mathbf{y}(N)]^\top$ ,  $\mathbf{w}_k = [\mathbf{w}_k^1 \cdots \mathbf{w}_k^D]^\top$ ,  $\mathbf{s}_k = [\mathbf{s}_k^1 \cdots \mathbf{s}_k^D]^\top$ , and  $\mathbf{X}_k = [\mathbf{x}_k[\Delta\tau_1]^\top \cdots \mathbf{x}_k[\Delta\tau_N]^\top]^\top$ .  $\mathbf{x}_k[\Delta\tau_n]$  is the circularly shifted sample.  $\mathbf{R}$  denotes the circular matrix generated by the shifted detection response  $\mathbf{R}_k[\Delta\tau_{u,v}]$ . The operator  $\top$  denotes the conjugate transpose operation. In order to improve computing efficiency, Eq. (7) is further transferred into the frequency domain as follows:

$$\begin{aligned} \hat{\mathcal{E}}(\mathbf{w}_k, \hat{\mathbf{g}}_k) &= \frac{1}{2} \|\hat{\mathbf{y}} - \hat{\mathbf{X}}_k \hat{\mathbf{g}}_k\|_2^2 + \frac{1}{2} \|\mathbf{s}_k \mathbf{w}_k\|_2^2 \\ &+ \frac{\gamma}{2} \|\hat{\mathbf{l}}_k - \hat{\mathbf{X}}_k^r \hat{\mathbf{g}}_k\|_2^2, \quad (8) \\ \text{s.t. } \hat{\mathbf{g}}_k &= \sqrt{N} \mathbf{F} \mathbf{w}_k \end{aligned}$$

where  $\hat{\mathbf{g}}_k \in \mathbb{C}^{DN \times 1}$  is introduced as an auxiliary variable. The superscript  $\hat{\cdot}$  denotes the discrete Fourier Transform (DFT) of a signal, *i.e.*,  $\hat{\alpha} = \sqrt{N} F \alpha$ . The matrix  $\hat{\mathbf{X}}_k$  and  $\hat{\mathbf{X}}_k^r$  are defined as  $\hat{\mathbf{X}}_k = [\text{diag}(\hat{\mathbf{x}}_k^1)^\top, \dots, \text{diag}(\hat{\mathbf{x}}_k^D)^\top]$  of size  $N \times DN$ ,  $\hat{\mathbf{X}}_k^r = [\text{diag}(\hat{\mathbf{r}}_k^1 \odot \hat{\mathbf{x}}_k^1)^\top, \dots, \text{diag}(\hat{\mathbf{r}}_k^D \odot \hat{\mathbf{x}}_k^D)^\top]$  of size  $N \times DN$ , respectively. The operator  $\odot$  indicates the element-wise multiplication.  $\hat{\mathbf{r}}_k$  is the discrete Fourier Transform of shifted detection response  $\mathbf{R}_k[\Delta\tau_{u,v}]$ .

Considering the convexity of Eq. (8), alternative direction method of multipliers (ADMM) is introduced to achieve a globally optimal solution efficiently. Hence Eq. (8) can be expressed in augmented Lagrangian form as follows:

$$\begin{aligned} \mathcal{L}(\mathbf{w}, \hat{\mathbf{g}}, \hat{\zeta}) &= \frac{1}{2} \|\hat{\mathbf{y}} - \hat{\mathbf{X}}_k \hat{\mathbf{g}}_k\|_2^2 + \frac{1}{2} \|\mathbf{s}_k \mathbf{w}_k\|_2^2 \\ &+ \frac{\gamma}{2} \|\hat{\mathbf{l}}_k - \hat{\mathbf{X}}_k^r \hat{\mathbf{g}}_k\|_2^2, \quad (9) \\ &+ \hat{\zeta}^\top (\hat{\mathbf{g}}_k - \sqrt{N} \mathbf{F} \mathbf{w}_k) \\ &+ \frac{\nu}{2} \|\hat{\mathbf{g}}_k - \sqrt{N} \mathbf{F} \mathbf{w}_k\|_2^2 \end{aligned}$$

where  $\hat{\zeta} \in \mathbb{C}^{DN \times 1}$  denotes the Lagrangian vector in the Fourier domain which is defined as  $\hat{\zeta} = [\hat{\zeta}^1 \top, \dots, \hat{\zeta}^D \top]^\top$  and  $\nu$  denotes a penalty factor. To learn filters for the  $(k+1)$ -th frame, ADMM algorithm should be adopted in the  $k$ -th frame. The augmented Lagrangian form can be solved by alternatingly solving subproblems  $\hat{\mathbf{g}}_{k+1}^*$  and  $\mathbf{w}_{k+1}^*$  as follows:

1) Subproblem  $\mathbf{w}_{k+1}^*$ :

$$\begin{aligned} \mathbf{w}_{k+1}^* &= \arg \min_{\mathbf{w}_k} \left\{ \frac{1}{2} \|\mathbf{s}_k \mathbf{w}_k\|_2^2 + \hat{\zeta}^\top (\hat{\mathbf{g}}_k - \sqrt{N} \mathbf{F} \mathbf{w}_k) \right. \\ &+ \left. \frac{\nu}{2} \|\hat{\mathbf{g}}_k - \sqrt{N} \mathbf{F} \mathbf{w}_k\|_2^2 \right\}, \\ &= (\mathbf{S}^\top \mathbf{S} + \nu)^{-1} (\zeta + \nu \hat{\mathbf{g}}_k) \end{aligned} \quad (10)$$

where  $\mathbf{S}$  denotes the  $DMN \times DMN$  diagonal matrix concatenating  $D$  diagonal matrices  $\text{diag}(\mathbf{s})$ .

2) Subproblem  $\hat{\mathbf{g}}_{k+1}^*$ :

$$\begin{aligned} \hat{\mathbf{g}}_{k+1}^* &= \arg \min_{\hat{\mathbf{g}}_k} \left\{ \frac{1}{2} \|\hat{\mathbf{y}} - \hat{\mathbf{X}}_k \hat{\mathbf{g}}_k\|_2^2 \right. \\ &+ \left. \frac{\gamma}{2} \|\hat{\mathbf{l}}_k - \hat{\mathbf{X}}_k^r \hat{\mathbf{g}}_k\|_2^2 \right. \\ &+ \left. \hat{\zeta}^\top (\hat{\mathbf{g}}_k - \sqrt{N} \mathbf{F} \mathbf{w}_k) \right. \\ &+ \left. \frac{\nu}{2} \|\hat{\mathbf{g}}_k - \sqrt{N} \mathbf{F} \mathbf{w}_k\|_2^2 \right\}, \quad (11) \end{aligned}$$

solving the subproblem  $\hat{\mathbf{g}}_{k+1}^*$  directly can bring heavy computational burden due to  $\hat{\mathbf{X}}_k^r \hat{\mathbf{g}}_k$  and  $\hat{\mathbf{X}}_k \hat{\mathbf{g}}_k$  in the function. Fortunately,  $\hat{\mathbf{X}}_k$  and  $\hat{\mathbf{X}}_k^r$  are sparse banded, and thus each element in  $\hat{\mathbf{y}}$ , *i.e.*,  $\hat{\mathbf{y}}(n)$  ( $n = 1, 2, \dots, N$ ) is only dependent on each  $\hat{\mathbf{x}}_k(n) = [\hat{\mathbf{x}}_k^1(n), \hat{\mathbf{x}}_k^2(n), \dots, \hat{\mathbf{x}}_k^D(n)]^\top$  and  $\hat{\mathbf{g}}_k(n) = [\text{conj}(\hat{\mathbf{g}}_k^1(n)), \dots, \text{conj}(\hat{\mathbf{g}}_k^D(n))]^\top$ . The operator  $\text{conj}(\cdot)$  denotes the complex conjugate. Therefore, the subproblem  $\hat{\mathbf{g}}_{k+1}^*$  can be divided into  $N$  independent objectives as:

$$\begin{aligned} \hat{\mathbf{g}}_{k+1}^*(n) &= \arg \min_{\hat{\mathbf{g}}_k(n)} \left\{ \frac{1}{2} \|\hat{\mathbf{y}}(n) - \hat{\mathbf{x}}_k(n)^\top \hat{\mathbf{g}}_k(n)\|_2^2 \right. \\ &+ \left. \frac{\gamma}{2} \|\hat{\mathbf{l}}_k(n) - \hat{\mathbf{x}}_k^r(n)^\top \hat{\mathbf{g}}_k(n)\|_2^2 \right. \\ &+ \left. \hat{\zeta}(n)^\top (\hat{\mathbf{g}}_k(n) - \hat{\mathbf{w}}_k(n)) \right. \\ &+ \left. \frac{\nu}{2} \|\hat{\mathbf{g}}_k(n) - \hat{\mathbf{w}}_k(n)\|_2^2 \right\}, \quad (12) \end{aligned}$$

where  $\hat{\mathbf{w}}_k(n) = [\hat{\mathbf{w}}_k^1(n), \dots, \hat{\mathbf{w}}_k^D(n)]$  and  $\hat{\mathbf{w}}_k^d = \sqrt{D} \mathbf{F} \mathbf{w}_k^d$ .

The solution to each problem is given as follows:

$$\begin{aligned} \hat{\mathbf{g}}_{k+1}^*(n)^* &= \frac{1}{1 + \gamma \hat{\mathbf{r}}(n) \hat{\mathbf{r}}(n)^\top} \left( \hat{\mathbf{x}}_k(n) \hat{\mathbf{x}}_k(n)^\top + \frac{\nu}{1 + \gamma \hat{\mathbf{r}}(n) \hat{\mathbf{r}}(n)^\top} \mathbf{I}_D \right)^{-1} \\ &\left( \hat{\mathbf{x}}_k(n) \hat{\mathbf{y}}_k(n) + \gamma \hat{\mathbf{x}}_k^r(n) \hat{\mathbf{l}}_k(n) - \hat{\zeta}(n) + \nu \hat{\mathbf{w}}_k(n) \right) \end{aligned} \quad (13)$$

**Remark 1:** The solving process of Eq. (12) is shown in the appendix.

To further increase the computation efficiency, the Sherman-Morrison formula is employed, *i.e.*,  $(\mathbf{A} + \mathbf{p} \mathbf{q}^\top)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{p} (\mathbf{I}_D + \mathbf{q}^\top \mathbf{A}^{-1} \mathbf{p})^{-1} \mathbf{q}^\top \mathbf{A}^{-1}$ . Thus, Eq. (13) is identically expressed as below:

$$\begin{aligned} \hat{\mathbf{g}}_{k+1}^*(n)^* &= \frac{1}{\nu} \left( \hat{\mathbf{x}}_k(n) \hat{\mathbf{y}}_k(n) + \hat{\mathbf{x}}_k^r(n) \hat{\mathbf{l}}_k(n) - \hat{\zeta}(n) + \nu \hat{\mathbf{w}}_k(n) \right) \\ &- \frac{\hat{\mathbf{x}}_k(n)}{\nu b} \left( \hat{\mathbf{s}}_{\mathbf{x}_k}(n) \hat{\mathbf{y}}_k(n) - \gamma \hat{\mathbf{s}}_{\mathbf{x}_k}(n) \hat{\mathbf{l}}_k(n) \hat{\zeta}(n) + \nu \hat{\mathbf{s}}_{\mathbf{w}_k}(n) \right) \end{aligned} \quad (14)$$

where

$$\begin{cases} \hat{\mathbf{s}}_{\mathbf{x}_k}(n) = \hat{\mathbf{x}}_k(n)^\top \hat{\mathbf{x}}_k(n) \\ \hat{\mathbf{s}}_{\zeta}(n) = \hat{\mathbf{x}}_k(n)^\top \hat{\zeta} \\ \hat{\mathbf{s}}_{\mathbf{w}_k}(n) = \hat{\mathbf{x}}_k(n)^\top \hat{\mathbf{w}}_k \\ b = \hat{\mathbf{x}}_k(n)^\top \hat{\mathbf{x}}_k(n) + \frac{\nu}{1 + \gamma \hat{\mathbf{r}}(n) \hat{\mathbf{r}}(n)^\top} \end{cases} \quad (15)$$

Lagrangian parameter  $\zeta$  is updated in each iteration according to the following equation:

$$\hat{\zeta}_{k+1}^{j+1} = \hat{\zeta}_{k+1}^j + \nu \left( \hat{\mathbf{g}}_{k+1}^{*(j+1)} - \hat{\mathbf{w}}_{k+1}^{*(j+1)} \right), \quad (16)$$

where the subscripts  $j$  and  $j+1$  indicate the  $j$ -th and  $(j+1)$ -th iteration, respectively.  $\hat{\mathbf{g}}_{k+1}^{*(j+1)}$  and  $\hat{\mathbf{w}}_{k+1}^{*(j+1)}$  denote the solution to subproblem  $\hat{\mathbf{g}}_{k+1}^*$  and  $\hat{\mathbf{w}}_{k+1}^*$  in the  $(j+1)$ -th iteration, respectively.

*E. Dynamic adjusting strategy for label  $\mathbf{l}_k$*

In the practical tracking process, the rate of target appearance variation is various in different scenarios. Therefore, an intelligent tracker should adjust the strength of consistency constraint according to the different tracking scenarios. On

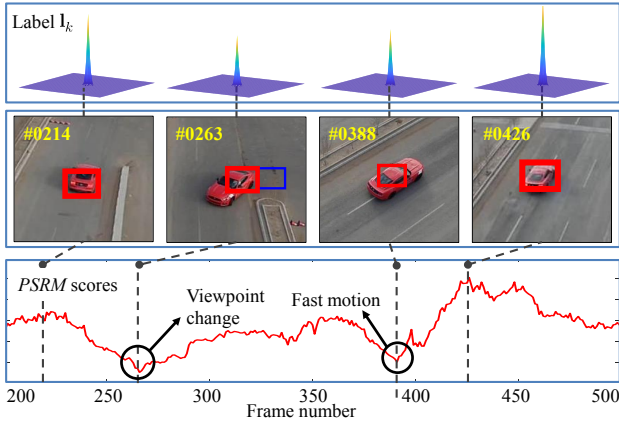


Fig. 3. Illustration of the adjusting process of the dynamic constraint label  $l_k$ . The sequence *car16\_2* in UAV123@10FPS benchmark is shown, and the blue bounding box denotes the baseline, which fails to track the car robustly due to the fast appearance changes of the car. The CPCF tracker, denoted by the red bounding box, appropriately adjusts the constraint strength for consistency in case of fast appearance changes and still track the target successfully.

the one hand, in terms of the fast appearance changes, constraints should be relaxed to render the tracker more authority to change and thus learn from new appearance changes. On the other hand, if the appearance change is smooth, a high-level constraint for consistency is required to enhance the robustness and accuracy. Thus, depending on the quality of the detection response  $R_k$ , the dynamic regulatory factor  $h_k$  is introduced as:

$$h_k = h_{min} + \frac{PSRM}{\alpha} (h_{max} - h_{min}), \quad (17)$$

where  $h_{min}$  and  $h_{max}$  denote the minimum and maximum magnitude of  $l_k$  respectively.  $\alpha$  is a normalized coefficient.  $PSRM$  denotes the quality scores of the response map as follows:

$$PSRM = \frac{R_{max} - \mu_{s1}}{\sigma_{s1}} + \beta R_{max}, \quad (18)$$

where the first and the second term denote the peak to side-lobe ratio (PSR) and the peak value ( $R_{max}$ ) in the response map respectively.  $\mu_{s1}$  and  $\sigma_{s1}$  denote the mean value and the standard deviation in the sidelobe, respectively.  $\beta$  is the weight coefficient to balance two evaluation criterions. The process of dynamic adjusting is shown in Fig. 3.

#### F. Model update

In order to improve the robustness for fast motion, viewpoint change and other challenges, an online adaptation strategy is introduced as follows:

$$\hat{x}_k^{model} = (1 - \eta) \hat{x}_{k-1}^{model} + \eta \hat{x}_k, \quad (19)$$

where  $\eta$  is the learning rate for the appearance model.  $k$  and  $(k-1)$  denote the  $k$ -th and the  $(k-1)$ -th frame, respectively.

## IV. EXPERIMENTS

In this section, the proposed CPCF tracker is evaluated comprehensively on three well-known and widely-used UAV object tracking benchmarks which are especially captured by UAV from the aerial view, *i.e.*,

UAV123@10FPS [24], UAVDT [25], and DTB70 [26], with 243 challenging image sequences. The results are compared with 25 state-of-the-art trackers, *i.e.*, STRCF [22], MCCT-H [27], KCC [28], Staple\_CA [29], SRDCF [5], SAMF\_CA [11], fDSST [30], ECO-HC [17], CSR-DCF [19], BACF [7], Staple [29], SRDCFdecon [20], SAMF [11], KCF [6], DSST [31], CFNet [32], MCCT [27], C-COT [18], ECO [17], IBCCF [33], UDT+ [34], MCPF [35], AD-Net [36], DeepSTRCF [22], and TADT [37]. Moreover, the original evaluation criteria defined in three benchmarks respectively is adopted.

#### A. Implementation details

CPCF is based on HOG [6] and CN [11] features. The consistency penalty  $\gamma$  in Eq. (6) is set to 0.9. For the dynamic constraint strategy,  $h_{min}$  and  $h_{max}$  in Eq. (17) are set to 0.6 and 1.2, respectively. The normalized coefficient  $\alpha$  and the weight coefficient  $\beta$  in Eq. (17) and Eq. (18) are set to 50 and 100, respectively. The learning rate  $\eta$  in Eq. (19) is set to 0.042. All the 26 trackers are performed with MATLAB R2018a on a computer with an i7-8700K CPU (3.7GHz), 32GB RAM and Nvidia GeForce RTX 2080. Note that the CPCF tracker is tested on a single CPU.

#### B. Comparison with hand-crafted based trackers

**Quantitative evaluation:** As shown in Fig. 4, the average overall performance of CPCF and other 15 state-of-the-art trackers utilizing hand-crafted features is demonstrated. The CPCF tracker has surpassed all compared trackers on three UAV benchmarks. In addition, the overall performance of each benchmark is shown in Fig. 5. Concretely, on UAV123@10FPS benchmark, CPCF (0.661) outperforms the second-best CSR-DCF (0.643) and the third-best ECO-HC (0.634) by 1.8% and 2.7%, respectively in precision, and has an advantage of 0.4% and 0.9% over the second-best (ECO-HC, 0.462) and the third-best (STRCF, 0.457), respectively in AUC. On UAVDT benchmark, CPCF (0.720, 0.444) surpasses the second-best (Staple\_CA, 0.695) and the third-best (BACF, 0.686) by 2.5% and 3.4%, respectively in precision, as well as an advancement of 1.1% and 2.5% over BACF (0.433) and SRDCF (0.419), respectively in AUC. On DTB70 benchmark, CPCF (0.710, 0.481) is followed by STRCF (0.649) and CSR-DCF (0.646) in precision, and by ECO-HC (0.453) and CSR-DCF (0.438) in AUC.

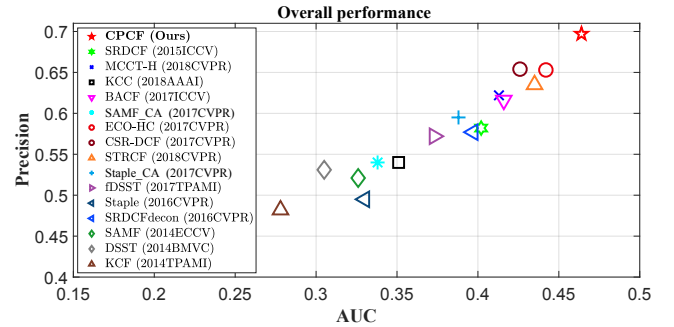


Fig. 4. Comparison between the proposed CPCF tracker and other 15 state-of-the-art trackers based on hand-crafted features. The average value of precision and success rate are calculated by averaging OPE [23] result from three benchmarks. The CPCF tracker favorably surpasses other 15 trackers in both precision and AUC.

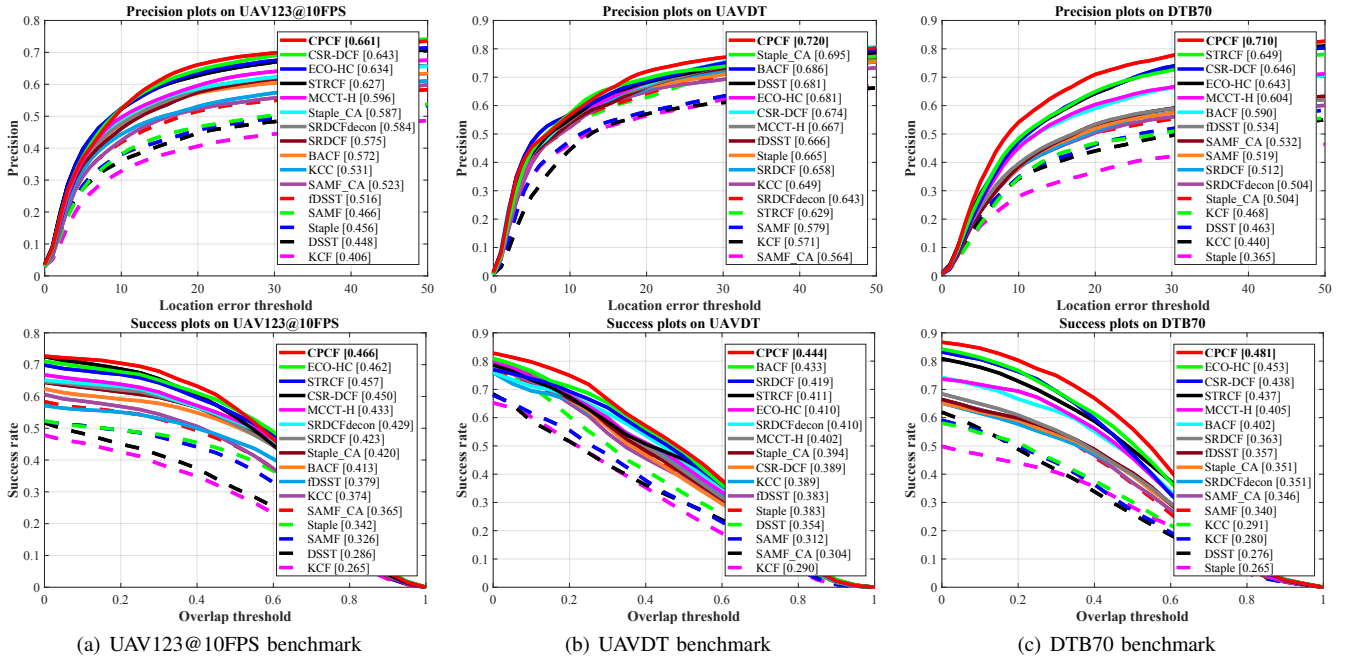


Fig. 5. Precision and success plots of CPCF tracker, as well as all other 15 trackers based on hand-crafted features on (a) UAV123@10FPS, (b) UAVDT and (c) DTB70. The proposed CPCF tracker has the best performance in terms of precision and success rate in three benchmarks.

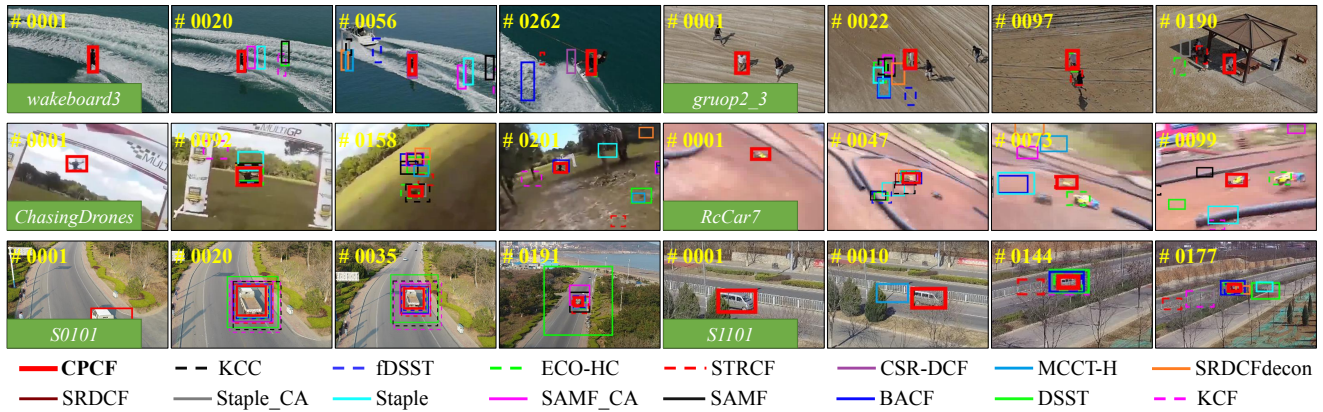


Fig. 6. Qualitative comparisons of the proposed CPCF tracker with other 15 state-of-the-art trackers. From top to bottom, the sequences are *wakeboard3* and *group2\_3* from UAV123@10FPS benchmark, *ChasingDrones* and *RcCar7* from DTB70 benchmark, *S0101* and *S1101* from UAVDT benchmark. Code and the UAV tracking video are: <https://github.com/vision4robotics/CPCF-Tracker> and <https://youtu.be/U004xI-3dl8>.

**Qualitative evaluation:** The comparisons of our approach with other trackers are visualized in Fig. 6. It can be seen that the CPCF tracker performs satisfactorily in different challenging scenarios.

**Speed comparison:** The speed of CPCF is sufficient for real-time UAV tracking applications, and the comparison can be seen in Tabel I.

**Attribute based comparison:** To better demonstrate the ability of CPCF to respond to different challenges, the three benchmarks classify them into different attributes, such as viewpoint change and object motion. Examples of attribute-based comparisons are shown in Fig. 7, which are ranked by AUC. The comparisons demonstrate that the CPCF tracker has a great improvement over its baseline SRDCF, due to the pursuit of consistency across frames.

TABLE I

THE AVERAGE FRAME PER SECOND (FPS) COMPARISON BETWEEN THE CPCF TRACKER AND OTHER 15 TRACKERS ON ALL THREE BENCHMARKS [24]–[26]. RED, GREEN AND BLUE COLOR DENOTE THE FIRST, SECOND AND THIRD PLACE RESPECTIVELY. NOTE THAT THE PRESENTED CPCF TRACKER IS TESTED ON A SINGLE CPU.

Tracker	FPS	Venue	Tracker	FPS	Venue
<b>CPCF</b>	<b>42.95</b>	<b>Ours</b>	CSR-DCF	12.09	CVPR'17
MCCT-H	59.72	CVPR'18	BACF	56.04	ICCV'17
KCC	46.12	AAAI'18	Staple	65.40	CVPR'16
Staple_CA	58.86	CVPR'17	SRDCFdecon	7.48	CVPR'16
SRDCF	14.01	ICCV'17	SAMF	12.76	ECCV'14
SAMF_CA	11.66	CVPR'17	KCF	<b>651.06</b>	TPAMI'14
fDSST	<b>168.06</b>	TPAMI'17	DSST	<b>106.49</b>	BMVC'14
ECO-HC	69.33	CVPR'17	STRCF	28.51	CVPR'18

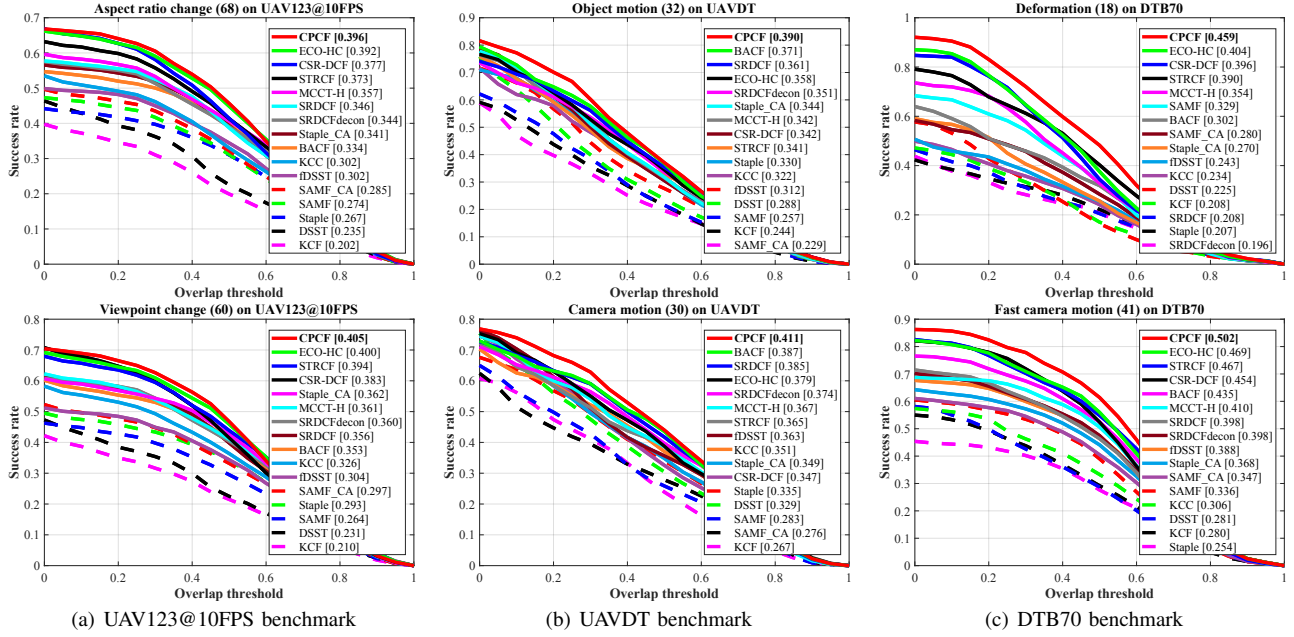


Fig. 7. Attribute-based comparisons between the CPCF tracker and other 15 trackers based on hand-crafted features on (a) UAV123@10FPS, (b) UAVDT and (c) DTB70 benchmarks. All the trackers are ranked by AUC. Details of these attributes can be seen in [24]–[26]. It can be seen that the CPCF tracker has a competitive performance in these attributes across the benchmarks.

TABLE II

THE FRAME PER SECOND (FPS), PRECISION AND AUC COMPARISON BETWEEN CPCF AND DEEP-BASED TRACKERS ON UAVDT [25] BENCHMARK.

**RED, GREEN AND BLUE** COLOR DENOTE THE FIRST, SECOND AND THIRD PLACE RESPECTIVELY. THE PRESENTED CPCF TRACKER PERFORMS FAVORABLY COMPARED WITH OTHER DEEP-BASED TRACKERS.

Tracker	CPCF	CFNet	MCCT	C-COT	ECO	TADT	IBCCF	UDT+	MCPF	ADNet	DeepSTRCF
FPS	<b>48.29</b>	41.05	8.60	1.10	16.38	32.48	3.39	<b>60.42</b>	3.63	7.55	6.61
Precision	<b>0.720</b>	0.680	0.671	0.656	<b>0.700</b>	0.677	0.603	<b>0.697</b>	0.660	0.683	0.667
AUC	<b>0.444</b>	0.428	0.437	0.406	<b>0.454</b>	0.431	0.388	0.416	0.399	0.429	<b>0.437</b>
Venue	Ours	2017CVPR	2018CVPR	2016ECCV	2017CVPR	2019CVPR	2017CVPR	2019CVPR	2017CVPR	2017CVPR	2018CVPR
GPU	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

### C. Comparison with deep-based trackers

In order to fully reflect the performance of CPCF tracker for UAV tracking applications, the CPCF tracker is also compared with deep-based trackers on the UAVDT benchmark. In terms of precision, success rate and speed, the CPCF tracker has performed favorably against other state-of-the-art compared trackers. The comparisons are shown in Table II. Note that all deep-based trackers are tested on the GPU, while the CPCF tracker is tested on a single CPU.

## V. CONCLUSIONS

In this work, a novel approach with dynamic consistency pursued correlation filters, *i.e.*, the CPCF tracker, is proposed. Generally, by exploiting the consistency across frames, rich temporal information in the response maps is introduced to enhance the discriminative power of the tracker. Besides, a dynamic consistency constraint is proposed to strengthen the adaptability in complex situations. Considerable experiments are conducted on three UAV object tracking benchmarks. The experimental results verify the outstanding performance of the presented tracker compared with 25 state-of-the-art trackers. Moreover, the CPCF tracker obtains a real-time speed ( $\sim 43$ FPS) on a single CPU. From our view,

the temporal information behind response maps is further explored by consistency representation, which can contribute to the object tracking on-board UAV.

## APPENDIX

Solving process from (12) to (13) :

$$\begin{aligned} \hat{\mathbf{g}}_{k+1}^*(n) = \arg \min_{\hat{\mathbf{g}}_k(n)} & \left\{ \frac{1}{2} \|\hat{\mathbf{y}}(n) - \hat{\mathbf{x}}_k(n)^\top \hat{\mathbf{g}}_k(n)\|_2^2 \right. \\ & + \frac{\gamma}{2} \|\hat{\mathbf{l}}_k(n) - \hat{\mathbf{x}}_k^r(n)^\top \hat{\mathbf{g}}_k(n)\|_2^2, \\ & + \hat{\zeta}(n)^\top (\hat{\mathbf{g}}_k(n) - \hat{\mathbf{w}}_k(n)) \\ & \left. + \frac{\nu}{2} \|\hat{\mathbf{g}}_k(n) - \hat{\mathbf{w}}_k(n)\|_2^2 \right\}, \end{aligned} \quad (20)$$

$$\begin{aligned} \text{Let } G = & \frac{1}{2} \|\hat{\mathbf{y}}(n) - \hat{\mathbf{x}}_k(n)^\top \hat{\mathbf{g}}_k(n)\|_2^2 \\ & + \frac{\gamma}{2} \|\hat{\mathbf{l}}_k(n) - \hat{\mathbf{x}}_k^m(n)^\top \hat{\mathbf{g}}_k(n)\|_2^2, \\ & + \hat{\zeta}(n)^\top (\hat{\mathbf{g}}_k(n) - \hat{\mathbf{w}}_k(n)) \\ & + \frac{\nu}{2} \|\hat{\mathbf{g}}_k(n) - \hat{\mathbf{w}}_k(n)\|_2^2, \end{aligned} \quad (21)$$

$$\begin{aligned} \frac{\partial G^\top}{\partial \hat{\mathbf{g}}_k(n)} = & - \left( \hat{\mathbf{y}}(n) - \hat{\mathbf{x}}_k(n)^\top \hat{\mathbf{g}}_k(n) \right)^\top \hat{\mathbf{x}}_k(n)^\top \\ & - \gamma \left( \hat{\mathbf{l}}_k(n) - \hat{\mathbf{x}}_k^m(n)^\top \hat{\mathbf{g}}_k(n) \right)^\top \hat{\mathbf{x}}_k^m(n)^\top, \\ & + \hat{\zeta}(n)^\top + \nu (\hat{\mathbf{g}}_k(n) - \hat{\mathbf{w}}_k(n))^\top = 0 \end{aligned} \quad (22)$$

then Eq. (13) is obtained. ■

## ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (No. 61806148).

## REFERENCES

- [1] G. J. Laguna and S. Bhattacharya, "Path planning with Incremental Roadmap Update for Visibility-based Target Tracking," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 1159–1164.
- [2] C. Fu, A. Carrio, M. A. Olivares-Mendez, and P. Campoy, "Online learning-based robust visual tracking for autonomous landing of Unmanned Aerial Vehicles," in *Proceedings of International Conference on Unmanned Aircraft Systems (ICUAS)*, 2014, pp. 649–655.
- [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 *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, 2014, pp. 5441–5446.
- [4] R. Bonatti, C. Ho, W. Wang, S. Choudhury, and S. Scherer, "Towards a Robust Aerial Cinematography Platform: Localizing and Tracking Moving Targets in Unstructured Environments," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 229–236.
- [5] M. Danelljan, G. Hager, F. Shahbaz Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 4310–4318.
- [6] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-Speed Tracking with Kernelized Correlation Filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 583–596, 2015.
- [7] H. Kiani Galoogahi, A. Fagg, and S. Lucey, "Learning background-aware correlation filters for visual tracking," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1135–1143.
- [8] 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 *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 4415–4422.
- [9] C. Fu, J. Xu, F. Lin, F. Guo, T. Liu, and Z. Zhang, "Object Saliency-Aware Dual Regularized Correlation Filter for Real-Time Aerial Tracking," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–12, 2020.
- [10] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 2544–2550.
- [11] Y. Li and J. Zhu, "A scale adaptive kernel correlation filter tracker with feature integration," in *Proceedings of European Conference on Computer Vision (ECCV) Workshops*, 2014, pp. 254–265.
- [12] Y. Li, J. Zhu, and S. C. H. Hoi, "Reliable Patch Trackers: Robust visual tracking by exploiting reliable patches," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 353–361.
- [13] C. Fu, Y. Zhang, Z. Huang, R. Duan, and Z. Xie, "Part-Based Background-Aware Tracking for UAV With Convolutional Features," *IEEE Access*, vol. 7, pp. 79997–80010, 2019.
- [14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2005, pp. 886–893.
- [15] M. Danelljan, F. Shahbaz Khan, M. Felsberg, and J. Van de Weijer, "Adaptive color attributes for real-time visual tracking," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 1090–1097.
- [16] Y. Li, C. Fu, Z. Huang, Y. Zhang, and J. Pan, "Intermittent Contextual Learning for Keyfilter-Aware UAV Object Tracking Using Deep Convolutional Feature," *IEEE Transactions on Multimedia*, pp. 1–13, 2020.
- [17] M. Danelljan, G. Bhat, F. Shahbaz Khan, and M. Felsberg, "ECO: efficient convolution operators for tracking," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 6638–6646.
- [18] M. Danelljan, A. Robinson, F. S. Khan, and M. Felsberg, "Beyond correlation filters: Learning continuous convolution operators for visual tracking," in *Proceedings of European Conference on Computer Vision (ECCV)*, 2016, pp. 472–488.
- [19] A. Lukežič, T. Vojir, L. C. Zajc, J. Matas, and M. Kristan, "Discriminative correlation filter with channel and spatial reliability," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4847–4856.
- [20] M. Danelljan, G. Häger, and F. Shahbaz Khan, "Adaptive decontamination of the training set: A unified formulation for discriminative visual tracking," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1430–1438.
- [21] Q. Guo, W. Feng, C. Zhou, R. Huang, L. Wan, and S. Wang, "Learning Dynamic Siamese Network for Visual Object Tracking," in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 1781–1789.
- [22] F. Li, C. Tian, W. Zuo, L. Zhang, and M. Yang, "Learning spatial-temporal regularized correlation filters for visual tracking," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4904–4913.
- [23] Y. Wu, J. Lim, and M.-H. Yang, "Online Object Tracking: A Benchmark," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.
- [24] M. Mueller, N. Smith, and B. Ghanem, "A benchmark and simulator for UAV tracking," in *Proceedings of European Conference on Computer Vision (ECCV)*, 2016, pp. 445–461.
- [25] D. Du, Y. Qi, H. Yu, Y. Yang, K. Duan, G. Li, W. Zhang, Q. Huang, and Q. Tian, "The unmanned aerial vehicle benchmark: Object detection and tracking," in *Proceedings of European Conference on Computer Vision (ECCV)*, 2018, pp. 370–386.
- [26] S. Li and D. Yeung, "Visual object tracking for unmanned aerial vehicles: A benchmark and new motion models," in *Proceedings of AAAI Conference on Artificial Intelligence (AAAI)*, 2017, pp. 4140–4146.
- [27] N. Wang, W. Zhou, Q. Tian, R. Hong, M. Wang, and H. Li, "Multi-cue correlation filters for robust visual tracking," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4844–4853.
- [28] C. Wang, L. Zhang, L. Xie, and J. Yuan, "Kernel cross-correlator," in *Proceedings of AAAI Conference on Artificial Intelligence (AAAI)*, 2018, pp. 4179–4186.
- [29] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, and P. H. Torr, "Staple: Complementary learners for real-time tracking," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1401–1409.
- [30] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg, "Discriminative scale space tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 8, pp. 1561–1575, 2016.
- [31] M. Danelljan, G. Häger, F. Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," in *Proceedings of British Machine Vision Conference (BMVC)*, 2014, pp. 1–11.
- [32] J. Valmadre, L. Bertinetto, J. Henriques, A. Vedaldi, and P. H. Torr, "End-to-end representation learning for correlation filter based tracking," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2805–2813.
- [33] F. Li, Y. Yao, P. Li, D. Zhang, W. Zuo, and M. Yang, "Integrating boundary and center correlation filters for visual tracking with aspect ratio variation," in *Proceedings of IEEE International Conference on Computer Vision (ICCV) Workshops*, 2017, pp. 2001–2009.
- [34] N. Wang, Y. Song, C. Ma, W. Zhou, W. Liu, and H. Li, "Unsupervised deep tracking," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 1308–1317.
- [35] T. Zhang, C. Xu, and M. Yang, "Multi-task correlation particle filter for robust object tracking," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4335–4343.
- [36] S. Yun, J. Choi, Y. Yoo, K. Yun, and J. Young Choi, "Action-decision networks for visual tracking with deep reinforcement learning," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2711–2720.
- [37] X. Li, C. Ma, B. Wu, Z. He, and M. Yang, "Target-aware deep tracking," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 1369–1378.