

Received November 21, 2019, accepted December 9, 2019, date of publication December 13, 2019, date of current version December 31, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2959406

An Anti-Drift Background-Aware Correlation Filter for Visual Tracking in Complex Scenes

SHANSHAN LUO^{1,2}, BAOQING LI¹, AND XIAOBING YUAN¹

¹Science and Technology on Microsystem Laboratory, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 201800, China

²Shanghai Institute of Microsystem and Information Technology, University of Chinese Academy of Sciences, Beijing 100049, China

Corresponding author: Baoqing Li (sinoiot@mail.sim.ac.cn)

ABSTRACT The model drift problem is an inevitable problem for online visual tracking. Model drift can amplify the false tracking result over time and lead to failed tracking finally. Background-aware correlation filter (BACF) tracker obtains the training samples from negative background patches instead of shifted foreground patches and can mitigate the drifting to some extent. However, in complex scenes, such as occlusion, deformation, the inappropriate update of BACF may lead to model drift. We propose an anti-drift background-aware correlation filter via introducing the temporal consistency constraint into BACF. That is, we combine the excellent ability of BACF to distinguish the foreground and background with the ability of the temporal consistency constraint to stabilize model changes, and improve anti-drift performance of BACF. To improve computation efficient, we offer a fast algorithm via alternative Direction Method of Multipliers (ADMM) in the frequency domain. Besides, we design a simple yet effective adaptive feature channel selection method, which can further improve the success rate and precision of the tracker in complex scenes. Our proposed tracker with hand-crafted features achieves a gain of 2.2%, 3.3%, and 5.3% in AUC success rate on OTB-2013, OTB-2015, and Temple Color-128 dataset, respectively. Our proposed tracker with deep features achieves AUC success rate of 69.1% on OTB-2013 dataset. Besides, our proposed tracker with hand-crafted features achieves a gain of 3.94% in EAO on VOT-2018 dataset. Moreover, the proposed tracker performs favorably against several representative state-of-the-art methods regarding precision and success rate.

INDEX TERMS Visual tracking, correlation filter, background-aware, temporal consistency, channel selection, model drift.

I. INTRODUCTION

Object tracking is an extensively studied problem in computer vision and has numerous advanced applications, such as precision guidance, human-computer interaction, automatic driving, and so on [5]–[7]. Its primary purpose is to continuously obtain the target's related motion parameters through analyzing a sequence of images given only its bounding box of the initial frame. Object tracking is not a simple task in actual application, because it inevitably faces the interference of complex scenes, such as non-rigid deformation, scale change, motion blur, illumination variation, occlusion, and so on. Although object tracking has made much progress in recent years, it remains a very challenging problem.

The associate editor coordinating the review of this manuscript and approving it for publication was Huazhu Fu¹.

Arguably, most existing methods of tracking can be classified into two categories: generative methods and discriminative methods [24]–[35]. Traditional generative tracking methods are trained based on object appearance without considering background information. By contrast, discriminative tracking methods consider both of them, usually treat tracking as a binary classification task which distinguishes the object from its surrounding background. Compared with the generative methods, the discriminative methods have more attention because the discriminant tracking methods use the background information in training and can achieve more accurate target position estimation. Recently, the trackers based discriminative correlation filter (DCF) [8]–[20] receive widespread attention due to their excellent accuracy and high speed, and have shown state-of-the-art performance in many standard benchmarks [1]–[4]. The DCF trackers belong to the discriminative tracking methods. They obtain the target

sample by cyclic shifts in a search window area instead of dense sampling. Taking the advantage of the property that the cyclic matrix can be diagonalize with the discrete fourier transform, the DCF trackers can transform complicated spatial correlation operation into efficient element wise operation in the Fourier domain and achieve extremely high tracking speed.

Although the DCF trackers have been widely used because of their excellent accuracy and high speed, the model drift still is their inevitable problem. In Fig. 2, we initialize the tracker in the first frame and let it track the target until the end of the sequence. When the average overlap is small, the tracker occurs model drift. And when the average overlap is zero, we consider tracking a failure. As shown in Fig. 2, model drift can amplify the false result over time and lead to failed tracking finally. The background-aware correlation filter (BACF) [20] tracker obtains the training samples from negative background patches instead of shifted foreground patches and can mitigate the drifting to some extent. BACF tracker has an outstanding performance among the state-of-the-art trackers because of the superior accuracy and real-time performance. However, in complex scenes, such as occlusion, deformation, the inappropriate update of BACF may lead to model drift.

In this paper, we focus on the problem of model drift in BACF tracker. Motivated by [36] that temporal consistency effectively enforces temporal consistency for visual tracking, we consider introducing the temporal consistency constraint into background-aware correlation filter. The temporal consistency constraint can inhibit the appearance variations of a target caused by the use of unstable temporal updating models for filters. Under the constraint, the filter model is updated around its historical value. Here, we consider combining the excellent ability of BACF to distinguish the foreground and background with the ability of the temporal consistency constraint to stabilize model changes. The problem of model drift can be mitigated effectively. In the optimization process, the problem can be solved efficiently via Alternative Direction Method of Multipliers [38] and fully carried out in the frequency domain to speed up. In practical application, different feature channels have different degrees of discrimination to the target. Some feature channels lead to a reduction in performance, because they are not useful to distinguish between target and backgrounds. Therefore, we design a simple yet effective adaptive feature channel selection method, which can mitigate model drift as well. We exploit two evaluation metrics (success rate and precision) to quantify the anti-drift capability of the tracker. Extensive experiments on the OTB-2013 [1], OTB-2015 [2], Temple Color-128 [3] and VOT2018 [4] datasets demonstrate that the proposed tracker outperforms the baseline BACF conspicuously in terms of success rate and precision. Especially in some complex scenes, the advantages of our proposed tracker are more obvious. Fig. 1 shows the qualitative comparison of our proposed tracker in occlusion, deformation, and background clutter. Moreover, the proposed tracker performs

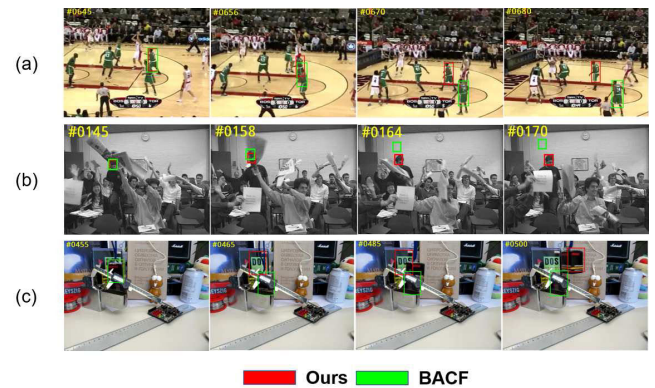


FIGURE 1. Qualitative comparison of our proposed tracker with BACF on the basketball, freeman, box videos, under occlusion, deformation, and background clutter.

favorably against several representative state-of-the-art methods regarding precision and success rate.

The main contributions of our work are summarized as follows:

(1) We introduce the temporal consistency constraint into BACF and propose a new tracker named BACF-TC. To improve computation efficient, we also offer a fast algorithm via alternative Direction Method of Multipliers in the frequency domain. The extensive experiments show that this method can mitigate model drift effectively.

(2) We design a simple yet effective adaptive feature channel selection method and obtain the response map for each channel by correlation calculation of the initial frame information and filter template. The value of the PSR indirectly reflects the ability of feature channel to distinguish between target and backgrounds. The experiment demonstrates that this method can mitigate model drift as well.

II. RELATED WORK

A. BASIC TRACKING METHODS

Generally, there exist two typical ways to solve the problem of object tracking: generative tracking methods and discriminative tracking methods. Since the proposal of the Lucas-Kanade algorithm [24]–[26], generative methods become popular in many computer vision tasks. The fundamental principle of generative tracking methods is to search for the most similar region to the target object within a neighborhood without considering background information. According to the target representation type, generative methods can be divided as follows: subspace representation and sparse representation. Generative methods based subspace representation [27], [28] cannot effectively resist the interference caused by occlusion. Generative methods based sparse representation [29]–[32] have excellent performance in case of occlusion, but most of them cannot realize the requirement of real-time.

Discriminant tracking methods regard tracking problem as a binary classification problem. They are distinguishing the object from its surrounding background via the object appearance and background information. The discriminant tracking

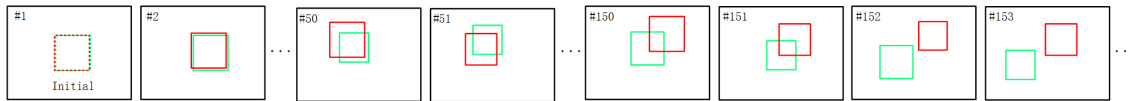


FIGURE 2. The illustration of model drift during tracking (the green box represents the ground-truth target location, and the red color denotes the tracker outputs. Dotted boxes represent the initialization of the tracker).

methods generally have a better performance compared with the generative tracking methods, because the discriminant tracking methods use the background information in training and can achieve more accurate target position estimation. The discriminant tracking methods are based on a discriminatively trained classifier, such as SVM [33], Adaboost [34], semi-supervised boosting [35], and so on. More recently, the discriminant tracking methods based discriminative correlation filters (DCF) are the latest development direction of the discriminative tracking methods because of its high speed and excellent performance.

B. TRACKING METHODS BASED DISCRIMINATIVE CORRELATION FILTERS

Since the proposal of MOSSE algorithm [8] which first adopt adaptive correlation filters to track, tracking methods based DCF become widely used because of high accuracy and computational efficiency. Many improved algorithms based DCF have been proposed. DSST [9] and SAMF [10] solve the target scale change issue, and achieve accurate scale adaptive visual tracking. STAPLE [11] is inherently robust to both color changes and deformations combining two image patch representations to learn a model. Danelljan *et al.* [12] introduce the CNN convolutional features into discriminative correlation filter based tracking frameworks. In [13], Ke Nai *et al.* learn the multiple correlation filters to capture different appearance patterns of the target object during the tracking.

Several works are utilizing spatial or temporal constraints to mitigate model drift and improve the overall performance of the tracker. SRDCF tracker [14] introduces a spatial regularization component in the learning to penalize correlation filter coefficients depending on their spatial location. Subsequently, C-COT [15], ECO [16] introduce a novel formulation for training continuous convolution filters and employ an implicit interpolation model to pose the learning problem in the continuous spatial domain. In [17], Kenan Dai *et al.* introduce an adaptive spatial regularization which could be effectively learned with respect to a specific object being tracked and result in more reliable filter coefficients during the tracking process. These methods only consider the boundary effects but do not consider the smoothing of model updates. In [18], Feng Li *et al.* propose STRCF tracker which incorporate both temporal and spatial regularization constraints, and achieve superior performance over SRDCF in terms of accuracy and speed. Xu *et al.* [19] design a low-dimensional discriminative manifold space by exploiting temporal consistency. These methods can mitigate

model drift effectively by considering the smoothing of model updates, but do not use real negative samples information. BACF [20] is learning/updating filter capable of from real negative examples densely extracted from the background, instead of shifted foreground patches. Therefore, BACF tracker can improve accuracy effectively, compared to prior CF-based trackers. However, in some complex scenes, such as occlusion, deformation, BACF tracker may occur model drift. In this paper, we propose an anti-drift background-aware correlation filter via introducing the temporal consistency constraint into BACF.

III. PROPOSED APPROACH

In this section, we first briefly introduce the background-aware correlation filter. Next, we describe how to introduce the global temporal consistency constraint to the background-aware correlation filter and present the proposed algorithm in detail.

A. BASELINE METHOD

Essentially, in the correlation filter (CF) framework, tracking is cast as a binary classification task which distinguishes the object from its surrounding background. In the training phase of CF framework, we need to learn a multi-channel correlation filter h from the sample x of the target appearance. Generally, we obtain the target sample x by cyclic shifts in a search window area. As the cyclic matrix can be diagonalized with the discrete Fourier transform, the CF tracker can processes each patch sample rapidly.

The CF in the spatial domain is formulated by minimizing the following objective:

$$\begin{aligned}
 E(h) &= \frac{1}{2} \left\| y - \sum_{l=1}^K h_l \star x_l \right\|_2^2 + \frac{\lambda}{2} \sum_{l=1}^K \|h_l\|_2^2 \\
 &= \frac{1}{2} \sum_{j=1}^D \left\| y(j) - \sum_{l=1}^K h_l^\top x_l[\Delta\tau_j] \right\|_2^2 + \frac{\lambda}{2} \sum_{l=1}^K \|h_l\|_2^2, \quad (1)
 \end{aligned}$$

where x_l and h_l refer to the l th channel of the vectorized image and filter respectively. $[\Delta\tau_1, \dots, \Delta\tau_D]$ represents the set of all circular shifts for a signal of length D . y is the desired correlation response. $y(j)$ refers to the j th element of the desired correlation response. λ is a regularization. K represents the number of feature channels. \star refers to the spatial correlation operator.

Eq. (1) is a linear least squares problem. We can obtain a closed-form solution in the Fourier domain using the

Parseval's formula,

$$\hat{h}_l^* = \frac{\hat{y} \odot \hat{x}_l^*}{\sum_{i=1}^K \hat{x}_i^* \odot \hat{x}_i + \lambda I}. \quad (2)$$

Here, the hat symbol denotes the Discrete Fourier Transform of corresponding variable, such as $\hat{x} = DFT(x)$. * expresses the complex-conjugate. \odot refers to the element-wise product.

These trackers based upon generic CF filter have an excellent performance to cope with the variations in lighting, scale, pose, and non-rigid deformations and can operate at high speed. However, they probably have over-fitting problems because of the training samples generated through the $[\Delta\tau_j]$ operator and can not classify the target from real non-target patches very well. In order to use the surrounding background part as negative samples at the training stage, Galoogahi *et al.* [20] propose the background-aware correlation filter (BACF) to directly learn from background patches. The BACF is formulated by minimizing the following objective:

$$E(h) = \frac{1}{2} \sum_{j=1}^D \left\| y(j) - \sum_{l=1}^K h_l^T P x_l[\Delta\tau_j] \right\|_2^2 + \frac{\lambda}{2} \sum_{l=1}^K \|h_l\|_2^2, \quad (3)$$

where P is a $D \times T$ binary matrix which crops the mid D elements of signal x_k , and $T \gg D$. T denotes the length of x . Furthermore, $x_k \in \mathbb{R}^T$, $y \in \mathbb{R}^T$, and $h \in \mathbb{R}^D$.

B. ANTI-DRIFT BACF VIA INTRODUCING THE TEMPORAL CONSISTENCY CONSTRAINT

We propose to learn the anti-drift BACF via introducing the temporal consistency constraint (BACF-TC) by minimizing the objective function as follows:

$$E(h, g) = \frac{1}{2} \sum_{j=1}^T \left\| y(j) - \sum_{l=1}^K g_l^T x_l[\Delta\tau_j] \right\|_2^2 + \frac{\lambda_1}{2} \sum_{l=1}^K \|h_l\|_2^2 + \frac{\lambda_2}{2} \sum_{l=1}^K \|g_l - g_{model,l}\|_2^2$$

s.t. $g = P^T h,$ (4)

where P is a $D \times T$ binary matrix which crops the mid D elements of signal x_l , and $T \gg D$. T is the length of x . h represents the parameters of the filter at the current moment. The model parameter g_{model} estimates from previous frames. $g_{model} = (1 - \varepsilon)g_{model} + \varepsilon g$, where ε denotes the learning rate. λ_1 and λ_2 are the regularization parameters. The transpose operator denotes the conjugate transpose of a complex vector or a matrix. K is the number of feature channels. The last term in the formula refers to the global temporal consistency constraint. By introducing the global temporal consistency constraint, our proposed tracker considers the target's previous appearance and imposes smooth variation between consecutive frames.

Similar to the general CF trackers [9], our proposed approach transforms its function into the frequency domain

for high computational efficiency. Hence, Eq. 4 can be expressed in the frequency domain as:

$$E(h, \hat{g}) = \frac{1}{2} \|\hat{y} - \hat{X}\hat{g}\|_2^2 + \frac{\lambda_1}{2} \|h\|_2^2 + \frac{\lambda_2}{2} \|\hat{g} - \hat{g}_{model}\|_2^2$$

s.t. $\hat{g} = \sqrt{T} (FP^T \otimes I_K) h.$ (5)

Here, we define the concatenated matrix $\hat{X} = [diag(\hat{x}_1)^T, \dots, diag(\hat{x}_K)^T]$, $h = [h_1^T, \dots, h_K^T]$ and $\hat{g} = [\hat{g}_1^T, \dots, \hat{g}_K^T]$. I_K refers to a $K \times K$ identify matrix, and \otimes denotes the Kronecker product. The hat symbol denotes the Discrete Fourier Transform of corresponding variable, such that $\hat{y} = \sqrt{T} F y$. F indicates the DFT matrix of T dimensional vectorized signal.

Employing Augmented Lagrangian method(ALM) [37], Eq. 5 can be formulated as:

$$\mathcal{S}(\hat{g}, h, \hat{\zeta}) = \frac{1}{2} \|\hat{y} - \hat{X}\hat{g}\|_2^2 + \frac{\lambda_1}{2} \|h\|_2^2 + \frac{\lambda_2}{2} \|\hat{g} - \hat{g}_{model}\|_2^2 + \hat{\zeta}^T (\hat{g} - \sqrt{T} (FP^T \otimes I_K) h) + \frac{\mu}{2} \|\hat{g} - \sqrt{T} (FP^T \otimes I_K) h\|_2^2, \quad (6)$$

where $\mu > 0$ is the corresponding penalty factor, and $\hat{\zeta}^T = [\hat{\zeta}_1^T, \dots, \hat{\zeta}_K^T]^T$ denotes the complex Lagrangian multiplier.

Here, our objective is to minimize $\mathcal{S}(\hat{g}, h, \hat{\zeta})$, and the model in Eq. 6 is convex. If we adopt the Lagrange Multiplier Method to solve, the augmented Lagrangian is minimized jointly with respect to two primal variables in Eq. 7.

$$\begin{cases} (\hat{g}, h) = \arg \min_{\hat{g}, h} \mathcal{S}(\hat{g}, \hat{\zeta}, \mu) \\ \hat{\zeta} = \arg \min_{\hat{\zeta}} \mathcal{S}(h, \hat{g}, \mu). \end{cases} \quad (7)$$

The common methods to solve the optimization of unconstrained multivariable functions are as follows: Stochastic Gradient Descent method (SGD), Newton's method, and so on. However, these iterative methods are time consuming. The alternating direction method of multipliers (ADMM) [38] is an algorithm that solves convex optimization problems by breaking them into smaller pieces, each of which are then easier to handle.

Here, we adopt ADMM algorithm to solve the following subproblems alternatingly:

$$\begin{cases} h = \arg \min_h \mathcal{S}(\hat{g}, \hat{\zeta}, \mu) \\ \hat{g} = \arg \min_{\hat{g}} \mathcal{S}(h, \hat{\zeta}, \mu) \\ \hat{\zeta} = \arg \min_{\hat{\zeta}} \mathcal{S}(h, \hat{g}, \mu). \end{cases} \quad (8)$$

We solve each subproblem separately as follows,

Subproblem h :

$$h = \arg \min_h \left\{ \frac{\lambda_1}{2} \|h\|_2^2 + \hat{\zeta}^\top \left(\hat{g} - \sqrt{T} \left(FP^\top \otimes I_K \right) h \right) + \frac{\mu}{2} \left\| \hat{g} - \sqrt{T} \left(FP^\top \otimes I_K \right) h \right\|_2^2 \right\}. \quad (9)$$

To solve for h , we set its complex gradient $\nabla_h \mathcal{L}$ to 0. Finally, we obtain the optimal solution as follows:

$$h = (\mu + \lambda_1)^{-1} (\mu g + \zeta), \quad (10)$$

where g and ζ are defined as $g = \frac{1}{\sqrt{T}} (PF^\top \otimes I_K) \hat{g}$ and $\zeta = \frac{1}{\sqrt{T}} (PF^\top \otimes I_K) \hat{\zeta}$.

Subproblem \hat{g} :

$$\hat{g} = \arg \min_{\hat{g}} \left\{ \frac{1}{2} \left\| \hat{y} - \hat{X} \hat{g} \right\|_2^2 + \frac{\lambda_2}{2} \left\| \hat{g} - \hat{g}_{model} \right\|_2^2 + \hat{\zeta}^\top \left(\hat{g} - \sqrt{T} \left(FP^\top \otimes I_K \right) h \right) + \frac{\mu}{2} \left\| \hat{g} - \sqrt{T} \left(FP^\top \otimes I_K \right) h \right\|_2^2 \right\}. \quad (11)$$

Here, we can express \hat{g} as T independent objectives $\hat{g}(t)$, and $\hat{g}(t) = [conj(\hat{g}_1(t)), \dots, conj(\hat{g}_K(t))]$, ($t = 1, \dots, T$).

$$\hat{g}(t) = \arg \min_{\hat{g}(t)} \left\{ \frac{1}{2} \left\| \hat{y}(t) - \hat{x}(t)^\top \hat{g}(t) \right\|_2^2 + \frac{\lambda_2}{2} \left\| \hat{g}(t) - \hat{g}_{model} \right\|_2^2 + \hat{\zeta}(t)^\top \left(\hat{g}(t) - \hat{h}(t) \right) + \frac{\mu}{2} \left\| \hat{g}(t) - \hat{h}(t) \right\|_2^2 \right\}, \quad (12)$$

where $\hat{x}(t) = [\hat{x}_1(t), \dots, \hat{x}_K(t)]$, $\hat{h}_k(t) = \sqrt{T} FP^\top h_k$, and $\hat{h}(t) = [\hat{h}_1(t), \dots, \hat{h}_K(t)]$.

We acquire the optimal solution by solving Eq. 12 and use Sherman-Morrison formula [39] to accelerate computation. Therefore, the optimal solution can be expressed as follows:

$$\hat{g}(t) = a_1 \left(\frac{1}{T} \hat{y}(t) \hat{x}(t) - \hat{\zeta}(t) + \mu \hat{h}(t) + \lambda_2 \hat{g}_{model} \right) - a_2 \left(\frac{1}{T} \hat{y}(t) \hat{s}_x(t) - \hat{s}_\zeta(t) + \mu \hat{s}_h(t) + \lambda_2 \hat{s}_{g_{model}}(t) \right), \quad (13)$$

where $a_1 = \frac{1}{\lambda_2 + \mu}$, $a_2 = \frac{\hat{x}(t)}{(\lambda_2 + \mu)b}$, $\hat{s}_x(t) = \hat{x}(t)^\top \hat{x}$, $\hat{s}_\zeta(t) = \hat{x}(t)^\top \hat{\zeta}$, $\hat{s}_h(t) = \hat{x}(t)^\top \hat{h}$, $\hat{s}_{g_{model}}(t) = \hat{x}(t)^\top \hat{g}_{model}$, and $b = \hat{s}_x(t) + T(\lambda_2 + \mu)$.

Subproblem $\hat{\zeta}$: Given \hat{h} , \hat{g} and μ , ζ can be updated by:

$$\hat{\zeta} = \hat{\zeta} + \mu \left(\hat{g} - \hat{h} \right), \quad (14)$$

where $\hat{h} = \sqrt{T} (PF^\top \otimes I_K) h$, and $\mu = \min(\mu_{max}, \beta\mu)$. μ_{max} denotes the maximum value of μ . β is the scale factor.

Complexity: The subproblem h has a closed form solution. $g = \frac{1}{\sqrt{T}} (PF^\top \otimes I_K) \hat{g}$ and $\zeta = \frac{1}{\sqrt{T}} (PF^\top \otimes I_K) \hat{\zeta}$ can be

broken into K IFFT computations of $g = \frac{1}{\sqrt{T}} PF^\top \hat{g}$ and $\zeta = \frac{1}{\sqrt{T}} PF^\top \hat{\zeta}$. $Tlog(T)$ is the cost of computing the IFFT of a signal with the length of T . Ultimately, the complexity of solving is $\mathcal{O}(KTlog(T))$, where K represents the number of channels. The subproblem \hat{g} has a closed form solution also. \hat{g} can be expressed as T independent objectives for $\hat{g}(t)$. $\hat{g}(t)$ has a closed form solution in Eq. 13 and its complexity is $\mathcal{O}(K)$. Finally, the complexity of solving subproblem \hat{g} is $\mathcal{O}(KT)$. Hence, the total complexity of our optimization framework is $\mathcal{O}(KTlog(T)I)$, where I represents the maximum number of iterations. A single tracking iteration is summarized in Algorithm 1.

Algorithm 1 Our BACF-TC Approach: Iteration at Time Step t

Input:

Image frame I_t .

Filter model g_{model} .

Previous target position p_{t-1} and scale s_{t-1} .

Output:

Position p_t , scale s_t and updated filter model g_{model} .

Localization and scale estimation:

1: Extract search windows with S scales from I_t at p_{t-1} .

2: Adaptive feature channel selection $[X(s)]_{s=1}^S$.

3: Compute correlation scores y_t .

4: Set p_t to the target position and s_t to the target scale.

Update:

1: Extract feature representations X from I_t at p_t and s_t .

2: Optimise g using Equ. 7 for K iterations.

3: Update filter model g_{model} .

C. ADAPTIVE FEATURE CHANNEL SELECTION

In this section, we design a simple yet effective adaptive feature channel selection method.

General tracking algorithms obtain the final response map via summing the responses for each channel. However, in our experiment, we find that different feature channels have different degrees of discrimination to the target. Some feature channels lead to a reduction in performance, because they are not useful to distinguish between goals and backgrounds.

In this paper, we obtain the response map for each channel by correlation calculation of the initial frame information and filter template. We adopt Peak to Sidelobe Ratio (PSR) [8] to evaluate the degree of feature channel. The PSR is defined as $\frac{g_{max} - \mu_{s1}}{\sigma_{s1}}$, where g_{max} is the peak values. μ_{s1} and σ_{s1} denote the mean and standard deviation of the sidelobe, respectively.

Fig. 3 shows the process of adaptive feature selection. When PSR drops to around the threshold, we discard the feature channel.

IV. EXPERIMENT

In this section, we extensively evaluate our proposed approach on four standard datasets, including OTB-2013 [1], OTB-2015 [2], Temple Color-128 [3], VOT-2018 [4].

TABLE 1. Analysis of our tracker on OTB-2015 dataset.

	BACF (baseline)	BACF + Adaptive feature channel selection	BACF + the temporal consistency constraint	BACF-TC (our tracker)
Success rate (at IOU>0.5)	0.776	0.782	0.797	0.800
Precision (at IOU>0.5)	0.824	0.834	0.868	0.872

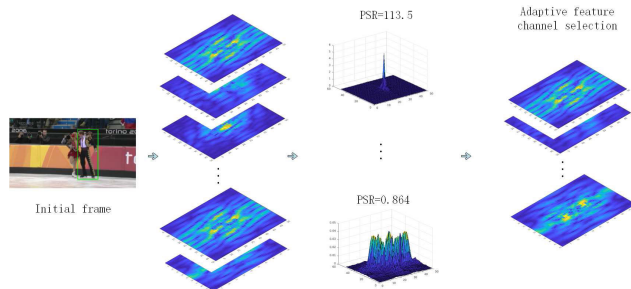


FIGURE 3. The process of adaptive feature channel selection.

Temple Color-128 consists of 128 color video sequences. VOT-2018 is composed of 60 challenging sequences. OTB-2013 and OTB-2015 have 11 attributes occurred in the video sequences, such as background clutter, motion blur, deformation, illumination variation, in-plane rotation, low resolution, occlusion, out-of-plane rotation, out of view, scale variation, fast motion. We compare our proposed approach with 16 state-of-the-art trackers, including trackers using hand-crafted features(i.e. SRDCF [14], ECOhc [15], STAPLE [11], DSST [9], KCF [21], CSR-DCF [22], BACF [20], SAMF [10], STAPLE_CA [23]), LADCF [19]) and trackers using CNN features(i.e. ECO [15], SDRCFdeep [14], STRCFdeep [18], CFCF [42], CFnet [43], SiamFC [44]).

A. EXPERIMENTAL SETUP

Our proposed approach is implemented in MATLAB 2017b on a computer with an Intel i5-8400 CPU at 2.80GHz×6 and Nvidia GTX 1060 GPU.

In this paper, we exploit two evaluation metrics—precision and success rate. Tracking precision is the center location error, which is defined as the average Euclidean distance between the center locations of the tracked targets and the manually labelled ground truths. The success rate refers to the ratios of successful frames at the thresholds of overlap S varied from 0 to 1. The overlap score is defined as $S = \frac{|r_t \cap r_a|}{|r_t \cup r_a|}$, where r_t and r_a represent the tracked bounding box and the ground truth bounding box. \cap and \cup denote the intersection and union of two regions, respectively. $|\cdot|$ denotes the number of pixels in the region. We use the area under the curve (AUC) of each success plot or precision plot and the success rate or precision at the conventional thresholds of 0.5 ($IOU > 0.5$) to rank the tracking algorithms. In this paper, we use the one-pass evaluation (OPE) criterion to evaluate the performance of trackers.

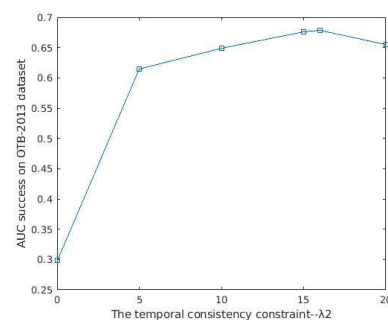
B. RESULTS AND ANALYSIS

1) OTB-2013 AND OTB-2015 DATASETS

a: COMPARISON WITH THE BASELINE BACF

Similar to BACF [7], we adopt the 31 dimensions of HOG feature with a square cell size 4×4 . In the process of scale search, the number of scales and scale step are set to 5 and 1.01, respectively. The regularization factor $\{\lambda_1, \lambda_2\}$ is set to $\{0.01, 16\}$. For the ADMM optimization, the number of iterations and the penalty factor μ are set to 2 and 1. The penalty factor at iteration $i+1$ is updated by $\mu^{i+1} = \min(\mu_{max}, \beta\mu^i)$, where $\mu = 1$, $\beta = 0.1$ and $\mu_{max} = 10^3$.

Table 1 shows an analysis of our contributions. Our primary contribution is introducing the temporal consistency constraint into the BACF. The secondary contribution is proposing an adaptive feature channel selection method, which may further improve overall performance. In Table 1, it has a significant performance improvement by the gains of 2.1% in success rate and 4.4% in precision compared to the baseline, while introducing the global temporal consistency constraint. Based on BACF, adopting an adaptive feature channel selection method can boost 0.6% in success rate and 1.0% in precision. On the whole, our tracker improves the performance of BACF tracker by 2.4% in success rate and 4.8% and in precision. Both of our contributions can improve the performance effectively, especially the former.

FIGURE 4. Impacts of the temporal consistency constraint λ_2 on OTB-2013 dataset.

Furthermore, we analyze the effect of the temporal consistency constraint λ_2 , the threshold of adaptive feature channel selection, and the learning rate ε on OTB-2013 dataset. Fig. 4 shows the AUC success rate in different temporal consistency constraint λ_2 . The AUC success rate is significantly influenced by the choice of λ_2 . The performance of our tracker achieves the best when $\lambda_2 = 16$. As shown in Fig. 5, the threshold of adaptive feature channel selection

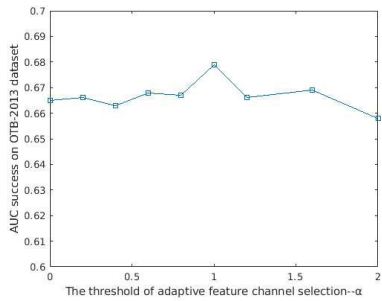


FIGURE 5. Impacts of the threshold of adaptive feature channel selection α on OTB-2013 dataset. fig13.

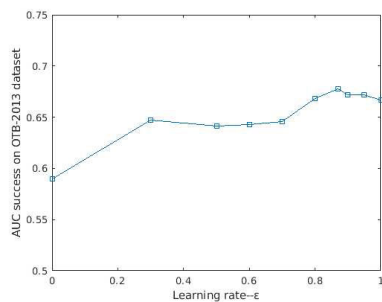


FIGURE 6. Impacts of the learning rate ϵ on OTB-2013 dataset.

α cannot be set too high or too low. Too high threshold will filter out useful channel features, but too low threshold can not reduce the interference channel features. Here, our tracker achieves the best performance when $\alpha = 1$. Fig. 6 shows the effect of the learning rate ϵ on OTB-2013 dataset. When $\epsilon = 1$, the temporal consistency constraint becomes $\|g_{l,t} - g_{l,t-1}\|_2^2$, and the form is the same as the temporal regularization constraint in STRCF [18]. However, our tracker achieves the best performance at $\epsilon = 0.87$. We think the reason is that a more robust model can be obtained by considering both interframe continuity and historical model information.

b: COMPARISON WITH HAND-CRAFTED FEATURES BASED TRACKERS

Here, we make an extensive comparison with the representative state-of-the-art trackers, including SRDCF [14], ECOhc [15], STAPLE [11], DSST [9], KCF [21], CSRDCF [22], BACF [7], SAMF [10], STAPLE_CA [23], STRCF [18]. SRDCF focuses on mitigating boundary effects by introducing spatial regularization. ECOhc adopts the efficient convolution operators and exploits available negative data by including all shifted versions of a training sample. The STAPLE tracker employs color histograms to improve the reliability of the final response. DSST and SAMF are committed to adapt to target scale changes in the tracking. CSR-DCF introduces the channel and spatial reliability concepts to DCF tracking and provides a novel learning algorithm for its efficient and seamless integration in the filter update and the tracking process. STAPLE_CA focus

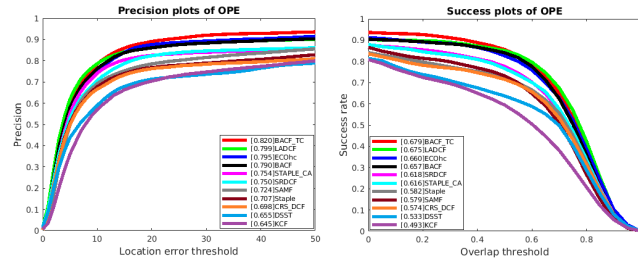


FIGURE 7. Success plots (right) and Precision plots (left) of OPE on OTB-2013 dataset.

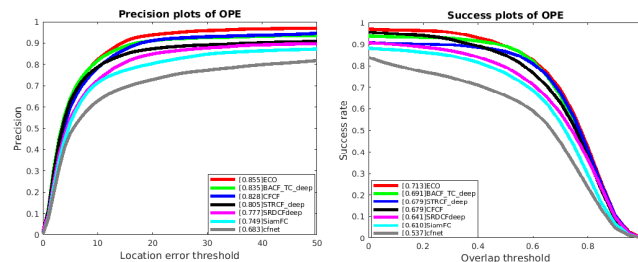


FIGURE 8. Success plots (right) and Precision plots (left) of OPE on OTB-2013 dataset.

on introducing background information in training. KCF employs circulant structure to solve a ridge regression problem in the frequency domain and have a high tracking speed. STRCF focuses on joint spatial-temporal filter learning in a lower dimensional discriminative manifold.

In Fig. 7, our tracker achieves success rate (at AUC) of 67.7% and precision (at AUC) of 82.0%. It is evident that our tracker significantly outperforms others in this comparison.

c: COMPARISON WITH DEEP FEATURES BASED TRACKERS

We analyze the performance of our proposed tracker with deep features. First, similar to ECO [15], our tracker uses VGG-M [40] network trained on the ILSVRC [41] dataset to extract deep features. In Table 2, we compare the performance of our tracker which adopts different convolutional layers. The middle convolutional layer (*conv3*) has the best performance compared with the low convolutional layer (*conv1*, *conv2*) and the high convolutional layer (*conv4*, *conv5*). Ultimately, our BACF-TC_deep tracker adopts the outputs of *conv3* layer from VGG-M network.

TABLE 2. Analysis of our tracker with deep features on OTB-2013 dataset.

	features	Success rate (at AUC)	Precision (at AUC)
BACF-TC_deep	conv1	0.610	0.747
BACF-TC_deep	conv2	0.623	0.769
BACF-TC_deep	conv3	0.691	0.835
BACF-TC_deep	conv4	0.679	0.827
BACF-TC_deep	conv5	0.672	0.807

In the process of scale search, the number of scales and scale step are set to 4.5 and 1.02, respectively. Other parameters remain the same. Here, we make an extensive

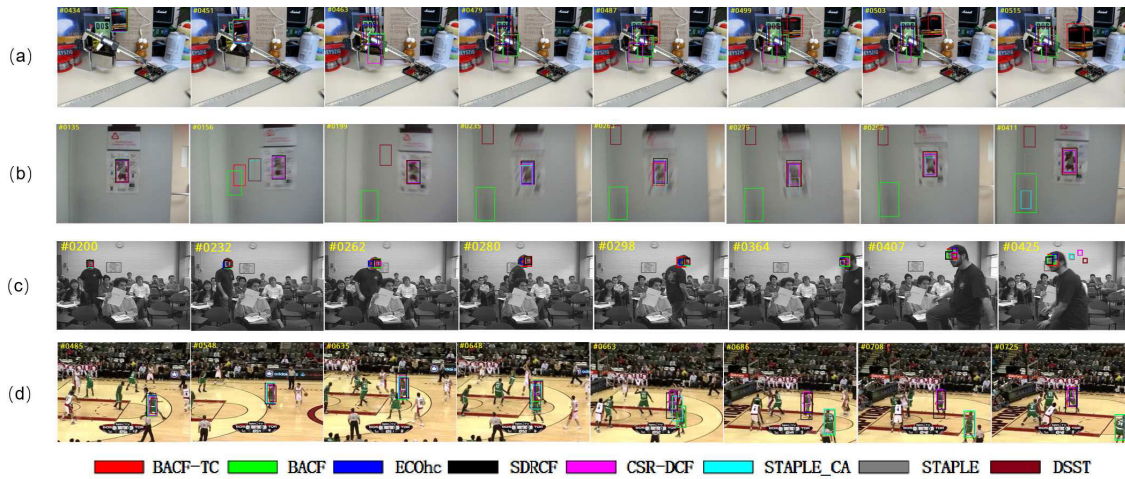


FIGURE 9. Qualitative comparison of our proposed tracker with state-of-the-art trackers on the box, blur owl, freeman and basketball videos, under occlusion, motion blur, out-of-plane rotation and background clutter.

TABLE 3. The success rate (at AUC) of 11 attributes occurred on OTB-2015 dataset.

	BC	MB	DEF	IV	IPR	LR	OCC	OPR	OV	SV	FM
BACF	0.625	0.596	0.582	0.642	0.584	0.512	0.576	0.584	0.552	0.579	0.601
BACF-TC	0.657	0.661	0.607	0.667	0.604	0.550	0.623	0.633	0.586	0.625	0.623

comparison with the representative state-of-the-art trackers, trackers using CNN features (i.e. ECO [15], SDR-CFdeep [14], STRCFdeep [18], CFCF [42], CFnet [43], SiamFC [44]).

Fig. 8 shows our improved tracker still has excellent performance compared with deep features based trackers. Our tracker achieves success rate (at AUC) of 69.1% and precision (at AUC) of 83.5%. Obviously, our tracker significantly outperforms the most tracker in this comparison.

d: ATTRIBUTES-BASED PERFORMANCE

For a comprehensive analysis of our proposed tracker, Fig. 9 illustrates the AUC scores of 11 attributes occurred on OTB-2015 dataset, including background clutter (BC), motion blur (MB), deformation (DEF), illumination variation (IV), in-plane rotation (IPR), low resolution (LR), occlusion (OCC), out-of-plane rotation (OPR), out of view (OV), scale variation (SV), fast motion (FM). In Fig. 11, our tracker BACF-TC achieves success rate (at AUC) of 65.4 % and precision (at AUC) of 80.2% on OTB-2015 dataset. From Table 3, we can see that our proposed tracker consistently outperforming BACF on the success rate intuitively.

In the following, we have a qualitative analysis in several challenging situations.

a) occlusion

In Fig. 9(a), the box video sequence occurs a full occlusion over frames #451 to #479. Only our proposed tracker can successfully track the box, while the other competing trackers occur model drift leading to failure in this experiment. It is evident that our proposed method outperforms its counterparts in this case.

b) motion blur

Due to the rapid movement of a camera or target, the video sequence may occur motion blur which can lead to failure. In Fig. 9(b), BACF, CSR-DCF and STAPLE_CA track the object unsuccessfully in the case of motion blur. However, BACF-TC achieves accurate tracking.

c) out-of-plane rotation

The movement of a camera or target often causes the out-of-plane rotation in the appearance. Fig. 9(c) shows the tracking performance of these trackers in the case of out-of-plane rotation. BACF, SDRCF and BACF-TC can track the walking man precisely, while the other tracker lost their target of tracking.

d) background clutter

Due to contains objects similar to the target in the background region, the tracking results may be affected likely. In Fig. 9(d), the basketball video sequence has background clutter over frames #485 to #725. It is an enormous challenge for tracking that people are wearing the same clothes as the tracking target in the background region. BACF, STAPLE_CA and Staple occur model drift leading to an inaccurate case that the similar objects are treated as tracking target mistakenly. By contrast, BACF-TC accurately estimate the position and scale of the basketballer which is labelled.

2) TEMPLE COLOR-128 DATASET

We perform experiments on Temple Color-128 dataset which consists of 128 color video sequences. The parameter settings are the same as on OTB-2013 and OTB-2015 datasets. In Fig. 12, our method achieves a prominent improvement

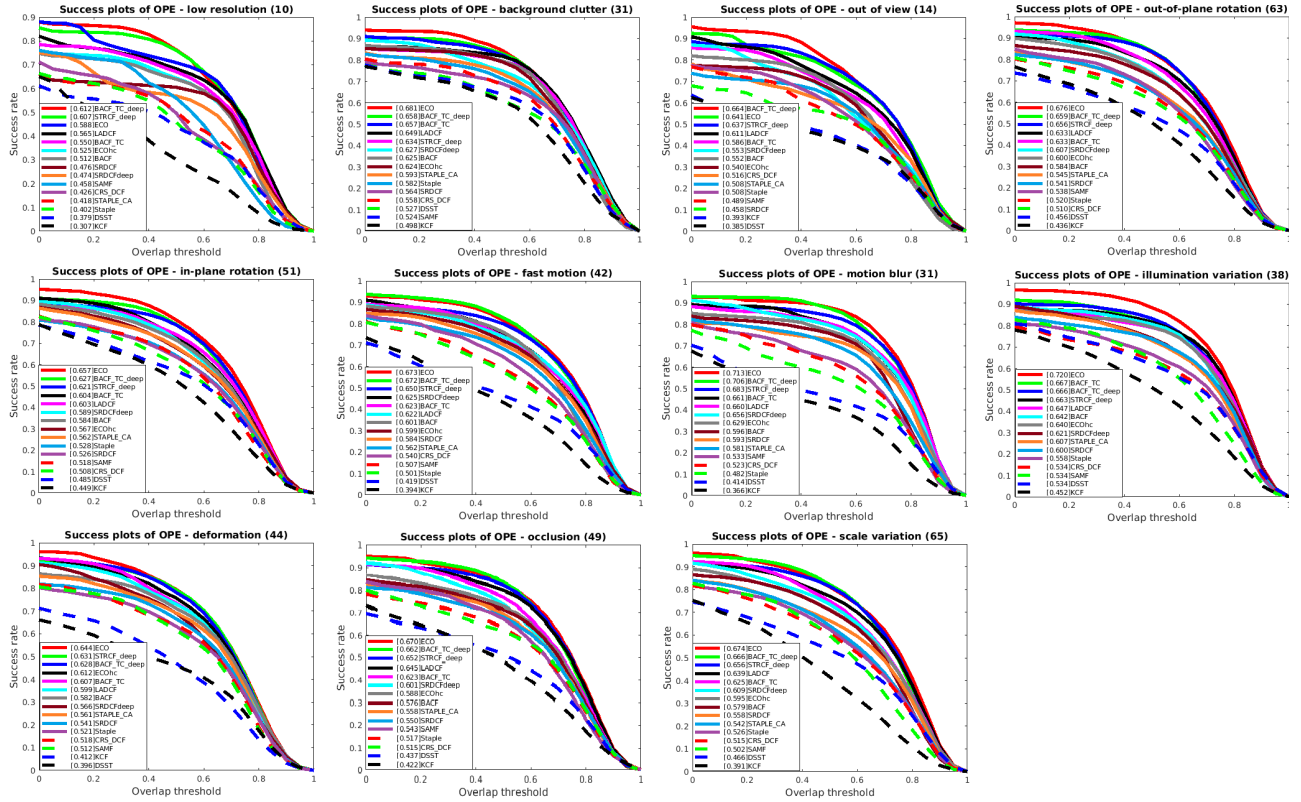


FIGURE 10. The success plots over 11 challenging attributes on OTB-2015 dataset.

TABLE 4. The EAO performance on VOT-2018 dataset.

	BACF-TC_deep (Ours)	BACF-TC (Ours)	BACF	SiamFC	SRDCF/deep	SRDCF	ECO	CSRDCF	CFCF	KCF	DSST	STAPLE
EAO	0.2595	0.1830	0.1436	0.1875	0.1544	0.1179	0.2809	0.2563	0.2806	0.1351	0.0793	0.1688

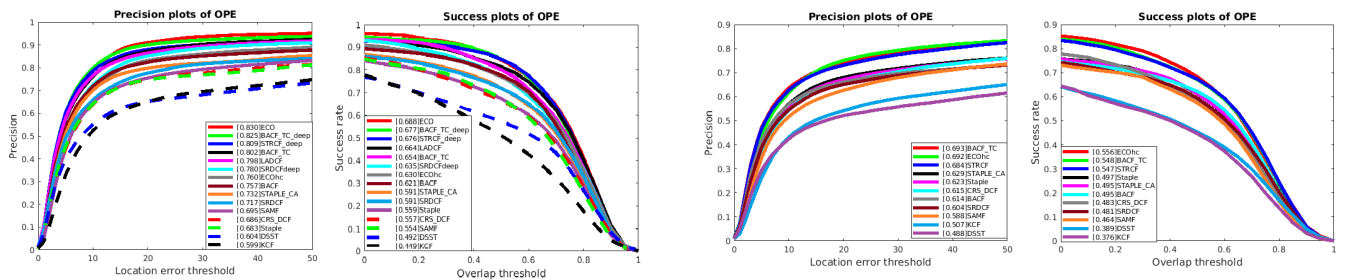


FIGURE 11. Success plots (right) and Precision plots (left) of OPE on OTB-2015 dataset.

over BACF with the gains of 5.3% in success rate (at AUC) and 7.9% in precision. The precision of our method is the highest of all of the others.

3) VOT-2018 DATASET

VOT-2018 dataset contains 60 challenging sequences and has 5 visual attributes, such as occlusion, illumination change, motion change, size change, and camera motion. In Table 4, we use the expected average overlap (EAO) to analyze the

FIGURE 12. success plots(right) and precision plots(left) of OPE on the Temple Color-128 dataset.

tracking performance. We can see from Table 4 that our proposed tracker performs significantly better than the baseline BACF in terms of the EAO metric. Our proposed tracker (BACF-TC) improves the performance of BACF tracker by 3.94% in EAO. Furthermore, our tracker with deep features (BACF-TC_deep) achieves a conspicuous improvement over BACF with a gain of 11.59% in EAO.

4) EVALUATIONS ON TRACKING SPEED

For practical tracking applications, tracking speed is a critical factor to determine whether the tracker can be widely used.

TABLE 5. The speeds of different trackers based hand-crafted features on OTB-2015 dataset.

	BACF-TC (Ours)	BACF	LADCF	ECOhc	SRDCF	STAPLE_CA	STAPLE	CRS_DCF	KCF	SAMF	DSST
Avg.FPS (CPU)	23.7	41.6	8.2	37.6	12.8	52.5	67.7	17.8	315.1	29.4	63.9

TABLE 6. The speeds of different trackers based deep features on OTB-2015 dataset.

	BACF-TC_deep (Ours)	SDRCF_deep	ECO	STRCF_deep	SiamFC	cfnet	CFCF
Avg.FPS (GPU)	18.4	0.2	16.1	6.6	28.8	42.0	1.3

For a full comparison, Table 5 and Table 6 list the average speed of our proposed tracker and these state-of-the-art trackers on OTB-2015 dataset. The average tracking speed of BACF-TC reaches 23.7 FPS on CPU, and the average tracking speed of BACF-TC_deep reaches 18.4 FPS on GPU. Considering combining the tracking speed and performance in practical application, our proposed tracker makes a good tradeoff between tracking speed and performance.

V. CONCLUSION

In this paper, we propose the improved BACF via the temporal consistency constraint (BACF-TC) and offer a fast algorithm via ADMM in the frequency domain. In addition, we propose a simple yet effective adaptive feature channel selection method. Our approach can mitigate model drift efficiently and achieve overall performance improvement. Extensive experiments demonstrate that the proposed tracker outperforms the baseline BACF in terms of success rate and precision. Moreover, the proposed tracker performs favourably against many representative state-of-the-art methods regarding precision and success rate.

REFERENCES

- [1] Y. Wu, J. Lim, and M. H. Yang, "Online object tracking: A benchmark," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2411–2418.
- [2] Y. Wu, J. Lim, and M. H. Yang, "Object tracking benchmark," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1834–1848, Sep. 2015.
- [3] P. Liang, E. Blasch, and H. Ling, "Encoding color information for visual tracking: Algorithms and benchmark," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5630–5644, Dec. 2015.
- [4] M. Kristan et al., "The sixth visual object tracking VOT2018 challenge results," in *Proc. Eur. Conf. Comput. Vis.*, 2018.
- [5] Z. Chen, Z. Hong, and D. Tao, "An experimental survey on correlation filter-based tracking," 2015, *arXiv:1509.05520*. [Online]. Available: <https://arxiv.org/abs/1509.05520>
- [6] H. Lu, P. Li, and D. Wang, "Visual object tracking: A survey," *Pattern Recognit. Artif. Intell.*, vol. 31, no. 1, pp. 61–76, 2018.
- [7] S. Ojha and S. Sakhare, "Image processing techniques for object tracking in video surveillance—A survey," in *Proc. Int. Conf. Pervas. Comput.*, 2015, pp. 1–6.
- [8] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2544–2550.
- [9] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg, "Discriminative scale space tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 8, pp. 1561–1575, Aug. 2017.
- [10] Y. Li and J. Zhu, "A scale adaptive kernel correlation filter tracker with feature integration," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 254–265.
- [11] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, and P. H. Torr, "Staple: Complementary learners for real-time tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1401–1409.
- [12] M. Danelljan, G. Hager, F. Shahbaz Khan, and M. Felsberg, "Convolutional features for correlation filter based visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Dec. 2015, pp. 58–66.
- [13] K. Nai, D. Xiao, Z. Li, S. Jiang, and Y. Gu, "Multi-pattern correlation tracking," *Knowl.-Based Syst.*, vol. 181, no. 1, 2019, doi: 10.1016/j.knosys.2019.05.032.
- [14] M. Danelljan, G. Hager, F. Shahbaz Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 4310–4318.
- [15] M. Danelljan, G. Bhat, F. S. Khan, and M. Felsberg, "Eco: Efficient convolution operators for tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 6638–6646.
- [16] M. Danelljan, A. Robinson, F. S. Khan, and M. Felsberg, "Beyond correlation filters: Learning continuous convolution operators for visual tracking," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 472–488.
- [17] K. Dai, D. Wang, H. Lu, C. Sun, and J. Li, "Visual tracking via adaptive spatially-regularized correlation filters," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 4670–4679.
- [18] F. Li, C. Tian, W. Zuo, L. Zhang, and M. H. Yang, "Learning spatial-temporal regularized correlation filters for visual tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4904–4913.
- [19] T. Xu, Z.-H. Feng, X.-J. Wu, and J. Kittler, "Learning adaptive discriminative correlation filters via temporal consistency preserving spatial feature selection for robust visual object tracking," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5596–5609, Nov. 2019.
- [20] H. Kiani Galoogahi, A. Fagg, and S. Lucey, "Learning background-aware correlation filters for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 1135–1143.
- [21] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 583–596, Mar. 2015.
- [22] A. Lukezic, T. Vojir, L. Cehovin Zajc, J. Matas, and M. Kristan, "Discriminative correlation filter with channel and spatial reliability," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 6309–6318.
- [23] M. Mueller, N. Smith, and B. Ghanem, "Context-aware correlation filter tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1396–1404.
- [24] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. Int. Joint Conf. Artif. Intell.*, 1981, p. 674.
- [25] S. Baker and I. Matthews, "Lucas-Kanade 20 years on: A unifying framework," *Int. J. Comput. Vis.*, vol. 56, no. 3, pp. 221–255, 2004.
- [26] S. Oron, A. Bar-Hillel, and S. Avidan, "Extended Lucas-Kanade tracking," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 142–156.

- [27] S. Oron, A. Bar-Hillel, D. Levi, and S. Avidan, "Locally orderless tracking," *Int. J. Comput. Vis.*, vol. 111, no. 2, pp. 213–228, Jan. 2015.
- [28] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," *Int. J. Comput. Vis.*, vol. 77, nos. 1–3, pp. 125–141, 2008.
- [29] X. Jia, H. Lu, and M. H. Yang, "Visual tracking via adaptive structural local sparse appearance model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1822–1829.
- [30] K. Nai, Z. Li, G. Li, and S. Wang, "Robust object tracking via local sparse appearance model," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 4958–4970, Jun. 2018.
- [31] W. Zhong, H. Lu, and M.-H. Yang, "Robust object tracking via sparse collaborative appearance model," *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 2356–2368, May 2014.
- [32] X. Mei and H. Ling, "Robust visual tracking using ℓ_1 minimization," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 1436–1443.
- [33] S. Avidan, "Subset selection for efficient SVM tracking," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2003, p. 1.
- [34] H. Grabner, M. Grabner, and H. Bischof, "Real-time tracking via on-line boosting," in *Proc. BMVC*, 2006, vol. 1, no. 5, p. 6.
- [35] H. Grabner, C. Leistner, and H. Bischof, "Semi-supervised on-line boosting for robust tracking," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2008, pp. 234–247.
- [36] T. Zhang, S. Liu, N. Ahuja, M.-H. Yang, and B. Ghanem, "Robust visual tracking via consistent low-rank sparse learning," *Int. J. Comput. Vis.*, vol. 111, no. 2, pp. 171–190, 2015.
- [37] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*. Philadelphia, PA, USA: SIAM, 1989, p. 9.
- [38] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
- [39] M. A. Akgun, J. H. Garcelon, and R. T. Haftka, "Fast exact linear and non-linear structural reanalysis and the Sherman Morrison Woodbury formulas," *Int. J. Numer. Methods Eng.*, vol. 5, no. 7, pp. 1587–1606, 2001.
- [40] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," in *Proc. Brit. Mach. Vis. Conf.*, 2014, pp. 2–22.
- [41] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, and A. C. Berg, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.
- [42] E. Gundogdu and A. A. Alatan, "Good features to correlate for visual tracking," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2526–2540, May 2018.
- [43] J. Valmadre, L. Bertinetto, J. Henriques, A. Vedaldi, and P. H. Torr, "End-to-end representation learning for correlation filter based tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2805–2813.
- [44] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. Torr, "Fully-convolutional siamese networks for object tracking," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 850–865.



machine learning, deep learning, and computer vision.

SHANSHAN LUO received the B.S. degree in communication engineering from Wuhan University, Wuhan, China, in 2015. She is currently pursuing the Ph.D. degree with the Science and Technology on Microsystem Laboratory, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, and with the University of Chinese Academy of Sciences. She has worked on visual tracking. Her interests include digital signal processing,



BAOQING LI received the Ph.D. degree from the State Key Laboratory of Transducer Technology, Shanghai Institute of Metallurgy, Chinese Academy of Sciences, Shanghai, China, in 2000. He is currently a Professor and the Ph.D. Supervisor with the Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences. His research interest is the application of wireless sensor networks.



XIAOBING YUAN received the Ph.D. degree from the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Shanghai, China, in 2000. He is currently a Professor and the Ph.D. Supervisor with the Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences. His research interests include wireless sensor networks, information transmission, and processing.

...