EURASIP Journal on Advances in Signal Processing

RESEARCH

Open Access

Learning spatial regularized correlation filters with response consistency and distractor repression for UAV tracking

Wei Zhang*

*Correspondence: zhangwei.personal@163.com

Department of Computer Science, Baoji University of Arts and Sciences, Baoji, China

Abstract

Correlation filter-based trackers have made significant progress in visual object tracking for various types of unmanned aerial vehicle (UAV) applications due to their promising performance and efficiency. However, the boundary effect remains a challenging problem. Several methods enlarge search areas to handle this shortcoming but introduce more background noise, and the filter is prone to learn from distractors. To address this issue, we present spatial regularized correlation filters with response consistency and distractor repression. Specifically, a temporal constraint is introduced to reinforce the consistency across frames by minimizing the difference between consecutive correlation response maps. A dynamic spatial constraint is also integrated by exploiting the local maximum points of the correlation response produced during the detection phase to mitigate the interference from background distractions. The proposed appearance model can optimize the temporal and spatial constraints together with a spatial regularization weight simultaneously. Meanwhile, the proposed appearance model can be solved effectively based on the alternating direction method of multipliers algorithm. The spatial and temporal information concealed in the response maps is fully taken into consideration to boost overall tracking performance. Extensive experiments are conducted on a public UAV benchmark dataset with 123 challenging sequences. The experimental results and analysis demonstrate that the proposed method outperforms 12 state-of-the-art trackers in terms of both accuracy and robustness while efficiently operating in real time.

Keywords: Visual object tracking, Unmanned aerial vehicle (UAV), Spatial–temporal information, Correlation filter, Response map

1 Introduction

Visual object tracking is widely used in many fields, especially in various types of unmanned aerial vehicle (UAV) applications, such as target following [1], autonomous landing [2, 3], and collision avoidance [4]. Although numerous visual tracking methods have been designed for UAVs [5], robust and accurate UAV tracking remains challenging due to numerous factors like aspect ratio change, fast motion, viewpoint change, low resolution, illumination variation, among others. Additionally, the inherent characteristics

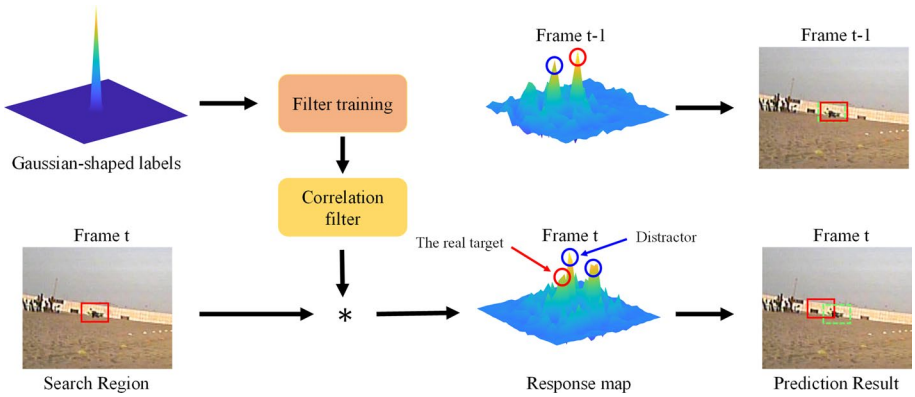

Fig. 1 An instance of the tracking results between consecutive frames of the baseline SRDCF tracker on the uav3 sequence from public UAV benchmark dataset. The red circle represents the real target, and the blue ones represent the distractors. The red bounding box denotes the baseline tracker output, while green dot one is the output of the ground truth

of UAVs, such as mechanical vibration, battery capacity, and limited computing power, also present great challenges for visual tracking.

In recent years, correlation filter (CF)-based trackers have gained increasing attention from researchers due to their satisfactory tracking performance and high computational efficiency [6–12]. Using the property of circulant matrices, the CF effectively transforms the correlation operation in the spatial domain into element-wise multiplication in the frequency domain to increase the computing speed. However, the cyclic shift operation brings undesired boundary effects, which introduces inaccurate negative samples and substantially degrades tracking performance. To address this issue, Danelljan et al. [8] introduced a spatial regularization to penalize filter coefficients in the background and proposed spatially regularized discriminative correlation filters (SRDCF) for object tracking. A larger set of negative samples are introduced to mitigate the boundary effect. In detection, a conventional CF produces a response map, and the object is believed to be located where the map's value is the highest. The quality of the response map reflects the similarity between the target appearance model trained in previous frames and the actual target detected in the current frame to some extent. In addition, the desired response map is unimodal and resembles Gaussian-shaped labels. However, during the practical detection process, the response map can be easily disturbed by complex factors in real scenarios, such as a similar object, partial or complete occlusion, and background clutter. Multiple peaks usually occur in the generated response map, and the tracker is prone to drift due to the interference from background distractions. Although the introduction of a spatial constraint in learning correlation filters improves tracking performance, this method lacks the consideration of spatial and temporal information hidden in response maps. As shown in Fig. 1, the tracking failure occurs if the response value of any distractor exceeds that of the actual target. The prediction result becomes an object that resembles the tracked target in appearance. In addition, the response maps between consecutive frames are not consistent. If the background distractors can be detected and suppressed, and the consistency between consecutive frames is constrained, the tracking accuracy can be improved to a certain extent.

Based on the aforementioned observations, this paper proposes spatial regularized correlation filters with response consistency and distractor repression for robust and efficient UAV tracking to thoroughly explore spatial and temporal information in response maps. Specifically, a temporal constraint is introduced to reinforce the response consistency between consecutive frames. By minimizing the difference between the correlation response from the current frame and the response map from the previous frame, consistency is sustained, and the temporal information in the response map is therefore efficiently integrated. Moreover, considering the disturbance of tracking scenario changes, a dynamic spatial constraint is integrated to suppress the impact of background distractions, which are automatically located by the local maximum points of the response map produced in the detection phase. Thus, the spatial information in the response map is incorporated in the learning phase to enhance the adaptability of the proposed appearance model in different UAV tracking scenarios. Compared to the baseline, the proposed method can suppress background distractors and ensure the quality of the response map, as shown in Fig. 2. The response maps between adjacent frames are also relatively continuous, which is attributed to the consideration of both spatial and temporal information hidden in response maps.

The main contributions of this work are summarized as follows:

(1) We propose a robust and efficient UAV tracking method by jointly learning spatial regularized correlation filters with response consistency and distractor repression. Spatial and temporal information hidden in response maps is taken into consideration to enhance the overall tracking performance.
(2) We apply the alternating direction method of multipliers (ADMM) algorithm to deduce the iteration solutions. Using the ADMM method, an efficient optimization algorithm is developed to find a solution for a spatially regularized CF with temporal and spatial constraints.
(3) The proposed method is evaluated and compared with 12 state-of-the-art trackers on a public UAV benchmark dataset with 123 challenging image sequences. Experimental results demonstrate that the proposed method outperforms other trackers in terms of accuracy and robustness while running efficiently in real time.

The remainder of this paper is organized as follows. Section 2 summarizes several related studies. Section 3 revisits the baseline SRDCF tracker and gives a detailed description of the proposed method. Experimental results are reported and analyzed in Sect. 4, and conclusions are finally drawn in Sect. 5.

2 Related work

2.1 Tracking with correlation filters

CF-based trackers have been widely applied in visual tracking tasks since the introduction of the minimum output sum of the squared error (MOSSE) filter [6], which can reach a leading speed of 669 frames per second (FPS). Following the introduction of the MOSSE tracker, researchers have improved the performance of CF-based trackers from different aspects by introducing the kernel method [13], multi-channel formulation [7], part-based strategy [14–16], scale estimation [17, 18], effective features [19–21], long-term re-detector

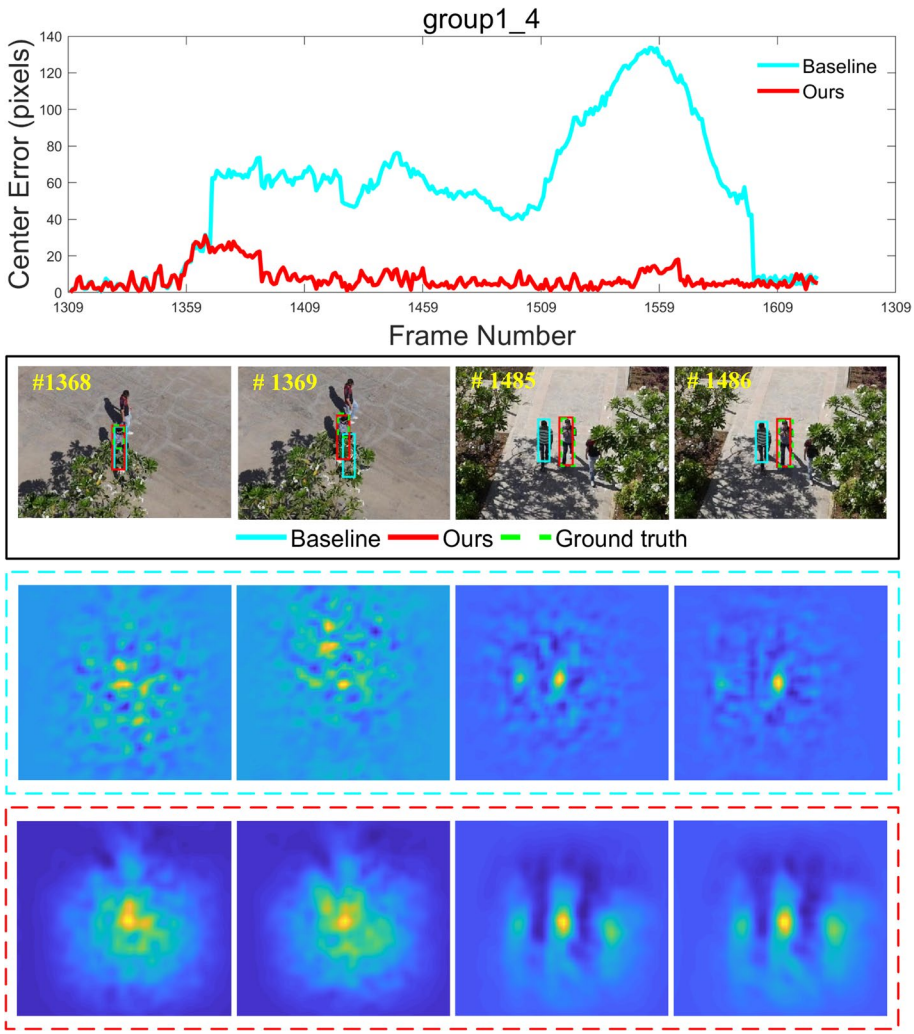

Fig. 2 Comparison of the proposed method with the baseline SRDCF tracker on the group4_1 sequence from public UAV benchmark dataset. From top to bottom are a frame-by-frame comparison of the center location error (in pixels), tracking results of baseline tracker (cyan), the proposed method (red), and the ground truth (green dot), the response maps of the baseline SRDCF tracker and the response maps of the proposed method

[22], and other techniques. Henriques et al. [7, 13] applied the kernel trick and multi-channel features to improve the CF-based trackers. Liu et al. [14], Li et al. [15], and Fu et al. [16] exploited the part-based strategy in the CF model. By identifying scale in a scaling pool, Li and Zhu [17] presented a scale adaptive with multiple features tracker (SAMF) for scale estimation. Danelljan et al. [18] trained a classifier on a scale pyramid for scale estimation and proposed a discriminative scale space tracker (DSST). For effective feature exploitation, Bertinetto et al. [19] utilized two complementary features to establish the target appearance model and proposed a real-time tracker staple. Moreover, to attain a more comprehensive object appearance, some works [20, 21] have incorporated deep features into the CF-based model. Nonetheless, the heavy computational load incurred by the deep features limits their application in real-time UAV tracking tasks. For long-term re-detection, Ma et al. [22] introduced online random fern and support vector machine (SVM) classifiers to recover

the target in case of tracking failure. Although various tracking methods have been put forth over time, it is still challenging to design a tracker with both favorable performance and satisfactory running speed.

2.2 Tracking with spatial and temporal information

Spatiotemporal information is known to offer essential cues for tracking tasks. To improve both tracking accuracy and robustness, some recent methods utilizing spatial information have been proposed [8, 9, 23]. Danelljan et al. [8] proposed SRDCF for visual tracking by incorporating spatial regularization to alleviate the boundary effect caused by the periodic assumption of training samples. Galoogahi et al. [23] trained a CF with limited boundaries (CFLB) to reduce the number of examples in a CF that are affected by boundary effects. Galoogahi et al. [9] further proposed to learn background-aware correlation filters (BACF) for tracking by effectively modeling the target and its background. To enhance the tracking of objects with irregular shapes, Alan Lukezic et al. [24] introduced an automatically estimated spatial reliability map and proposed a discriminative correlation filter with channel and spatial reliability (CSRDCF) method. However, the enhancement brought by spatial information alone is insufficient. In addition to spatial information, the effective addition of temporal information has rekindled increasing interest in the CF-based tracking community. SRDCFdecon [25] 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 large number of historical samples, thereby sacrificing its tracking efficiency. Li et al. [10] incorporated temporal regularization into the SRDCF and proposed spatial–temporal regularized correlation filters (STRCF) for object tracking. Li et al. [26] suggested learning augmented memory correlation filters (AMCF) for UAV tracking. Multiple historical views were selected and stored to be used in training so that they would have more historical appearance information. Huang et al. [11] introduced a regularization term to BACF to restrict the alteration rate of response maps and proposed aberrance repressed correlation filters (ARCF) for UAV tracking. Compared to [10, 11, 25, 26], our method fully exploits the rich spatiotemporal information concealed in response maps to improve the accuracy and robustness of the UAV tracking process.

3 Proposed method

3.1 Revisit the SRDCF tracker

Unlike the conventional kernelized correlation filter (KCF) tracker, the SRDCF tracker introduced a spatial regularization in the learning process to penalize filter coefficients. This allows SRDCF to be learned on a significantly larger set of negative training samples, without corrupting the positive samples, which greatly mitigates the boundary effect and achieves greater performance. The overall objective of SRDCF is formulated by minimizing the following objective:

$$\varepsilon(\boldsymbol{f}) = \sum_{k=1}^{T} \alpha_k \left\| \sum_{d=1}^{D} \boldsymbol{x}_k^d * \boldsymbol{f}^d - \boldsymbol{y}_k \right\|^2 + \sum_{d=1}^{D} \left\| \boldsymbol{w} \circ \boldsymbol{f}^d \right\|^2, \tag{1}$$

where $D = \{(\boldsymbol{x}_k, \boldsymbol{y}_k)\}_{k=1}^{T}$ indicates a set of training samples, each sample $\boldsymbol{x} = [\boldsymbol{x}_k^1, \ldots, \boldsymbol{x}_k^D]$ consists of D feature maps extracted from an image region with dimensions $M \times N$, $*$

denotes the convolution operator, $\circ$ stands for the Hadamard product, $\boldsymbol{y}_k$ represents the desired Gaussian-shaped labels, $\boldsymbol{f}$ and $\boldsymbol{w}$ are the correlation filter and spatial regularization matrix, respectively, the superscript d denotes the d-th channel, and the weight α_k indicates the impact of each sample $\boldsymbol{x}_k^d$; it is set to emphasize more the recent ones. In [8], Danelljan et al. employ the Gauss–Seidel method to iteratively update the CF $\boldsymbol{f}$.

Although SRDCF is effective in mitigating boundary effects, it lacks consideration of spatial and temporal information hidden in response maps. In addition, its failure to exploit the circulant matrix structure, the large linear equations, and the Gauss–Seidel solver also increases the computational burden. More details on implementation can be found in [8].

3.2 Overall formulation

Motivated by the above discussion, we propose spatial regularized correlation filters with response consistency and distractor repression to enhance model stability and accuracy. The overall framework and flowchart of the proposed method are presented in Figs. 3 and 4, respectively. The proposed method based on the SRDCF tracker introduces distractor-repressed and response-consistent constraints to improve the overall tracking performance.

The overall objective of the proposed method is to minimize the following loss function:

$$\varepsilon(\boldsymbol{f}) = \frac{1}{2}\left\| \sum_{d=1}^{D} \boldsymbol{x}^d * \boldsymbol{f}^d - \boldsymbol{y} \circ \mathcal{C}_d \right\|^2 + \frac{\lambda}{2} \sum_{d=1}^{D} \left\| \boldsymbol{w} \circ \boldsymbol{f}^d \right\|^2 + \mathcal{C}_r, \tag{2}$$

where $\mathcal{C}_d$ and $\mathcal{C}_r$ denote the distractor-repressed and response-consistent constraint terms, respectively, and $\boldsymbol{w}$ and λ, respectively, denote the spatial regularization weight and parameter.

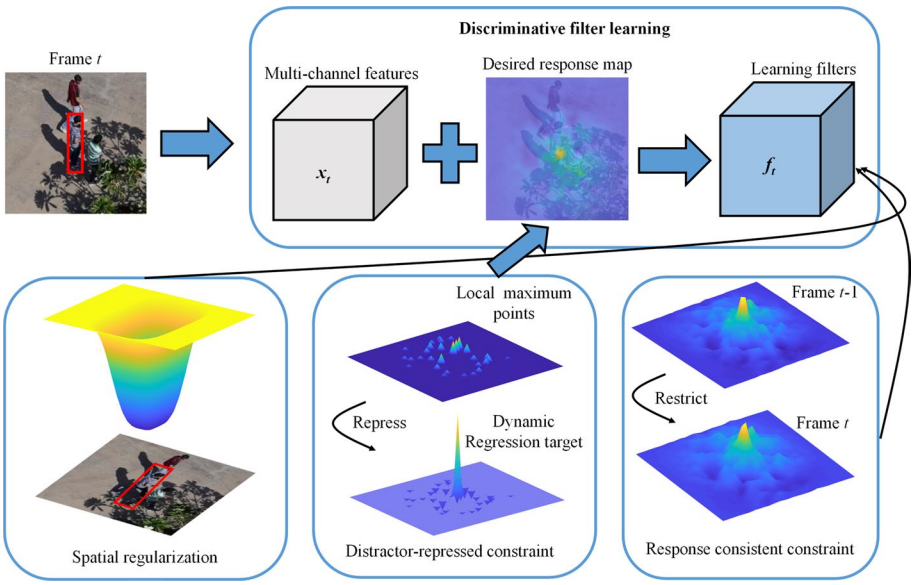

Fig. 3 Overall framework of the proposed method. The second row from left to right shows the spatial regularization, the distractor-repressed constraint, and the response-consistent constraint. In the discriminative filter learning, they are incorporated to constrain the current correlation filters

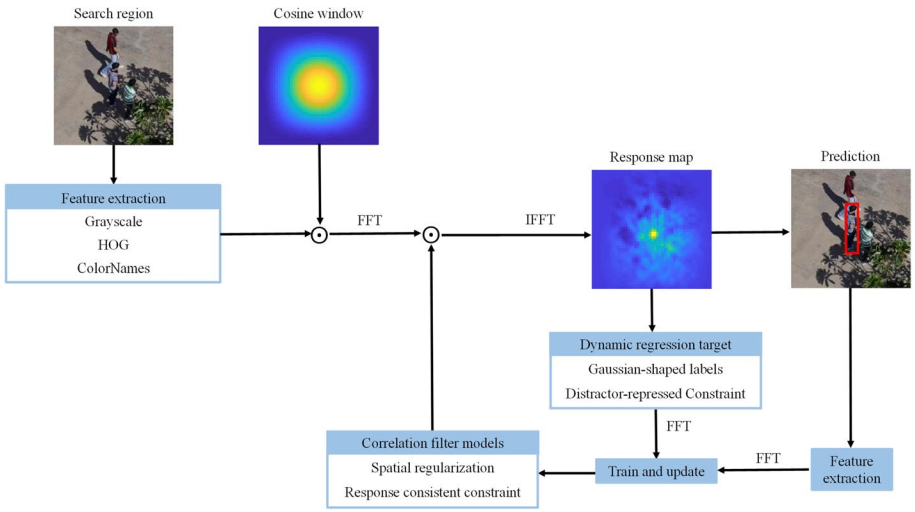

Fig. 4 Flowchart of the proposed method

3.2.1 Distractor-repressed Constraint

Ideally, the response map is unimodal and resembles Gaussian-shaped labels. However, the response map usually has multiple peaks because background distractors exist in actual detection. If the response of the background distractor transcends that of the target object, the tracker will drift to the distractor [12, 27]. In this work, we adopt distractor-repressed constraint to suppress the interferences from background distractions, which is obtained by:

$$\mathcal{C}_d = \boldsymbol{I} - \delta \boldsymbol{P}^{\mathrm{T}} \Delta(\boldsymbol{R}[\varphi_{p,c}]), \tag{3}$$

where $\Delta(\cdot)$ denotes the local maximum cropping function. The local maximum points in the response map $\boldsymbol{R}$ indicate the presence of distractors. Only the top N_d local maxima are selected and counted as distractors, discarding the low response values. The cropping matrix $\boldsymbol{P}^{\mathrm{T}}$ cuts the central area of $\Delta(\cdot)$ to remove the maximum point within the object area. Factor δ controls the repression strength. $\boldsymbol{I}$ is an identity vector. $[\varphi_{p,c}]$ denotes a shift operator to match the peaks of the response map and regression target. The subscripts p and c denote the location difference of the two peaks. The distractor-repressed constraint term $\mathcal{C}_d$ generates a dynamic regression target compared to the fixed target label. The first term in (2) is the ridge regression term that convolves the training samples $\boldsymbol{x} = [\boldsymbol{x}^1, \ldots, \boldsymbol{x}^D]$ with the filter $\boldsymbol{f} = [\boldsymbol{f}^1, \ldots, \boldsymbol{f}^D]$ to fit the distractor-repressed label $\boldsymbol{y}$. It acts as a dynamic spatial constraint to suppress the local maxima of the response map in training phase.

3.2.2 Response consistent constraint

Ideally, the appearance of the target and its context change very little between adjacent frames as the time interval is short. Therefore, there is not much of a change in the correlation response of two consecutive frames. However, abrupt changes in appearance caused by partial or full occlusion and background clutter will lead to response anomalies. In this study, we introduce a response-consistent constraint $\mathcal{C}_r$ to

mitigate the influence of abrupt changes in response maps between two consecutive frames:

$$\mathcal{C}_r = \frac{\gamma}{2}\left\| \sum_{d=1}^{D} \boldsymbol{x}^d * \boldsymbol{f}^d - \sum_{d=1}^{D} \boldsymbol{x}_{t-1}^d * \boldsymbol{f}_{t-1}^d[\varphi_{p,q}] \right\|^2, \tag{4}$$

where γ is the regularization parameter and $\sum_{d=1}^{D} \boldsymbol{x}_{t-1}^d * \boldsymbol{f}_{t-1}^d$ denotes the response map obtained in the $(t-1)$-th frame. The operator $[\varphi_{p,q}]$ shifts two peaks of both response maps to coincide with each other. When abrupt appearance changes occur, the similarity between consecutive frames will suddenly drop and thus, the value of the response-consistent constraint term will be high. This term indicates that the desired response difference between consecutive frames should be zero, which can help suppress the response inconsistency in the training process. It also acts as a temporal constraint to penalize filter coefficients when the response difference is unusually high.

3.3 Optimization of formulation

Considering the convexity of (2), ADMM is introduced to obtain the globally optimal solution. To this end, we first introduce an auxiliary variable $\boldsymbol{g}$, by requiring $\boldsymbol{f} = \boldsymbol{g}$, and the step size parameter μ. The augmented Lagrangian form of (2) can be formulated as:

$$\begin{aligned}\mathcal{L}(\boldsymbol{f}, \boldsymbol{g}, \boldsymbol{\rho}) = {} & \frac{1}{2}\left\| \sum_{d=1}^{D} \boldsymbol{x}^d * \boldsymbol{f}^d - \boldsymbol{y} \circ \mathcal{C}_d \right\|^2 + \frac{\lambda}{2} \sum_{d=1}^{D} \left\| \boldsymbol{w} \circ \boldsymbol{g}^d \right\|^2 + \sum_{d=1}^{D} (\boldsymbol{f}^d - \boldsymbol{g}^d)^{\mathrm{T}} \boldsymbol{\rho}^d \\ & + \frac{\mu}{2} \sum_{d=1}^{D} \left\| \boldsymbol{f}^d - \boldsymbol{g}^d \right\|^2 + \frac{\gamma}{2} \left\| \sum_{d=1}^{D} \boldsymbol{x}^d * \boldsymbol{f}^d - \sum_{d=1}^{D} \boldsymbol{x}_{t-1}^d * \boldsymbol{f}_{t-1}^d[\varphi_{p,q}] \right\|^2,\end{aligned} \tag{5}$$

where $\boldsymbol{\rho}$ is the Lagrange multiplier. By introducing $\boldsymbol{h} = \frac{1}{\mu}\boldsymbol{\rho}$, (5) can be reformulated as:

$$\begin{aligned}\mathcal{L}(\boldsymbol{f}, \boldsymbol{g}, \boldsymbol{h}) = {} & \frac{1}{2}\left\| \sum_{d=1}^{D} \boldsymbol{x}^d * \boldsymbol{f}^d - \boldsymbol{y} \circ \mathcal{C}_d \right\|^2 + \frac{\lambda}{2} \sum_{d=1}^{D} \left\| \boldsymbol{w} \circ \boldsymbol{g}^d \right\|^2 \\ & + \frac{\mu}{2} \sum_{d=1}^{D} \left\| \boldsymbol{f}^d - \boldsymbol{g}^d + \boldsymbol{h}^d \right\|^2 + \frac{\gamma}{2} \left\| \sum_{d=1}^{D} \boldsymbol{x}^d * \boldsymbol{f}^d - \sum_{d=1}^{D} \boldsymbol{x}_{t-1}^d * \boldsymbol{f}_{t-1}^d[\varphi_{p,q}] \right\|^2,\end{aligned} \tag{6}$$

The ADMM algorithm is then adopted by alternately solving the following subproblems,

$$\begin{cases} \boldsymbol{f}^{i+1} = \arg\min_{\boldsymbol{f}} \mathcal{L}(\boldsymbol{f}, \boldsymbol{g}^i, \boldsymbol{h}^i) \\ \boldsymbol{g}^{i+1} = \arg\min_{\boldsymbol{g}} \mathcal{L}(\boldsymbol{f}^{i+1}, \boldsymbol{g}, \boldsymbol{h}^i) \\ \boldsymbol{h}^{i+1} = \boldsymbol{h}^i + \mu(\boldsymbol{f}^{i+1} - \boldsymbol{g}^{i+1}) \end{cases}. \tag{7}$$

The solution to each subproblem is detailed as follows:

Subproblem $\boldsymbol{f}$: Using $\boldsymbol{g}$ and $\boldsymbol{h}$ obtained in the last iteration, the optimal $\boldsymbol{f}$ can be determined by:

$$\begin{aligned}\boldsymbol{f} = \arg\min_{\boldsymbol{f}} \Bigg\{ & \frac{1}{2}\left\| \sum_{d=1}^{D} \boldsymbol{x}^d * \boldsymbol{f}^d - \boldsymbol{y} \circ \mathcal{C}_d \right\|^2 + \frac{\mu}{2} \sum_{d=1}^{D} \left\| \boldsymbol{f}^d - \boldsymbol{g}^d + \boldsymbol{h}^d \right\|^2 \\ & + \frac{\gamma}{2}\left\| \sum_{d=1}^{D} \boldsymbol{x}^d * \boldsymbol{f}^d - \sum_{d=1}^{D} \boldsymbol{x}_{t-1}^d * \boldsymbol{f}_{t-1}^d[\varphi_{p,q}] \right\|^2 \Bigg\}.\end{aligned} \tag{8}$$

Based on the convolution theorem, the cyclic convolution operation in the spatial domain can be replaced by element-wise multiplication in the Fourier domain, and (8) can therefore be rewritten as:

$$\begin{aligned}\hat{\boldsymbol{f}} = \arg\min_{\hat{\boldsymbol{f}}} \Bigg\{ & \frac{1}{2}\left\| \sum_{d=1}^{D} \hat{\boldsymbol{x}}^d \circ \hat{\boldsymbol{f}}^d - \widehat{(\boldsymbol{y} \circ \mathcal{C}_d)} \right\|^2 + \frac{\mu}{2} \sum_{d=1}^{D} \left\| \hat{\boldsymbol{f}}^d - \hat{\boldsymbol{g}}^d + \hat{\boldsymbol{h}}^d \right\|^2 \\ & + \frac{\gamma}{2}\left\| \sum_{d=1}^{D} \hat{\boldsymbol{x}}^d \circ \hat{\boldsymbol{f}}^d - \sum_{d=1}^{D} \hat{\boldsymbol{x}}_{t-1}^d \circ \hat{\boldsymbol{f}}_{t-1}^d[\varphi_{p,q}] \right\|^2 \Bigg\},\end{aligned} \tag{9}$$

where $\hat{\boldsymbol{f}}$ denotes the discrete Fourier transform (DFT) of the filter $\boldsymbol{f}$. Considering the independence of each pixel, the solution can be, respectively, obtained across all channels for every pixel. The optimization in the $\boldsymbol{j}$th pixel can be further reformulated as:

$$\begin{aligned}\mathcal{V}_j(\hat{\boldsymbol{f}}) = \arg\min_{\mathcal{V}_j(\hat{\boldsymbol{f}})} \Bigg\{ & \frac{1}{2}\left\| \mathcal{V}_j(\hat{\boldsymbol{x}})^{\mathrm{T}} \mathcal{V}_j(\hat{\boldsymbol{f}}) - \widehat{(\boldsymbol{y} \circ \mathcal{C}_d)}_j \right\|^2 + \frac{\mu}{2}\left\| \mathcal{V}_j(\hat{\boldsymbol{f}}) - \mathcal{V}_j(\hat{\boldsymbol{g}}) + \mathcal{V}_j(\hat{\boldsymbol{h}}) \right\|^2 \\ & + \frac{\gamma}{2}\left\| \mathcal{V}_j(\hat{\boldsymbol{x}})^{\mathrm{T}} \mathcal{V}_j(\hat{\boldsymbol{f}}) - (\hat{\boldsymbol{R}}_{t-1}^s)_j \right\|^2 \Bigg\},\end{aligned} \tag{10}$$

where $\hat{\boldsymbol{R}}_{t-1}^s$ denotes the DFT of the shifted detection response from the previous frame.

Setting the derivative of (10) to zero, the closed-form solution for $\mathcal{V}_j(\hat{\boldsymbol{f}})$ can be obtained:

$$\mathcal{V}_j(\hat{\boldsymbol{f}}) = \frac{1}{1+\gamma}\left(\mathcal{V}_j(\hat{\boldsymbol{x}}) \mathcal{V}_j(\hat{\boldsymbol{x}})^{\mathrm{T}} + \frac{\mu}{1+\gamma} \boldsymbol{I} \right)^{-1} \boldsymbol{q}, \tag{11}$$

where the vector $\boldsymbol{q}$ takes the form $\boldsymbol{q} = \mathcal{V}_j(\hat{\boldsymbol{x}}) \widehat{(\boldsymbol{y} \circ \mathcal{C}_d)}_j + \mu \mathcal{V}_j(\hat{\boldsymbol{g}}) - \mu \mathcal{V}_j(\hat{\boldsymbol{h}}) + \gamma \mathcal{V}_j(\hat{\boldsymbol{x}}) \hat{\boldsymbol{R}}_{t-1}^s$. Since $\mathcal{V}_j(\hat{\boldsymbol{x}}) \mathcal{V}_j(\hat{\boldsymbol{x}})^{\mathrm{T}}$ is a rank-1 matrix, (11) can be solved with the Sherman–Morrison formula [28], i.e.,$(\boldsymbol{A} + \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}})^{-1} = \boldsymbol{A}^{-1} - \frac{\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}}{1+\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u}}$. In this case, $\boldsymbol{A} = \frac{\mu}{1+\gamma}\boldsymbol{I}$, and $\boldsymbol{u} = \boldsymbol{v} = \mathcal{V}_j(\hat{\boldsymbol{x}})$. As a result, (11) is equivalent to:

$$\begin{aligned}\mathcal{V}_j(\hat{\boldsymbol{f}}) = & \gamma^* \left(\mathcal{V}_j(\hat{\boldsymbol{x}}) \widehat{(\boldsymbol{y} \circ \mathcal{C}_d)}_j + \mu \mathcal{V}_j(\hat{\boldsymbol{g}}) - \mu \mathcal{V}_j(\hat{\boldsymbol{h}}) + \gamma \mathcal{V}_j(\hat{\boldsymbol{x}}) \hat{\boldsymbol{R}}_{t-1}^s \right) \\ & - \gamma^* \frac{\mathcal{V}_j(\hat{\boldsymbol{x}})}{b} \left(\hat{\mathcal{S}}_{\boldsymbol{x}\boldsymbol{x}} \widehat{(\boldsymbol{y} \circ \mathcal{C}_d)}_j + \mu \hat{\mathcal{S}}_{\boldsymbol{x}\boldsymbol{g}} - \mu \hat{\mathcal{S}}_{\boldsymbol{x}\boldsymbol{h}} + \gamma \hat{\mathcal{S}}_{\boldsymbol{x}\boldsymbol{x}} \hat{\boldsymbol{R}}_{t-1}^s \right),\end{aligned} \tag{12}$$

where $b = \frac{\mu}{1+\gamma} + \mathcal{V}_j(\hat{\boldsymbol{x}})^{\mathrm{T}} \mathcal{V}_j(\hat{\boldsymbol{x}})$, $\gamma^* = \frac{1}{1+\gamma}(\frac{\mu}{1+\gamma})^{-1}$, $\hat{\mathcal{S}}_{\boldsymbol{x}\boldsymbol{x}} = \mathcal{V}_j(\hat{\boldsymbol{x}})^{\mathrm{T}} \mathcal{V}_j(\hat{\boldsymbol{x}})$, $\hat{\mathcal{S}}_{\boldsymbol{x}\boldsymbol{g}} = \mathcal{V}_j(\hat{\boldsymbol{x}})^{\mathrm{T}} \mathcal{V}_j(\hat{\boldsymbol{g}})$, and $\hat{\mathcal{S}}_{\boldsymbol{x}\boldsymbol{h}} = \mathcal{V}_j(\hat{\boldsymbol{x}})^{\mathrm{T}} \mathcal{V}_j(\hat{\boldsymbol{h}})$. Note that (12) contains only the vector sum-product operation and thus can be computed efficiently. The filter $\boldsymbol{f}$ can be further obtained by the inverse DFT of $\hat{\boldsymbol{f}}$.

Subproblem $\boldsymbol{g}$: Given $\boldsymbol{f}$ and $\boldsymbol{h}$, the optimal $\boldsymbol{g}$ can be obtained by:

$$\boldsymbol{g} = \arg\min_{\boldsymbol{g}} \left\{ \frac{\lambda}{2} \sum_{d=1}^{D} \left\| \boldsymbol{w} \circ \boldsymbol{g}^d \right\|^2 + \frac{\mu}{2} \sum_{d=1}^{D} \left\| \boldsymbol{f}^d - \boldsymbol{g}^d + \boldsymbol{h}^d \right\|^2 \right\}. \tag{13}$$

Each element of $\boldsymbol{g}$ can be computed independently and thus, the closed-form solution of $\boldsymbol{g}$ can be computed by:

$$\boldsymbol{g} = (\lambda \boldsymbol{W}^{\mathrm{T}} \boldsymbol{W} + \mu \boldsymbol{I})^{-1} (\mu \boldsymbol{f} + \mu \boldsymbol{h}). \tag{14}$$

As a result, the subproblems $\boldsymbol{f}$ and $\boldsymbol{g}$ are solved.

Updating step size parameter μ: The step size parameter μ is updated as follows:

$$\mu^{(i+1)} = \min(\mu^{\max}, \beta \mu^{(i)}), \tag{15}$$

where $\mu^{\max}$ and β denote the maximum value of μ and the scale factor, respectively.

3.4 Update of appearance model

The appearance model $\hat{\boldsymbol{x}}^{\mathrm{model}}$ is updated as follows:

$$\hat{\boldsymbol{x}}_t^{\mathrm{model}} = (1 - \eta) \hat{\boldsymbol{x}}_{t-1}^{\mathrm{model}} + \eta \hat{\boldsymbol{x}}, \tag{16}$$

where η is the learning rate, $\hat{\boldsymbol{x}}$ is the object feature extracted at frame t, and $\hat{\boldsymbol{x}}_t^{\mathrm{model}}$ is the model feature.

3.5 Object location

When a new frame arrives, the filter trained in the last frame $\hat{\boldsymbol{f}}_{t-1}$ is used to localize the object by searching for the peak in the response map calculated as follows:

$$\boldsymbol{R} = \mathcal{F}^{-1} \left(\sum_{d=1}^{D} \hat{\boldsymbol{z}}_t^d \circ \hat{\boldsymbol{f}}_{t-1}^d \right), \tag{17}$$

where $\boldsymbol{z}_t^d$ denotes the feature map of the search area patch in the d-th channel and $\mathcal{F}^{-1}$ represents the IDFT. The target location $\boldsymbol{l}_t$ in frame t can be found at the maximum response value.

3.6 Tracking with response consistency and distractor repression

The details of our proposed method are summarized in Algorithm 1.

Algorithm1 The proposed tracking method.

Input: The image frame t, location $\boldsymbol{l}_{t-1}$ and the scale $\boldsymbol{s}_{t-1}$ of the tracked object on frame $t-1$,the appearance model $\hat{\boldsymbol{x}}_{t-1}^{model}$, and the filter $\boldsymbol{f}_{t-1}$.

Output: Location $\boldsymbol{l}_t$ and scale $\boldsymbol{s}_t$ of the tracked object on frame t.

If $t = 1$ **then**

Extract $\mathcal{C}_d^1$ centered at the ground truth $\boldsymbol{l}_1$ using (3);

Use (12), (14) and (15) to initialize the filters $\boldsymbol{f}_1$ and $\boldsymbol{g}_1$;

Else

Crop the search image patch $\boldsymbol{z}$ with S scales on the frame t centered at $\boldsymbol{l}_{t-1}$;

Extracted gray-scale, color names (CN) [29], and histogram of oriented gradient (HOG) [30] features $[\boldsymbol{z}(s)]_{s=1}^{S}$ of the patch;

Generate the response map $\boldsymbol{R}_t$ using (17) ;

Algorithm1 The proposed tracking method.

Estimate object location $\boldsymbol{l}_t$ on frame t by searching for the highest value in $\boldsymbol{R}_t$;
Extract $\mathcal{C}_d^t$ centered at location $\boldsymbol{l}_t$ using (3);
Update and the filters $\boldsymbol{f}_t$ and $\boldsymbol{g}_t$ using (12), (14) and (15);
Update the appearance model using (16).
End

4 Experiments

In this section, we conduct experiments with the proposed method and 12 other state-of-the-art trackers for comparison, using a publicly available UAV benchmark dataset. The experimental settings are first introduced, the overall performance and attribute-based evaluation of all trackers on the UAV benchmark dataset are then presented, and the qualitative evaluations, ablation study, parameters, and tracking speed analysis are finally discussed.

4.1 Experimental settings

Parameter Settings For the ADMM hyperparameters, we set $\gamma = 0.8$, $\lambda = 0.01$, the maximum value of the step size parameter $\mu^{\max} = 10000$, $\beta = 10$, and $\mu^0 = 1$. The number of iterations for the ADMM N is set to 5, and the learning rate η is 0.0192. The number of top local maxima N_d is set to 30 and δ is 0.25. All parameters are the same for the following experiments. The experiments were carried out in MATLAB 2017b on an Intel(R) Core(TM) i7-7700 CPU (3.6 GHz) with 8 GB RAM.

Features Only the hand-crafted features are utilized for appearance representations, namely gray-scale, CN, and HOG features. The cell dimensions for feature extraction are 4×4 and the HOG orientation bin number is 9.

Datasets To evaluate the tracking performance, experiments were conducted using the public UAV benchmark dataset UAV123@10FPS [31]. The dataset is made up of 123 challenge sequences (37885 frames).

Evaluation Methodology To analyze and evaluate the performance of our proposed method, precision and success rate based on one-pass evaluation (OPE) are employed as the main evaluation criteria. More details can be found in [32].

4.2 Overall performance

The overall performance of our proposed method is evaluated with 12 state-of-the-art trackers, including KCF [7], TLD [33], LCT [21], SAMF [17], Struck [34], BACF [9], Staple [19], AMCF [26], SRDCF [8], SRDCFdecon [25], MEEM [35], and STRCF [10]. Figure 5 shows the precision and success plots of all the trackers on 123 challenging UAV sequences. In precision plots, the distance precision scores at the 20-pixel threshold are 0.673 (our method), 0.627 (STRCF), 0.584 (MEEM), 0.584 (SRDCFdecon), 0.575 (SRDCF), 0.574 (AMCF), 0.573 (Staple), 0.572 (BACF), 0.509 (Struck), 0.465 (SAMF), 0.442 (LCT), 0.415 (TLD), and 0.406 (KCF), respectively. Our proposed method achieves the best precision among all tracking algorithms and outperforms the second and third best trackers by 4.6% and 8.9%, respectively. Likewise, in success plots, the success scores for the area under the curve (AUC) of all trackers are 0.476 (our method), 0.457

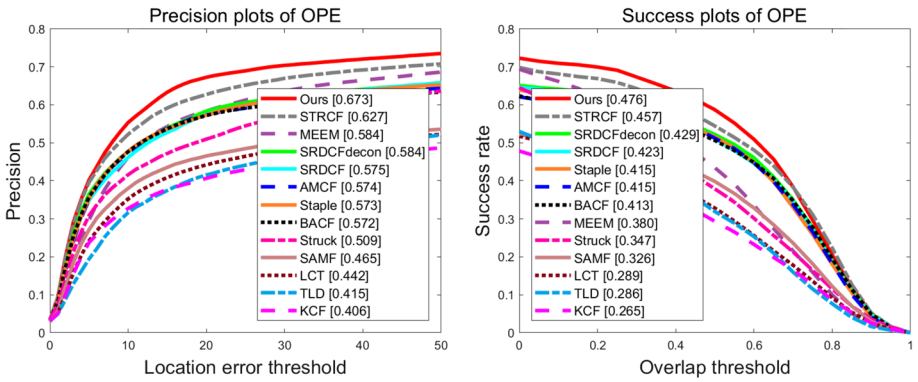

Fig. 5 Precision and success plots of all trackers on 123 UAV sequences

(STRCF), 0.429 (SRDCFdecon), 0.423 (SRDCF), 0.415 (Staple), 0.415 (AMCF), 0.413 (BACF), 0.380 (MEEM), 0.347 (Struck), 0.326 (SAMF), 0.289 (LCT), 0.286 (TLD) and 0.265 (KCF), respectively. Our proposed method also achieves an advantage of 1.9% and 4.7% over the second best tracker STRCF and the third best tracker SRDCFdecon. Thus, it can be summarized that our proposed method outperforms the other 12 state-of-the-art trackers in terms of precision and success rate on 123 UAV sequences.

4.3 Attribute-based evaluation

The benchmark sequences are annotated with 12 attributes, namely scale variation (SV), partial occlusion (POC), camera motion (CM), aspect ratio change (ARC), viewpoint change (VC), low resolution (LR), similar object (SOB), full occlusion (FOC), illumination variation (IV), out-of-view (OV), fast motion (FM), and background clutter (BC). These attributes affect the performance of a tracker and are used to evaluate the tracker in different scenarios. Figures 6 and 7 depict the precision and success plots of different attributes of all trackers on 123 UAV sequences.

In the precision plots, our proposed method performs competitively among the challenging attributes compared with other state-of-the-art trackers. Our method ranks first among ten attributes out of the twelve in the UAV123@10FPS benchmark, namely SV, POC, CM, ARC, VC, LR, SOB, FOC, IV, and FM. Tables 1 and 2 also present the precision and success scores of all trackers for the above attributes. As shown in Table 1, our method performs 5.1%, 3.2%, 2.1%, 7.3%, 5.2%, 4.8%, 2.1%, 1.2%, 8.4%, and 3.1% better in these attributes compared to the second best trackers in precision scores at the 20-pixel threshold.

Similarly, in the success plots, our method ranks first among nine attributes, namely SV, POC, ARC, VC, LR, SOB, FOC, IV, and FM. As shown in Table 2, our method performs 2.2%, 2.2%, 2.9%, 1.5%, 4.2%, 0.5%, 0.6%, 4.1%, and 1.7% better in these attributes compared to the second best trackers in AUC-based success scores.

Figures 8 and 9 present the analysis of precision and success scores for all trackers with different attributes. As can be seen in these figures, our proposed method ranks first in general among all trackers. Therefore, it can be concluded that the proposed method has achieved better tracking performance compared to other state-of-the-art trackers.

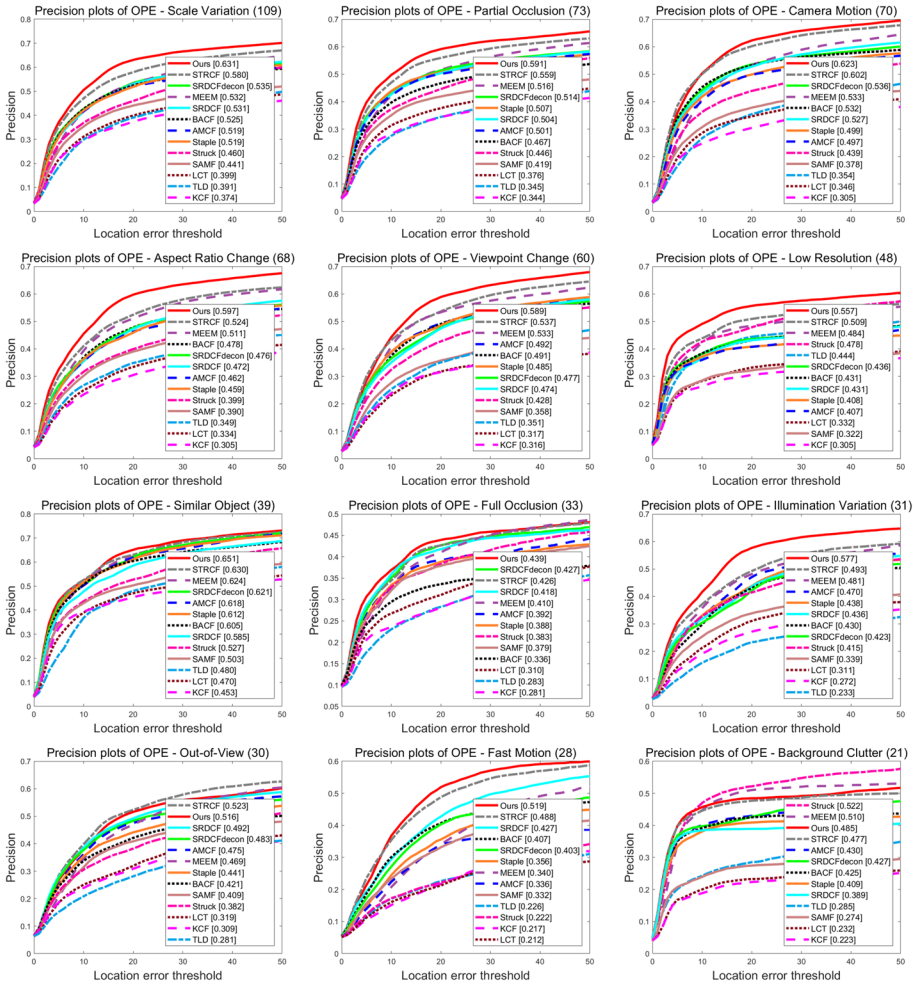

Fig. 6 Precision plot of different attributes of all trackers on 123 UAV sequences

4.4 Qualitative evaluation

For clearer visualization, Fig. 10 further exhibits the tracking results obtained by all trackers with several challenging sequences. Table 3 shows the number of frames and relevant attributes of these sequences. The tracking results show that the proposed method outperforms other state-of-the-art trackers.

4.5 Ablation study

To validate its effectiveness, our method is compared to itself with different modules enabled. The overall evaluation is presented in Table 4. With the response consistency (RC) and distractor repression (DR) modules added to the baseline (SRDCF), the performance is smoothly improved. Furthermore, our final tracker improves the baseline method by 5.3% and 9.8% in terms of success rate and precision criteria, respectively.

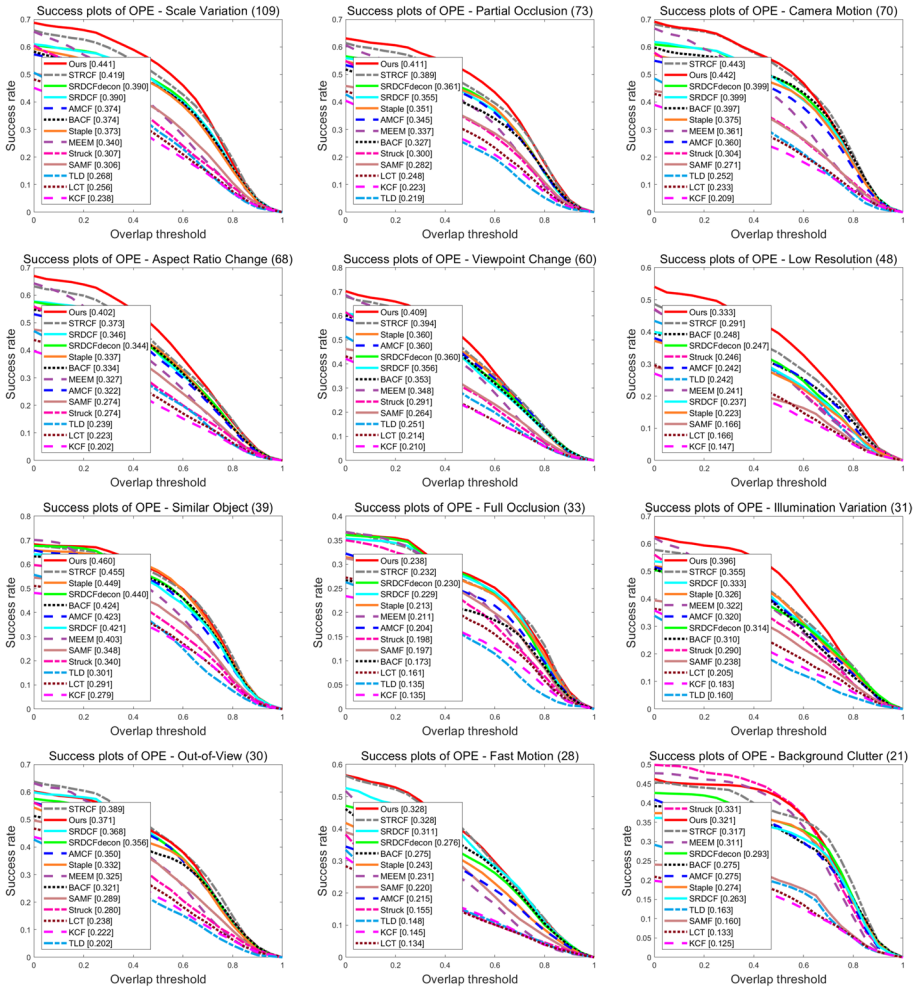

Fig. 7 Success plots of different attributes of all trackers on 123 UAV sequences

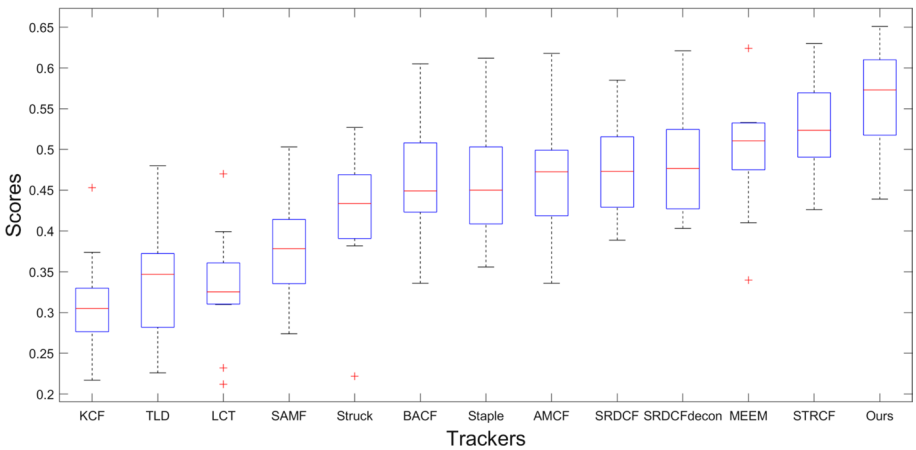

Fig. 8 Analysis of precision scores for all trackers with different attributes

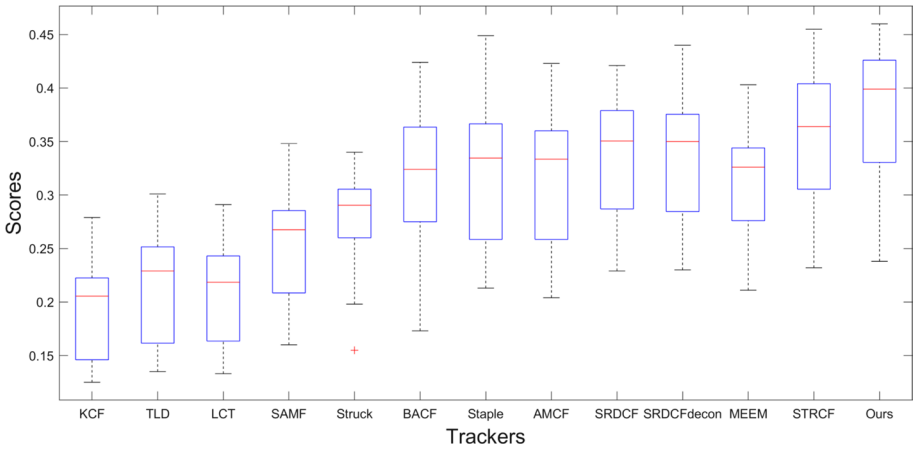

Fig. 9 Analysis of AUC success scores for all trackers with different attributes

Table 1 Precision scores (with a 20-pixel threshold) of the different attributes of all trackers on the UAV123@10fps dataset

Trackers	SV	POC	CM	ARC	VC	LR	SOB	FOC	IV	OV	FM	BC
KCF	0.374	0.344	0.305	0.305	0.316	0.305	0.453	0.281	0.272	0.309	0.217	0.223
TLD	0.391	0.345	0.354	0.349	0.351	0.444	0.480	0.283	0.233	0.281	0.226	0.285
LCT	0.399	0.376	0.346	0.334	0.317	0.332	0.470	0.310	0.311	0.319	0.212	0.232
SAMF	0.441	0.419	0.378	0.390	0.358	0.322	0.503	0.379	0.339	0.409	0.332	0.274
Struck	0.460	0.446	0.439	0.399	0.428	0.478	0.527	0.383	0.415	0.382	0.222	**0.522**
BACF	0.525	0.467	0.532	0.478	0.491	0.431	0.605	0.336	0.430	0.421	0.407	0.425
Staple	0.519	0.507	0.499	0.459	0.485	0.408	0.612	0.388	0.438	0.441	0.356	0.409
AMCF	0.519	0.501	0.497	0.462	0.492	0.407	0.618	0.392	0.470	0.475	0.336	0.430
SRDCF	0.531	0.504	0.527	0.472	0.474	0.431	0.585	0.418	0.436	*0.492*	*0.427*	0.389
SRDCFdecon	*0.535*	0.514	*0.536*	0.476	0.477	0.436	0.621	0.427	0.423	0.483	0.403	0.427
MEEM	0.532	*0.516*	0.533	*0.511*	*0.533*	*0.484*	*0.624*	0.410	*0.481*	0.469	0.340	0.510
STRCF	0.580	0.559	0.602	0.524	0.537	0.509	0.630	*0.426*	0.493	**0.523**	0.488	0.477
Ours	**0.631**	**0.591**	**0.623**	**0.597**	**0.589**	**0.557**	**0.651**	**0.439**	**0.577**	0.516	**0.519**	*0.485*

The top three ranking trackers in each attribute are marked in bold, underlined, and italic fonts, respectively

4.6 Parameters analysis

We investigate the parameter sensitivity of the distractor repression factor δ in (3), response-consistent constraint parameter γ in (4), and the number of local maximums for suppression N_d described in Sect. 3.2.1. We conduct experiments on our tracker using the UAV123@10fps dataset for different parameters. The precision and success scores are employed as evaluation criteria for tracking performance. We fix the other parameters while changing the value of the analyzed one. The values of both δ and N_d have an influence on the distractors repression and affect the tracking performance. In Fig. 11, we exhibit the precision and success scores of our tracker with various values in δ and N_d. Ranging from 0.663 to 0.673 and 0.468 to 0.476, the precision and success scores are slightly influenced by the value of N_d, when N_d is above 20. As only the top local maxima have significant interference, the suppression of other lower local maxima has little impact on the tracking performance. As for δ, when its

Table 2 Success scores (AUC) of the different attributes of all trackers on the UAV123@10fps dataset

Trackers	SV	POC	CM	ARC	VC	LR	SOB	FOC	IV	OV	FM	BC
KCF	0.238	0.223	0.209	0.202	0.210	0.147	0.279	0.135	0.183	0.222	0.145	0.125
TLD	0.268	0.219	0.252	0.239	0.251	0.242	0.301	0.135	0.160	0.202	0.148	0.163
LCT	0.256	0.248	0.233	0.223	0.214	0.166	0.291	0.161	0.205	0.238	0.134	0.133
SAMF	0.306	0.282	0.271	0.274	0.264	0.166	0.348	0.197	0.238	0.289	0.220	0.160
Struck	0.307	0.300	0.304	0.274	0.291	0.246	0.340	0.198	0.290	0.280	0.155	**0.331**
BACF	0.374	0.327	0.397	0.334	0.353	*0.248*	0.424	0.173	0.310	0.321	0.275	0.275
Staple	0.373	0.351	0.375	0.337	*0.360*	0.223	*0.449*	0.213	0.326	0.332	0.243	0.274
AMCF	0.374	0.345	0.360	0.322	*0.360*	0.242	0.423	0.204	0.320	0.350	0.215	0.275
SRDCF	*0.390*	0.355	*0.399*	*0.346*	0.356	0.237	0.421	*0.229*	*0.333*	*0.368*	0.311	0.263
SRDCFdecon	*0.390*	*0.361*	*0.399*	0.344	*0.360*	0.247	0.440	*0.230*	0.314	0.356	*0.276*	0.293
MEEM	0.340	0.337	0.361	0.327	0.348	0.241	0.403	0.211	0.322	0.325	0.231	0.311
STRCF	0.419	0.389	**0.443**	0.373	0.394	0.291	0.455	0.232	0.355	**0.389**	**0.328**	*0.317*
Ours	**0.441**	**0.411**	0.442	**0.402**	**0.409**	**0.333**	**0.460**	**0.238**	**0.396**	0.371	**0.328**	0.321

The top three ranking trackers in each attribute are marked in bold, underlined, and italic fonts, respectively

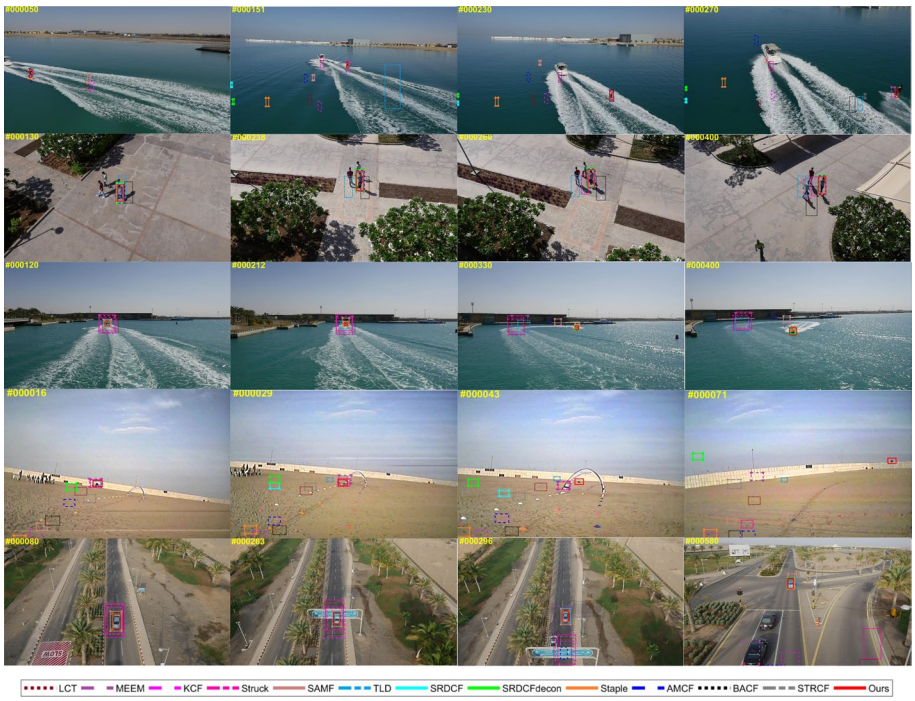

Fig. 10 Visualization of the tracking results of all trackers on five challenging sequences. (# Number on the left corner of each image denotes the frame index. From top to bottom are wakeboard3, group1_1, boat9, uav3, and car9 sequences)

value increases from 0.05 to 0.25, there is an increase in tracking performance. When δ varies from 0.25 to 1, both precision and the success scores decrease. Specifically, when δ is bigger than 0.35, the variation of δ has a relatively small impact on tracking performance, and the precision and success scores are mostly within the range of 0.655 to 0.641 and 0.466 to 0.457, respectively. When the local maximums are fixed, further increasing the value of factor δ does not improve the tracking performance.

Table 3 Typical UAV sequences

Video sequences	Number of frames	Attributes
wakeboard3	275	SV, ARC, LR, VC, CM
group1_1	445	SV, POC, SO
boat9	467	SV, ARC, LR, POC, VC
uav3	89	SV, LR, FM, CM
car9	627	SV, ARC, LR, FM, POC, CM, SO

Table 4 Ablation analysis on the UAV123@10fps dataset

Tracker	Ours	Baseline + DR	Baseline + RC	Baseline
Success scores (AUC)	0.476	0.465	0.466	0.423
Precision scores	0.673	0.648	0.654	0.575

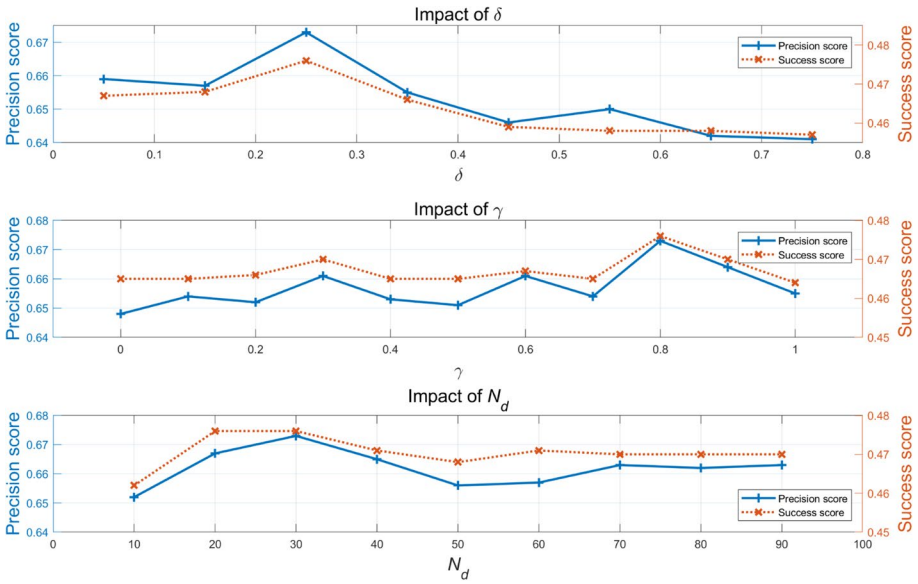

Fig. 11 Tracking performance of precision and success scores versus three varying parameters on the UAV123@10fps dataset

The response-consistent constraint parameter γ works when the abrupt appearance changes occur and is introduced as a temporal constraint. The value of γ varies from 0 to 1 with a step size of 0.1. Notably, our tracker with $\gamma = 0$ is the 'Baseline+DR' tracker. As γ gradually increases from 0, both precision and success scores exhibit an upward trend. When the value arrives $\gamma = 0.8$, both precision and success scores reach the maximum value (0.673 and 0.476). As γ continues to increase, both precision and success scores decrease. The introduction of response-consistent constraint maintains temporal smoothness, and further improves the tracking performance on the basis of distractors repression. To achieve satisfactory performance, we set δ, γ, and N_d, as 0.25, 0.8, and 30, respectively. All parameters are the same for the overall experiments.

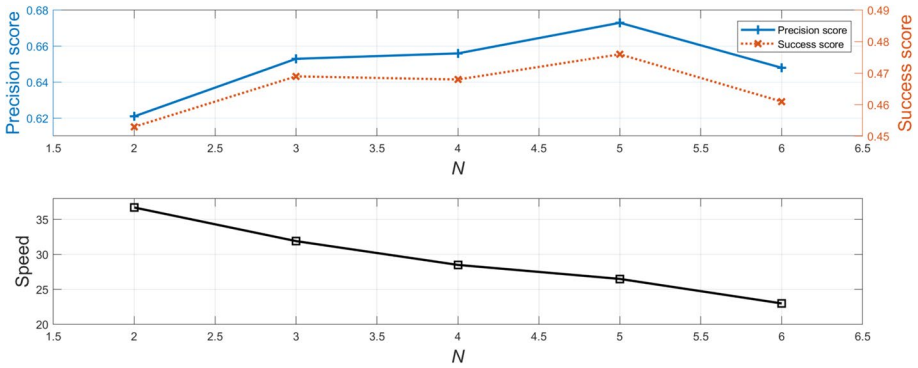

Fig. 12 Tracking performance of precision score, success score, and speed on the UAV123@10fps dataset with different numbers of ADMM iterations

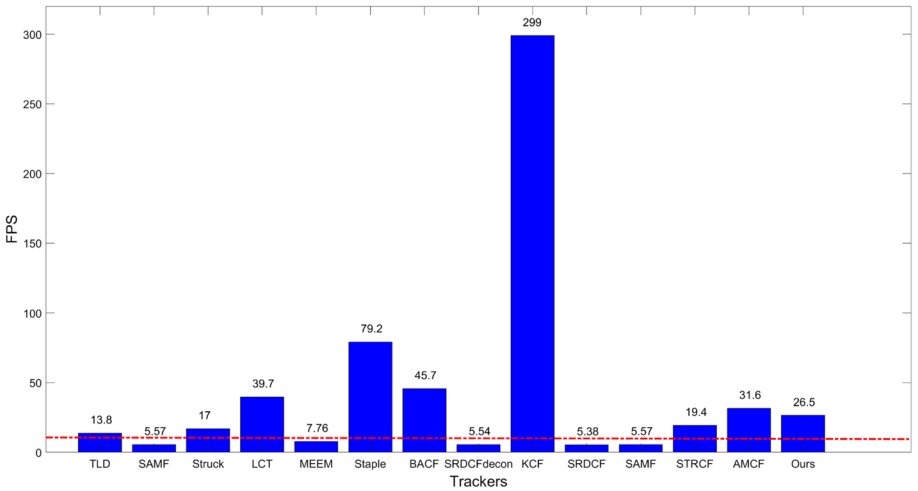

Fig. 13 FPS comparison results of all trackers on 123 UAV sequences. The red dashed line represents the capturing speed of the original UAV sequences

In addition, we also conducted a comparative experiment on the number of iterations of ADMM. Figure 12 presents the precision score, success score, and speed under different numbers of iterations. By comprehensively considering the accuracy and tracking speed, we finally set the number of iterations of ADMM to 5.

4.7 Tracking speed analysis

Figure 13 presents the tracking speed comparison results in terms of the number of processed frames per second (FPS), for all trackers on 123 challenging UAV sequences. The capturing speed of the original UAV sequences is 10 FPS. From Fig. 13, the speed of our method ranks sixth among all trackers and is greater than 10 FPS. Therefore, our method can meet the real-time requirement in UAV tracking tasks.

5 Conclusions

In this study, spatial regularized correlation filters with response consistency and distractor repression were proposed in the context of UAV-based tracking. By exploiting the response-consistent constraint to limit the correlation response variations, the temporal consistency across adjacent frames was pursued to enhance the discriminative power of the proposed appearance model. In addition, the distractor-repressed constraint was incorporated in the learning phase. It served as a dynamic spatial constraint to suppress the influence of distractors. An ADMM algorithm was developed to solve the appearance model efficiently. Spatial and temporal cues in response maps were explored and encoded in the conventional CF framework to facilitate UAV tracking in complex scenarios and boost overall performance. Comparative experimental results over 123 challenging UAV sequences demonstrated that the proposed method outperforms 12 state-of-the-art trackers in terms of accuracy, robustness, and efficiency.

Abbreviations

CF	Correlation filter
UAV	Unmanned aerial vehicle
ADMM	Alternating direction method of multipliers
FPS	Frames per second
DFT	Discrete Fourier transform
CN	Color name
HOG	Histogram of oriented gradient
OPE	One-pass evaluation
AUC	Area under the curve
RC	Response consistency
DR	Distractor repression

Acknowledgements
The author thanks the editor and anonymous reviewers for their helpful comments and valuable suggestions.

Author contributions
WZ received the B.S. degree in network engineering from Shaanxi University of Science and Technology, Xi'an, China, in 2010, the M.S. degree in computer science and technology from Shaanxi Normal University, Xi'an, China, in 2013, and the Ph.D. degree in computer science and technology from Northwest University, Xi'an, China, in 2019. She is currently a lecturer of the Department of Computer Science, Baoji University of Arts and Sciences. Her research interests include object tracking and machine learning.The author have read and approved the final version of the manuscript.

Funding
This work was supported by the scientific research program of the Education Department of Shaanxi Provincial Government (20JK0487) and the Key R &D Program of the Shaanxi Province of China (No.2022GY-071).

Availability of data and materials
Experiments were conducted using the public UAV benchmark dataset UAV123@10FPS to evaluate the tracking performance. The UAV123@10fps dataset can be found in https://cemse.kaust.edu.sa/ivul/uav123.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 12 November 2022 Accepted: 1 March 2023
Published online: 15 March 2023

References

1. R. Li, M. Pang, C. Zhao, G. Zhou, L. Fang, Monocular long-term target following on uavs, in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2016, Las Vegas, NV, USA, June 26–July 1, 2016*, p. 29–37 (2016). https://doi.org/10.1109/CVPRW.2016.11
2. C. Fu, A. Carrio, M.A. Olivares-Mendez, P. Campoy, Online learning-based robust visual tracking for autonomous landing of unmanned aerial vehicles, in *2014 International Conference on Unmanned Aircraft Systems (ICUAS), Orlando, FL, USA*, p. 649–655 (2014)
3. S. Lin, M.A. Garratt, A.J. Lambert, Monocular vision-based real-time target recognition and tracking for autonomously landing an UAV in a cluttered shipboard environment. Auton. Robots **41**(4), 881–901 (2017)
4. C. Fu, A. Carrio, M.A. Olivares-Méndez, R. Suarez-Fernandez, P.C. Cervera, Robust real-time vision-based aircraft tracking from unmanned aerial vehicles, in *2014 IEEE International Conference on Robotics and Automation, ICRA 2014, Hong Kong, China, May 31–June 7, 2014*, p. 5441–5446 (2014). https://doi.org/10.1109/ICRA.2014.6907659
5. C. Fu, B. Li, F. Ding, F. Lin, G. Lu, Correlation filters for unmanned aerial vehicle-based aerial tracking: a review and experimental evaluation. IEEE Geosci. Remote Sens. Mag. **10**(1), 125–160 (2021)
6. D.S. Bolme, J.R. Beveridge, B.A. Draper, Y.M. Lui, Visual object tracking using adaptive correlation filters, in *The Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010, San Francisco, CA, USA, 13–18 June 2010*, p. 2544–2550 (2010). https://doi.org/10.1109/CVPR.2010.5539960
7. J.F. Henriques, R. Caseiro, P. Martins, J. Batista, High-speed tracking with kernelized correlation filters. IEEE Trans. Pattern Anal. Mach. Intell. **37**(3), 583–596 (2015)
8. M. Danelljan, G. Häger, F.S. Khan, M. Felsberg, Learning spatially regularized correlation filters for visual tracking, in *2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7–13, 2015*, p. 4310–4318 (2015). https://doi.org/10.1109/ICCV.2015.490
9. H.K. Galoogahi, A. Fagg, S. Lucey, Learning background-aware correlation filters for visual tracking, in *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22–29, 2017*, p. 1144–1152 (2017). https://doi.org/10.1109/ICCV.2017.129
10. F. Li, C. Tian, W. Zuo, L. Zhang, M. Yang, Learning spatial-temporal regularized correlation filters for visual tracking, in *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18–22, 2018*, p. 4904–4913 (2018). https://doi.org/10.1109/CVPR.2018.00515
11. Z. Huang, C. Fu, Y. Li, F. Lin, P. Lu, Learning aberrance repressed correlation filters for real-time UAV tracking, in *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27–November 2, 2019*, p. 2891–2900 (2019). https://doi.org/10.1109/ICCV.2019.00298
12. C. Fu, F. Ding, Y. Li, J. Jin, C. Feng, Dr^{2}track: towards real-time visual tracking for UAV via distractor repressed dynamic regression, in *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2020, Las Vegas, NV, USA, October 24, 2020–January 24, 2021*, p. 1597–1604 (2020). https://doi.org/10.1109/IROS45743.2020.9341761
13. J.F. Henriques, R. Caseiro, P. Martins, J.P. Batista, Exploiting the circulant structure of tracking-by-detection with kernels, in *Computer Vision—ECCV 2012—12th European Conference on Computer Vision, Florence, Italy, October 7–13, 2012, Proceedings, Part IV. Lecture Notes in Computer Science*, vol. 7575, p. 702–715 (2012). https://doi.org/10.1007/978-3-642-33765-9_50
14. T. Liu, G. Wang, Q. Yang, Real-time part-based visual tracking via adaptive correlation filters, in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7–12, 2015*, p. 4902–4912 (2015). https://doi.org/10.1109/CVPR.2015.7299124
15. Y. Li, J. Zhu, S.C.H. Hoi, Reliable patch trackers: Robust visual tracking by exploiting reliable patches, in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7–12, 2015*, p. 353–361 (2015). https://doi.org/10.1109/CVPR.2015.7298632
16. C. Fu, Y. Zhang, Z. Huang, R. Duan, Z. Xie, Part-based background-aware tracking for UAV with convolutional features. IEEE Access **7**, 79997–80010 (2019)
17. Y. Li, J. Zhu, A scale adaptive kernel correlation filter tracker with feature integration, in *Computer Vision—ECCV 2014 Workshops—Zurich, Switzerland, September 6–7 and 12, 2014, Proceedings, Part II. Lecture Notes in Computer Science*, vol. 8926, p. 254–265 (2014). https://doi.org/10.1007/978-3-319-16181-5_18
18. M. Danelljan, G. Häger, F.S. Khan, M. Felsberg, Accurate scale estimation for robust visual tracking, in *British Machine Vision Conference, BMVC 2014, Nottingham, UK, September 1–5, 2014* (2014). http://www.bmva.org/bmvc/2014/papers/paper038/index.html
19. L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, P.H.S. Torr, Staple: complementary learners for real-time tracking, in *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27–30, 2016*, p. 1401–1409 (2016). https://doi.org/10.1109/CVPR.2016.156
20. M. Danelljan, G. Häger, F.S. Khan, M. Felsberg, Convolutional features for correlation filter based visual tracking, in *2015 IEEE International Conference on Computer Vision Workshop, ICCV Workshops 2015, Santiago, Chile, December 7–13, 2015*, p. 621–629 (2015). https://doi.org/10.1109/ICCVW.2015.84
21. Y. Li, C. Fu, Z. Huang, Y. Zhang, J. Pan, Intermittent contextual learning for keyfilter-aware UAV object tracking using deep convolutional feature. IEEE Trans. Multimed. **23**, 810–822 (2021)
22. C. Ma, X. Yang, C. Zhang, M. Yang, Long-term correlation tracking, in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7–12, 2015*, p. 5388–5396 (2015). https://doi.org/10.1109/CVPR.2015.7299177
23. H.K. Galoogahi, T. Sim, S. Lucey, Correlation filters with limited boundaries, in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7–12, 2015*, p. 4630–4638 (2015). https://doi.org/10.1109/CVPR.2015.7299094
24. A. Lukezic, T. Vojír, L.C. Zajc, J. Matas, M. Kristan, Discriminative correlation filter tracker with channel and spatial reliability. Int. J. Comput. Vis. **126**(7), 671–688 (2018)
25. M. Danelljan, G. Häger, F.S. Khan, M. Felsberg, Adaptive decontamination of the training set: a unified formulation for discriminative visual tracking, in *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27–30, 2016*, p. 1430–1438 (2016). https://doi.org/10.1109/CVPR.2016.159

26. Y. Li, C. Fu, F. Ding, Z. Huang, J. Pan, Augmented memory for correlation filters in real-time UAV tracking, in *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2020, Las Vegas, NV, USA, October 24, 2020–January 24, 2021*, p. 1559–1566 (2020). https://doi.org/10.1109/IROS45743.2020.9341595
27. W. Zhang, B. Kang, S. Zhang, Enhanced occlusion handling and multipeak redetection for long-term object tracking. J. Electron. Imaging **27**(03), 033005 (2018)
28. K.B. Petersen, M.S. Pedersen, The Matrix Cookbook. Technical University of Denmark (2012). http://www2.compute.dtu.dk/pubdb/pubs/3274-full.html
29. F.S. Khan, R.M. Anwer, J. van de Weijer, A.D. Bagdanov, M. Vanrell, A.M. López, Color attributes for object detection, in *2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, June 16–21, 2012*, p. 3306–3313 (2012). https://doi.org/10.1109/CVPR.2012.6248068
30. N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), 20–26 June 2005, San Diego, CA, USA*, p. 886–893 (2005). https://doi.org/10.1109/CVPR.2005.177
31. M. Mueller, N. Smith, B. Ghanem, A benchmark and simulator for UAV tracking, in *Computer Vision—ECCV 2016—14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I. Lecture Notes in Computer Science*, vol. 9905, p. 445–461 (2016). https://doi.org/10.1007/978-3-319-46448-0_27
32. Y. Wu, J. Lim, M. Yang, Object tracking benchmark. IEEE Trans. Pattern Anal. Mach. Intell. **37**(9), 1834–1848 (2015)
33. Z. Kalal, K. Mikolajczyk, J. Matas, Tracking-learning-detection. IEEE Trans. Pattern Anal. Mach. Intell. **34**(7), 1409–1422 (2012). https://doi.org/10.1109/TPAMI.2011.239
34. S. Hare, S. Golodetz, A. Saffari, V. Vineet, M. Cheng, S.L. Hicks, P.H.S. Torr, Struck: structured output tracking with kernels. IEEE Trans. Pattern Anal. Mach. Intell. **38**(10), 2096–2109 (2016)
35. J. Zhang, S. Ma, S. Sclaroff, MEEM: robust tracking via multiple experts using entropy minimization, in *Computer Vision—ECCV 2014—13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part VI. Lecture Notes in Computer Science*, vol. 8694, p. 188–203 (2014). https://doi.org/10.1007/978-3-319-10599-4_13

Publisher's Note