Object Tracking Based on a Time-Varying Spatio-Temporal Regularized Correlation Filter With Aberrance Repression

Junnan Wang, Zhenhong Jia, Huicheng Lai, Jie Yang, *Senior Member, IEEE*, and Nikola K. Kasabov, *Life Fellow, IEEE*

***Abstract*—When used for object tracking, the discriminative correlation filter (DCF) is effective, but its performance is often burdened by undesirable boundary effects. Meanwhile, when there is too much background information in training samples of the DCF, it will be easier to learn the area deviating from the tracking object. Further, illumination variation, partial/full occlusion, and appearance variations, render the response map aberrance of the correlation filter (CF) more prone to occur. To overcome these problems, an object tracking model based on a time-varying Spatio-temporal regularized correlation filter with aberrance repression is proposed in this paper. Firstly, by adding a regularized term to the traditional CFs to limit the change rate of the response map generated in the object detection phase, the proposed tracker can obviously repress the aberrance of the response maps; secondly, by adjusting the filter to the object regions suitable for tracking with high confidence scores with a time-varying spatial reliability map, the proposed tracker effectively overcomes the adverse effects caused by the boundary effect; and finally, by introducing a temporal regularized term, the proposed tracker also has superior tracking ability for the partial occluded objects and those with large appearance variations. Significant experiments on the OTB100, VOT2016, TC128, and UAV 123 datasets have revealed that the performance thereof outperformed many state-of-the-art trackers based on DCF and deep-based frameworks in terms of tracking accuracy, tracking success rate, and A-R rank, etc.**

***Index Terms*—Visual tracking, correlation filter, spatial reliability map, temporal regularized, aberrance repression.**

I. Introduction

AMONGST the background of improvements in computer hardware performance and the rapid development of artificial intelligence technology, visual tracking has made more development, and has progressively received an increasing amount of attention. Presently, although several visual tracking technologies have achieved good results when applied in video surveillance, action recognition, and automatic driving, etc., how to track an object quickly, accurately, and robustly in real situations is still a very challenging task.

Recently, the use of deep convolutional neural networks (CNNs) in visual tracking has emerged as a favorable option [1], [2], [3]. Trackers based on deep features such as CNN [4], ResNet [5], and VGG-Net [6] can capture complex and hierarchical object features, and the tracking accuracy thereof is generally high. Yet, these trackers require a substantial amount of convolution operations, which contribute to the poor real-time performance thereof. Further, the demand for substantial number of training data limits the application thereof in visual tracking.

In 2010, a filtering algorithm for the minimum output sum of squared error filter (MOSSE) was proposed [9], which was the first time the CF was introduced into visual tracking. Compared with traditional tracking methods, MOSSE exhibited strong competitiveness. Subsequently, Henriques et al. [7] solved the problem of the number of samples being limited, and reduced the computational complexity by applying the circular matrix and the kernel technique, thereby rendering the tracker more versatile.

Among the existing tracking methods, the discriminative correlation filter paradigm has been proved to have excellent performance in visual tracking [7], [8]. Although the correlation filter possesses a myriad of advantages, owing to the use of circular training samples in the tracker, the periodic Fourier transform of the left and right upper and lower image boundaries will produce noise, resulting in an undesirable boundary effect, thereby limiting the object tracking performance of the correlation filter in a number of aspects. Firstly, the lack of negative training samples will lead to over fitting of the learning model, significantly affecting the tracking ability with regard to the deformed objects of the tracker. Secondly, the lack of real negative training samples will substantially reduce the tracking robustness of the tracker with regard to the objects with background clutter, and the risk of tracking drift will increase especially when the object and background display similar visual cues. Thirdly, the simple expansion of the image region employed to train filters results

Manuscript received 23 July 2022; revised 14 November 2022; accepted 1 December 2022. Date of publication 6 December 2022; date of current version 16 December 2022. This work was supported in part by the National Natural Science Foundation of China under Grant U1803261, and in part by the International Science and Technology Cooperation Project of the Ministry of Education of the Peoples Republic of China under Grant DICE 2016-2196. *(Corresponding author: Zhenhong Jia.)*

Junnan Wang, Zhenhong Jia, and Huicheng Lai are with the School of Information Science and Engineering, and the Key Laboratory of Signal Detection and Processing, Xinjiang University, Urumqi 830046, China (e-mail: 1254982138@qq.com; jzhh@xju.edu.cn; lai590@qq.com).

Jie Yang is with the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200400, China (e-mail: jieyang@sjtu.edu.cn).

Nikola K. Kasabov is with the School of Engineering, Computing and Mathematical Sciences, Auckland University of Technology, Auckland 1010, New Zealand (e-mail: nkasabov@aut.ac.nz).

Digital Object Identifier 10.1109/JPHOT.2022.3227118

in a profusion of background information being contained in the positive training samples, and these corrupted training samples considerably reduce the object recognition ability of the model. In addition, if the objects move too fast, the DCF based trackers will encounter issues because of the limited search areas. Finally, the DCF based trackers also generally have other problems, such as insufficient expression ability of the selected features and insufficient consideration of the influence of the time continuity of tracking on tracking results, etc.

Regarding the aforementioned problems in DCF trackers [10], [11], [12], a plethora of methods have been proposed to solve them. To solve the boundary effects of the tracker when using circular shift samples for training, a spatial regularized DCF model (SRDCF) was proposed [13], which imposed a spatial penalty on the DCF coefficients. Meanwhile, after a temporal regularized term being introduced into SRDCF, a Spatio-temporal regularized correlation filter was proposed [14], which compared with SRDCF, could handle the boundary effects without losing efficiency. However, they are all based on a fixed spatial regularization mode and only consider the information compression along the spatial dimensions, which make them unable to make full use of the diverse information of the object. Yuan et al. [47] incorporated channel and spatial confidence into DCF and proposed a DCF with channel and spatial reliabilities, which could not only allow for the searching of the objects in any range area, but could also effectively track the objects with irregular shape. However, as the influence of time continuity on tracking results is not fully considered, its tracking ability to the occluded objects and the objects with large appearance variations is weak. In addition, Xu et al. [16] enhanced the object recognition and interpretation ability of the filter by means of reducing the information redundancy and un-correlation of high-dimensional multichannel features through group feature selection in two dimensions of space and channel. But it requires a large amount of calculation since the complex features are used in filter training.

Ning et al. [17] employed discriminant correlation filters (DCFs) combined with different feature types to construct multiple experts, and each expert independently tracked the object. Here, the optimal expert was selected to optimize the tracking results and superior tracking results were obtained. However, due to the complex structure of the tracker, the real-time object tracking performance cannot be achieved. In the adaptive spatial regularized correlation filter proposed by Dai et al. [18], two correlation filter models were used, in which one used complex features for object localization, and the other used shallow features for scale estimation, with the tracking performance thereof being considerably high. Nevertheless, as the set of shallow and deep features are used to train the correlation filter to determine the object position, it requires a large amount of calculation.

From conducting a detailed comparison and analysis on the advantages and disadvantages of the current DCF based tracking models, in this paper, a novel visual object tracking model is proposed, i.e., object tracking based on a time-varying Spatio-temporal regularized correlation filter with aberrance repression (ARSTCF). The proposed tracking model can better solve the boundary effect of the DCF-based tracking models in visual tracking and improve the tracking ability to the occluded objects and the objects with large appearance variations. Extensive experimental results conducted on four general video sequence data sets (OTB100 [22], VOT2016 [23], TC128 [24], and UAV123 [25]) indicate that the proposed tracking model has certain advantages in some evaluation indicators compared with several state-of-the-art tracking models.

The major contributions of this paper are summarized below:

1) A regularized term is added into the standard DCF tracking model, which can limit the change rate of the response map in the object detection stage, thereby allowing the proposed tracking model to obviously repress the aberrance of the object response map, and improve the tracking robustness and accuracy, instead of using complex combinations of the object features.
2) Differing from the current DCF based tracking models, which generally use a fixed spatial regularized matrix in the objective function to repress the response of the correlation filter outside the object region by adjusting the filter coefficients. A time-varying spatial reliability map is employed to adjust the CF to the regions with high confidence scores so that the filter can reflect the characteristics and appearance variations of the object in real-time, thereby effectively overcoming the adverse effects caused by the boundary effect.
3) A correlation filter temporal regularized term is introduced into the standard DCF tracking model, so as to allow the proposed tracking model to effectively deal with the boundary effect. Meanwhile, in the case of occlusion and large appearance variations, the proposed tracking model can alleviate tracking drift by passively updating the DCFs to maintain closeness to the previous ones, and the tracking ability for the occluded objects and the objects with large appearance variations have also been further improved.
4) Extensive experiments in the image sequences of the OTB100, VOT2016, TC128, and UAV 123 data sets show that the proposed low-illumination object tracking model outperforms state-of-the-art trackers.

An overview of our framework for Object tracking is shown in Fig. 1.

The remainder is structured below: the related works of visual tracking in recent years are reviewed in Section II; the proposed tracking model is elaborated in Section III; in Section IV, the effectiveness and robustness of the proposed tracking model have been proved by significant experiments; the experimental analysis is given in Section V; the conclusion is given in Section VI.

II. Related Work

In recent years, object tracking models that are based on deep learning have been increasingly applied. Wei et al. [26] proposed an object-aware network that includes a background filtering module, channel complementary module, and template adaptive network to improve feature representation and to realize

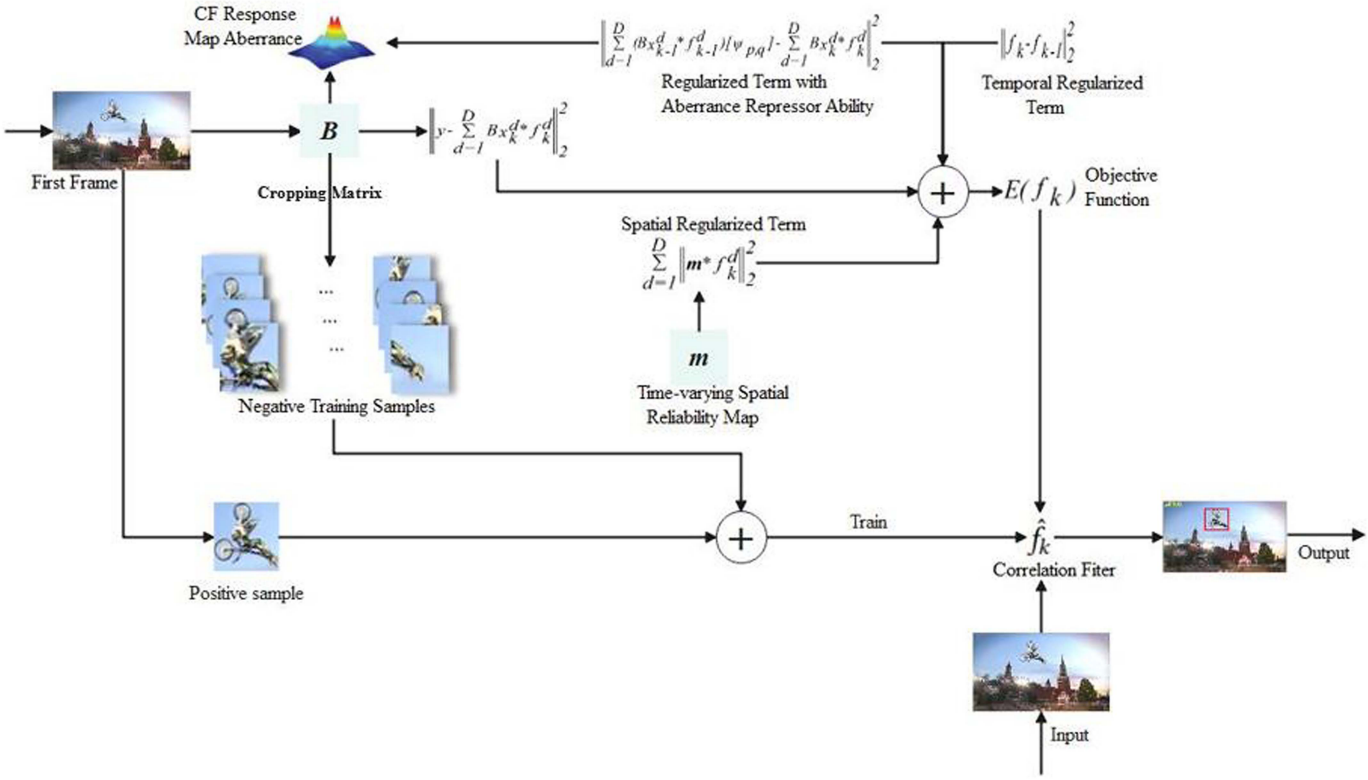

Fig. 1. An overview of our framework.

robust object tracking. Li et al. [27] successfully trained an object tracker based on SiamRPN using ResNet as the backbone network. Zhang et al. [28] built a tracker that contains both semantic models and appearance models by using Siamese networks and multi-granularity color features. Chandrakar et al. [29] proposed an enhancement system for automatic moving object detection and object tracking using a radial basis function-based, filtered, deep learning neural network and a counterfactual, regret minimization algorithm. Yu et al. [30] combined the discriminant correlation filter with the matching-based segmentation paradigm as a local tracker and used a memory attention network based on partial cost to address the problem of object appearance change. Zhu et al. [31] proposed a multilevel predictive Siamese network composed of a Siamese feature extraction module and multilevel predictive module for object tracking in unmanned aerial vehicle (UAV) video. For small-size object tracking, a residual feature fusion block is designed, and the low-level feature representation is constrained by high-level abstract semantics. Fu et al. [32] proposed a tracking algorithm for learning discrimination and adaptive feature representation by using a hard-balanced, focus loss function and embedding an off-line training, guidance domain adaptive module in a Siamese network. Fu et al. [33] proposed a visual target tracking method based on a Siamese modulation network on the basis of Resnet to extract the multi-layer fusion features of a given object in the first frame and the current frame. Meng et al. [34] proposed a hierarchical correlation Siamese network for object tracking. The network uses the convolution features of each layer to compare the correlation between the object and the search area, and determines the location of the tracking object according to the maximum correlation.

Xie et al. [36] found that the features extracted by Siamese-like networks cannot completely distinguish between the tracked target and the distractor objects. Therefore, by deeply embedding cross-image feature correlation in multiple layers of the feature network, a new target-dependent feature network is proposed, and it allows the features of the search and template images to be deeply fused for tracking. Considering that the temporal contexts among consecutive frames are far from being fully utilized in the existing object trackers, Cao et al. [37] proposed an object tracking framework for exploiting temporal contexts in Siamese-based networks. In terms of feature extraction, an online temporally adaptive convolution is proposed to extract features with convolution weights dynamically calibrated by the previous frames; At the level of similarity map, an adaptive temporal transformer is proposed to refine the similarity map according to the temporal information. To improve the expressibility of the tracking network, Mayer et al. [38] proposed a tracking framework using a Transformer-based model prediction module. The weights of the framework are obtained using a Transformer-based model predictor, which allows it to learn more powerful object models. Ye et al. [39] developed an unsupervised domain adaptive framework for night-time aerial tracking. In the the framework, to solve the problem of domain discrepancy, a Transformer-based bridging layer is used on the feature extractor to align the image features from the day-time and night-time. Through the Transformer

day/night feature discriminator, the day-time tracking model is adversarially trained to track at night. Considering that previous unsupervised tracking methods are unable to track the objects with strong variation over a long time span, Shen et al. [40] proposed a novel unsupervised tracking framework, which is composed of three components: consistency propagation transformation, region mask operation, and mask-guided loss reweighting, and aims to learn the temporal correspondence both on the classification and regression branchs. As the current single-object tracking method and multi-object tracking method are usually not easily adapted to the other, Ma et al. [41] developed a Unifed Transformer Tracker to solve these two tracking tasks, in which the correlation between the target feature and the tracking frame feature is used to locate the target.

In addition, DCF has also been widely used in visual tracking in these years. With a view to realize an adaptive spatially-regularized correlation filter, Xu et al. [42] embedded dynamic spatial feature selection into the filter learning phase. An effective updating strategy was proposed by Lukežic et al. [43], which could minimize the short term visual model pollution and activate re-detection in time, so as to solve the model degradation problem caused by occlusion. Xie et al. [44] established the motion appearance model through precise Spatio-temporal segmentation in combination with motion information between consecutive frames for online object tracking, while Bertinetto [45] used the complementarity of multiple features to train a DCF tracker, which had strong robustness in the case of severe object deformation and occlusion.

With the introduction of the channel and spatial reliability into DCF, a discriminant correlation filter (CSR-DCF) was proposed [15], which allowed the object to be searched in any area, in addition to effectively tracking the irregular shaped object. For overcoming certain negative effects produced by generated samples, an object-focusing convolutional regression model was proposed [47]. Meanwhile, a temporal constrained background aware CF with saliency map was proposed by Liao et al. [48], which enhanced the robustness of the tracker.

Sun et al. [49] proposed an adaptive kernel CF tracking model to improve the tracking accuracy in complex scenes. To solve the tracking drift problem, Guan et al. [50] proposed a reliability redetermination correlation filter to adjust the weight of each feature, in which two different weight solvers were formulated to determine the weights, thereby representing the importance of each feature more accurately.

Taking into account that the existing approaches treat deep features produced by different network layers independently, which limits the representation power thereof, Zheng et al. [51] proposed a multi-task deep dual CFs based tracking method for robust visual tracking. To enhance tracking ability to the occluded and deformed objects, Pu et al. [52] constructed a spatial map with deep features and introduced temporal regularization into DCF training. A lightweight particle filter was proposed by Li et al. [53], which not only retained the robust tracking ability of the particle filter, but also reduced the time cost in sampling with the use of correlation filters.

Compared with the existing DCF based tracking models, the proposed object tracking model has the following differences:

1) To make the DCF in this paper obtain stronger discrimination to background clutter, we introduce a background clipping matrix to the objective function of the standard correlation filter to clip the real background around the tracking object as negative training samples to train the CFs, instead of using the negative training samples with false background information generated by the circular shift of a base sample. For the aberrance of the CF response map that may be caused by the use of the clipping matrix, we introduce a regularized term with aberrance repressor ability to the objective function to deal with, instead of leaving it alone. As far as we know, it is the first time that the the DCF training method and the response map aberrance repression method are combined to use.
2) Different from the current tracking models based on DCF and using spatial regularization, which use a fixed spatial regularization matrix in their objective functions, we use a time-varying spatial reliability map to weight the training samples. By dynamically weighting the training samples, the DCF can be adjusted to the regions suitable for tracking with high confidence scores for learning, which effectively suppressed the response of the correlation filter in the non-object area. To our knowledge, our proposed tracking model is the first to use the spatial reliability map in this way.

III. Proposed Tracking Model

A. Background Aware Correlation Filter

The tracking performance of a traditional CF based tracking model is often affected by an undesirable boundary effect. For this reason, a background aware CF (BACF) for real-time object tracking model was proposed by Galoogahi et al. [10]. By introducing a clipping matrix in the CF training process, BACF could use the object as a positive sample, in addition to also being able to densely sample the real background as negative samples to train the filter.

The objective function of BACF in the spatial domain is

$$E(f) = \frac{1}{2}\left\| y - \sum_{d=1}^{D} Bx^d * f^d \right\|_2^2 + \frac{\lambda}{2}\sum_{d=1}^{D}\left\| f^d \right\|_2^2 \qquad (1)$$

where $x = [x^1, x^2, ..., x^D]$ is the vectorized sample with D feature maps, $x^d \in R^N$ is the d-th channel sample, D is the number of feature channels, y is the vectorized expected output, $f^d \in R^N$ is the correlation filter to be learned in the d-th channel, $*$ is the convolution operator, B is a cropping matrix, which is used to choose central M elements of each channel sample, λ is a regularized factor.

(1) can be expressed in the frequency domain as follows:

$$\hat{E}(f) = \frac{1}{2}\left\| \hat{y} - B\hat{f}^H \operatorname{diag}(\hat{x}) \right\|_2^2 + \frac{\lambda}{2}\|\hat{f}\|_2^2 \qquad (2)$$

where the superscript (^) represents the discrete fourier transform (DFT) of a vector, that is, $\widehat{a} = \sqrt{N}F(a)$ is the discrete Fourier transform of the vector a, $(\cdot)^H$ is a Hermitian transpose.

A drawback exists in that more background clutter will be introduced when using too much background information to train CFs in BACF, thereby rendering the filter learning more background noise instead of the object. Further, a similar object in the context is more likely to be suggested as the real object than some previous CF based trackers. Along with the object appearance variations caused by full/partial OCC, IV, etc., the response map in the object detection process is more likely to distort, which could substantially degrade its tracking credibility.

B. The Proposed Tracking Model

In this paper, a novel object tracking model, i.e., object tracking model based on a time-varying Spatio-temporal regularized correlation filter with aberrance repression, is proposed for better solving the problems of the boundary effect, the aberrance of the response map and the rectangular object hypothesis in the traditional correlation filter tracking model.

Firstly, to eliminate the boundary effects, the spatial reliability map varying with time was introduced into the objective function of the standard CF tracking model. Here, the spatial reliability map could shift the filter to the regions with high confidence score, thereby alleviating the circular shift problem of the correlation filter in the arbitrary search range and the limitation of the tracking object rectangular hypothesis. By adopting the spatial reliability map, the proposed tracking model could solve the adverse effects caused by the boundary effect well.

Secondly, for the purpose of repressing the aberrances of the response map, the proposed tracking model integrated the repression of the occurrences thereof to the training process of correlation filters. By introducing a regularized term that could limit the change rate of the response map generated in the object detection stage into the objective function of standard CFs, the aberrances of the response map were obviously repressed, and the robustness and accuracy of the object tracking were improved.

Finally, by introducing a temporal regularized term into the objective function of the standard DCF tracking model, the synchronization of the DCF learning and updating was realized in the proposed tracking model. In this paper, the tracking drift was avoided by passively updating the DCF to keep it close to the previous, at the same time, its tracking ability to the occluded objects and the objects with considerable appearance variations was also enhanced at some extent.

The objective function of the proposed tracking model in the spatial domain is:

$$E\left(f_k\right) = \frac{1}{2}\left\| y - \sum_{d=1}^{D} Bx_k^d * f_k^d \right\|_2^2 + \frac{\lambda}{2}\sum_{d=1}^{D}\left\|\boldsymbol{m} * f_k^d\right\|_2^2 + \frac{\gamma}{2} \left\| \sum_{d=1}^{D}\left(Bx_{k-1}^d * f_{k-1}^d\right)[\psi_{p,q}] - \sum_{d=1}^{D} Bx_k^d * f_k^d \right\|_2^2 + \frac{\varepsilon}{2}\left\|f_k - f_{k-1}\right\|_2^2 \tag{3}$$

where $\boldsymbol{m}$ in the second term of (3) is the spatial reliability map; the third term is the regularized term to repress the aberrances of the response map; the fourth term is the introduced temporal regularized term; λ and ε are regularized factors; γ is a aberrance penalty factor; $f^d$is the CF learned in d-th channel; subscripts k and $(k-1)$ represent the k-th and $(k-1)$-th frame. The regularized term $\| \sum_{d=1}^{D}(Bx_{k-1}^d * f_{k-1}^d)[\psi_{p,q}] - \sum_{d=1}^{D} Bx_k^d * f_k^d\|_2^2$ is introduced to represent the difference between the response maps of the two frames before and after, and p and q represent the position difference of two peaks in both response maps in two-dimensional space, $[\psi_{p,q}]$ represents the shifting operation to make the two peaks coincide with each other. When an aberrance occurs, it is proved that the similarity between the two frames before and after would suddenly drop, resulting in the value of the regularized term will be high, the aberrance can be suppressed by optimizing and reducing it. The $\|f_k - f_{k-1}\|_2^2$ is the introduced temporal regularized term.

In order to solve the objective function in the frequency domain more conveniently, (3) is expressed in matrix form as follows:

$$E\left(f_k\right) = \frac{1}{2}\left\| y - X_k\left(I_D \otimes B^T\right) f_k\right\|_2^2 + \frac{\lambda}{2}\left\|\boldsymbol{m} * f_k\right\|_2^2 + \frac{\gamma}{2} \left\| M_{k-1}\left[\psi_{p,q}\right] - X_k\left(I_D \otimes B^T\right) f_k\right\|_2^2 + \frac{\varepsilon}{2}\left\|f_k - f_{k-1}\right\|_2^2 \tag{4}$$

where X_K is the matrix form of input sample x_k; I_D is an identity matrix with a size D*D; the operator $\otimes$ represents Kronecker product, and the superscript $(*)^T$ represents conjugate transpose operation; M_{k-1} is the previous response map, whose value is equal to $[X_{K-1}(I_D \otimes B^{\top})F_{K-1}]$; f_k and f_{k-1} are the CFs learned in k-th and $(k-1)$-th frames respectively.

To get high order of computational efficiency, (4) is transformed to the frequency domain by discrete Fourier transform as follows:

$$\hat{E}\left(f_k, \hat{g}_k\right) = \frac{1}{2}\left\|\hat{y} - \hat{X}_k\hat{g}_k\right\|_2^2 + \frac{\lambda}{2}\left\|\boldsymbol{m} * f_k\right\|_2^2 + \frac{\gamma}{2}\left\|\hat{M}_{k-1} - \hat{X}_k\hat{g}_k\right\|_2^2 + \frac{\varepsilon}{2}\left\|f_k - f_{k-1}\right\|_2^2 \tag{5}$$

$$\hat{g}_k = \sqrt{N}\left(I_D \otimes B^T\right) f_k \tag{6}$$

where the auxiliary variables of (6) are introduced to further optimize (5).

To achieve the global optimal solution of (5), and considering the convexity of it, the alternative direction method of multipliers (ADMM) is used to accelerate computing. Therefore, (5) should be written as follows:

$$\hat{E}\left(f_k, \hat{g}_k, \hat{\zeta}\right) = \frac{1}{2}\left\|\hat{y} - \hat{X}_k\hat{g}_k\right\|_2^2 + \frac{\lambda}{2}\left\|\boldsymbol{m} * f_k\right\|_2^2 + \frac{\gamma}{2}\left\|\hat{M}_{k-1} - \hat{X}_k\hat{g}_k\right\|_2^2 + \frac{\varepsilon}{2}\left\|f_k - f_{k-1}\right\|_2^2 + \hat{\zeta}^T\left[\hat{g}_k - \sqrt{N}\left(I_D \otimes FB^T\right) f_k\right] + \frac{\mu}{2}\left\|\hat{g}_k - \sqrt{N}\left(I_D \otimes FB^T\right) f_k\right\|_2^2 \tag{7}$$

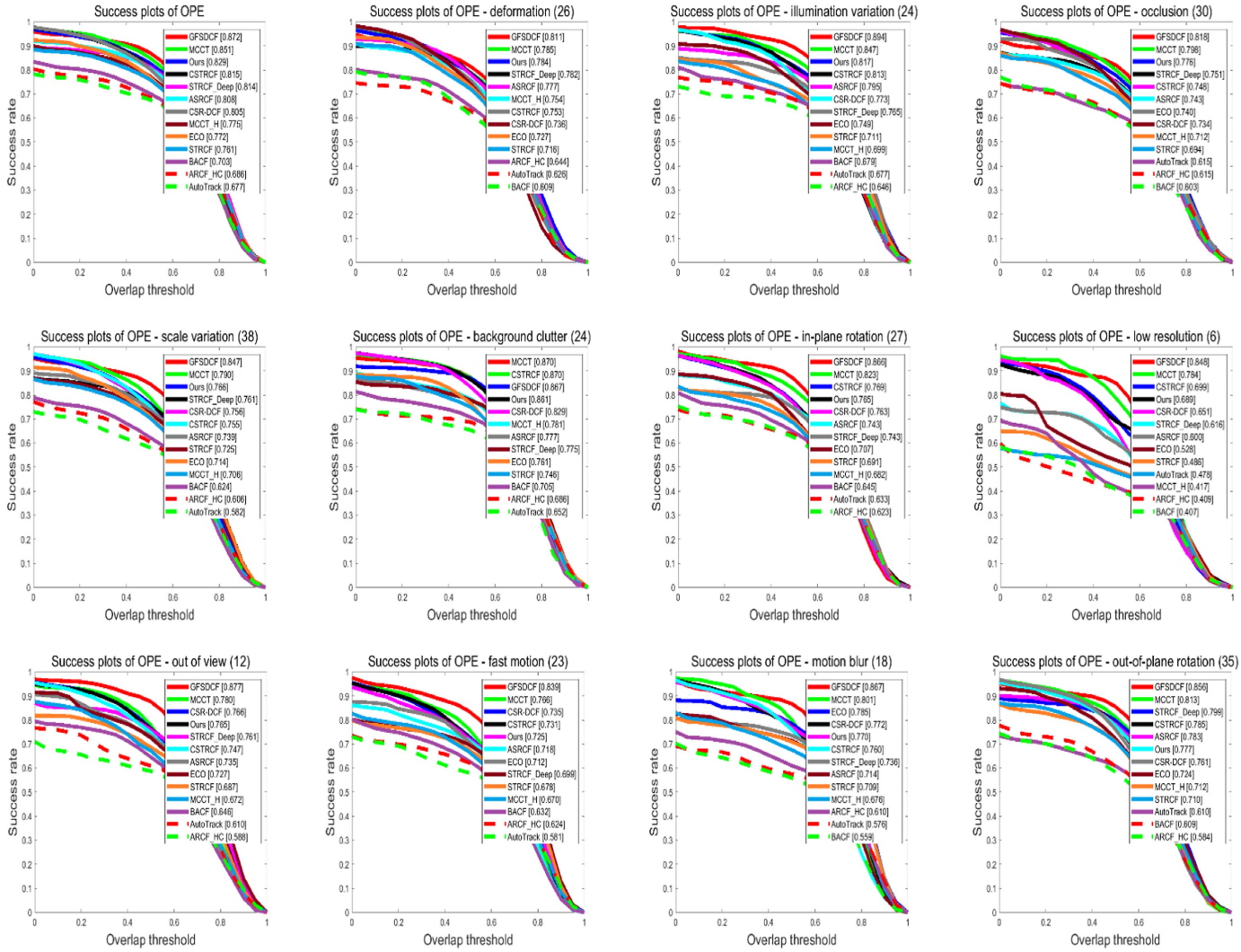

Fig. 2. Success plots on the OTB100.

where μ and the lagrange vector $\hat{\zeta} = [\hat{\zeta}^{1T}, \hat{\zeta}^{2T} \ldots, \hat{\zeta}^{DT}]$ in Fourier domain are the introduced penalty factor and the auxiliary variable.

Employing ADMM in k-th frame to calculate the correlation filters in $(k+1)$-th means that (7) can be solved by solving the following subproblem 1 (solve (8)) and subproblem 2 (solve (9)) iteratively. In each iteration, the values of the filter f and auxiliary variable $\hat{g}_k$ can be obtained by the partial derivatives of (8) to f_k and (9) to $\hat{g}_k$ being equal to zero, respectively.

$$\hat{f}_{k+1} = \underset{f_k}{\operatorname{argmin}} \left\{ \frac{\lambda}{2} \|\boldsymbol{m} * f_k\|_2^2 + \frac{\varepsilon}{2} \|f_k - f_{k-1}\|_2^2 + \hat{\zeta}^T \left[\hat{g}_k - \sqrt{N}\left(I_D \otimes FB^T\right) f_k\right] + \frac{\mu}{2} \|\hat{g}_k - \sqrt{N}\left(I_D \otimes FB^T\right) f_k\|_2^2 \right\} \tag{8}$$

$$\hat{g}_{k+1} = \underset{\hat{g}_k}{\operatorname{argmin}} \left\{ \frac{1}{2} \left\|\hat{y} - \hat{X}_k \hat{g}_k\right\|_2^2 + \frac{\gamma}{2} \left\|\hat{M}_{k-1} - \hat{X}_k \hat{g}_k\right\|_2^2 + \hat{\zeta}^T \left[\hat{g}_k - \sqrt{N}\left(I_D \otimes FB^T\right) f_k\right] + \frac{\mu}{2} \|\hat{g}_k - \sqrt{N}\left(I_D \otimes FB^T\right) f_k\|_2^2 \right\} \tag{9}$$

1) Solution to subproblem 1: The partial derivative of (8) to f_k is

$$\begin{aligned} \frac{\partial \hat{f}_k + 1}{\partial f_k} &= \lambda m(m)^T f_k + \varepsilon\left(f_k - f_{k-1}\right) - \sqrt{N}\left(I_D \otimes FB^T\right) \hat{\zeta} - \sqrt{N}\mu\left(I_D \otimes BF^T\right)\left[\hat{g}_k - \sqrt{N}\left(I_D \otimes FB^T\right) f_k\right] \\ &= \lambda m(m)^T f_k + \varepsilon f_k - \varepsilon f_{k-1} - \sqrt{N}\left(I_D \otimes DTF^T\right)\hat{\zeta} - \sqrt{N}\mu\left(I_D \otimes BF^T\right)\hat{g}_k + N\mu f_k \end{aligned} \tag{10}$$

Let (10) equal to 0, then the closed form solution of (8) can be obtained,

$$\hat{f}_k = \frac{1}{\lambda m(m)^T + \varepsilon + N\mu}\left[\varepsilon f_{k-1} + N\zeta + N\mu g_k\right] \tag{11}$$

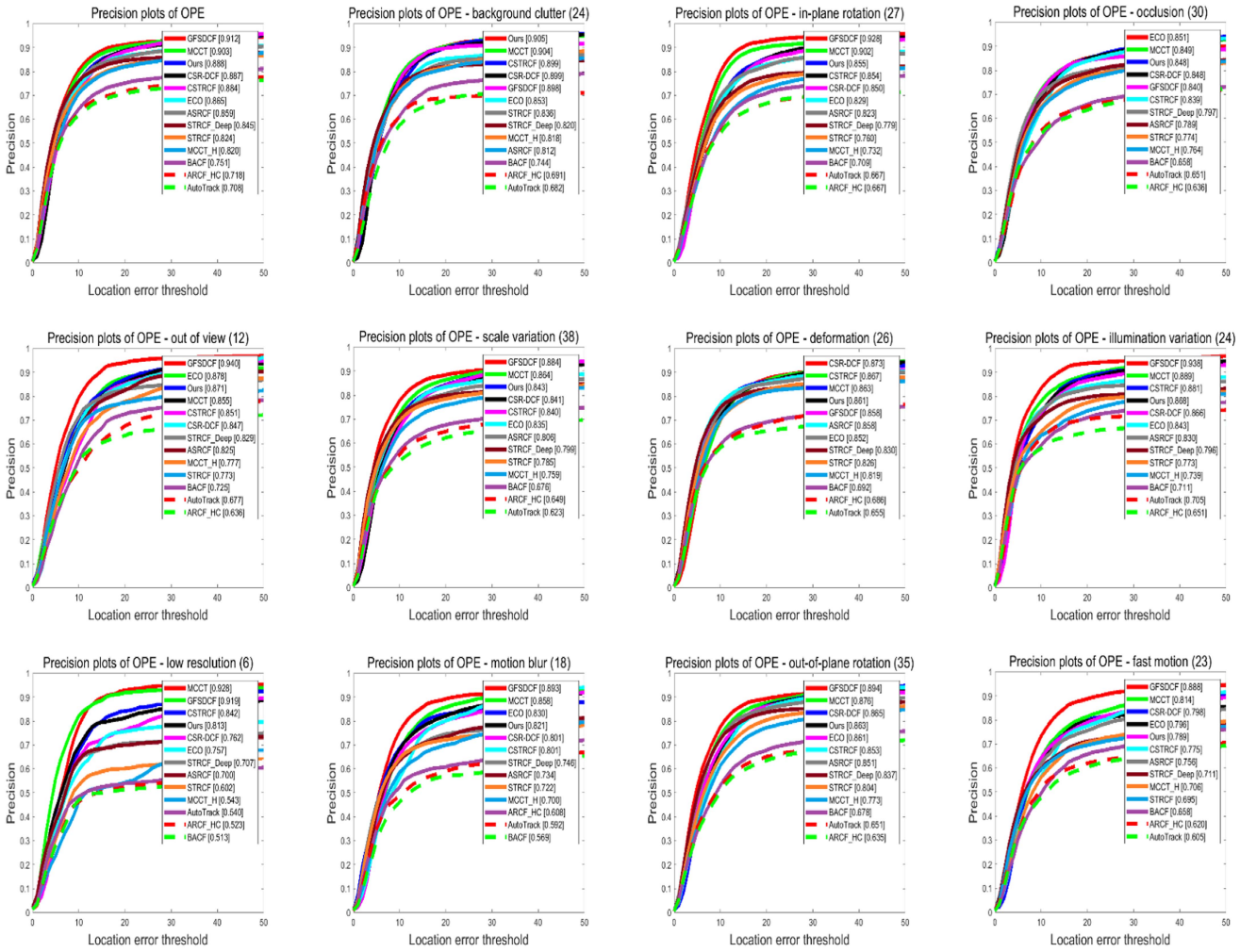

Fig. 3. Precision plots on the OTB100.

2) Solution to subproblem 2: Unlike solving subproblem 1, solving subproblem 2 is highly time consuming. For accelerating the computation, the sparsity of $\hat{X}_k$ is exploited here, it supposes that each element $\hat{y}(n)$ of $\hat{y}$, $n = 1, 2, ..., N$, is solely dependent on each $\hat{X}_k^T = [\hat{X}_k^1(n), \hat{X}_k^2(n), ..., \hat{X}_k^D(n)]^T$ and $\hat{g}_k(n) = [conj(\hat{g}_k^1(n)), conj(\hat{g}_k^2(n)), ..., conj(\hat{g}_k^D(n))]^T$, conj (*) represents a complex conjugate operation, then (9) can be divided into N smaller problems,

$$\hat{g}_{k+1}(n) = \underset{\hat{g}_k(n)}{\operatorname{argmin}} \left\{ \frac{1}{2} \left\| \hat{y}(n) - \hat{X}_k^T(n)\hat{g}_k(n) \right\|_2^2 + \frac{\gamma}{2} \left\| \hat{M}_{k-1}^S - \hat{X}_k^T(n)\hat{g}_k(n) \right\|_2^2 + \hat{\zeta}^T \left[g_k(n) - \hat{f}_k(n) \right] + \frac{\mu}{2} \left\| \hat{g}_k(n) - \hat{f}_k(n) \right\|_2^2 \right\} \tag{12}$$

The partial derivation of (12) to $\hat{g}_k(n)$ is

$$\begin{aligned} \frac{\partial \hat{g}_{k+1}(n)}{\partial \hat{g}_k(n)} = & -\hat{X}_k(n) \left[\hat{y}(n) - \hat{X}_k^T(n)\hat{g}_k(n) \right] - \gamma \hat{X}_k(n) \left[\hat{M}_{k-1} - \hat{X}_k^T(n)\hat{g}_k(n) \right] + \hat{\zeta} + \mu \left[\hat{g}_k(n) - \hat{f}_k(n) \right] \\ = & -\hat{X}_k(n)\hat{y}(n) + \hat{X}_k(n)\hat{X}_k^T(n)\hat{g}_k(n) - \gamma \hat{X}_k(n)\hat{M}_{k-1} + \gamma \hat{X}_k(n)\hat{X}_k^T(n)\hat{g}_k(n) + \hat{\zeta} + \mu \hat{g}_k(n) - \mu \hat{f}_k(n) \end{aligned} \tag{13}$$

Let (13) equal to 0, then the closed form solution of (9) can be obtained,

$$\begin{aligned} \hat{g}_k(n) = & \frac{1}{\mu} \left[\hat{X}_k(n)\hat{y}(n) + \gamma \hat{X}_k(n)\hat{M}_{k-1} - \hat{\zeta} + \mu \hat{f}_k(n) \right] \\ & - \frac{1}{\mu} \frac{\hat{X}_k(n)}{\frac{\mu}{1+\gamma} + \hat{X}_k^T(n)\hat{X}_k(n)} \left[\hat{X}_k^T(n)\hat{X}_k(n)\hat{y}(n) + \gamma \hat{X}_k^T(n)\hat{X}_k(n)\hat{M}_{k-1} - \hat{X}_k^T(n)\hat{\zeta} + \mu \hat{X}_k^T(n)\hat{f}_k(n) \right] \end{aligned} \tag{14}$$

Let $b = \frac{\mu}{1+\gamma} + \hat{X}_k^T(n)\hat{X}_k(n)$, $\hat{S}_{xk}(n) = \hat{X}_k^T(n)\hat{X}_k(n)$, $\hat{S}_{fk}(n) = \hat{X}_k^T(n)\hat{f}_k(n)$, $\hat{S}_\zeta(n) = \hat{X}_k^T(n)\hat{\zeta}$, then (14) can be

TABLE I
THE AVERAGE OVERLAP RATE AND PIXEL ERROR VALUES OF TRACKERS ON THE OTB100 (THE RED, BLUE, AND GREEN NUMBERS DENOTE THE **1ST**, **2ND**, AND **3RD** RANKS SEPARATELY)

Tracker	ARCF_HC	ASRCF	AutoTrack	BACF	CSR-DCF	CSTRCF	ECO	GFSDCF	MCCT	MCCT_H	STRCF	STRCF_Deep	Ours
Overlap Rate (%)	0.570	0.650	0.554	0.576	0.658	**0.659**	0.647	**0.685**	**0.684**	0.630	0.627	0.658	**0.685**
Pixel Error (pixel)	45.7	21.4	53.4	42.1	**16.1**	16.4	21.8	**13.6**	**12.5**	32.1	28.3	23.3	16.8

TABLE II
THE TRACKING SPEED OF TRACKERS ON THE OTB100 (THE RED, BLUE, AND GREEN NUMBERS DENOTE THE **1ST**, **2ND**, AND **3RD** RANKS SEPARATELY)

Tracker	ARCF_HC	ASRCF	AutoTrack	BACF	CSR-DCF	CSTRCF	ECO	GFSDCF	MCCT	MCCT_H	STRCF	STRCF_Deep	Ours
Tracking Speed (fps)	20.9	2.0	**42.5**	**36.3**	3.9	14.6	2.5	0.3	0.9	**35.5**	21.5	6.2	**16.7**

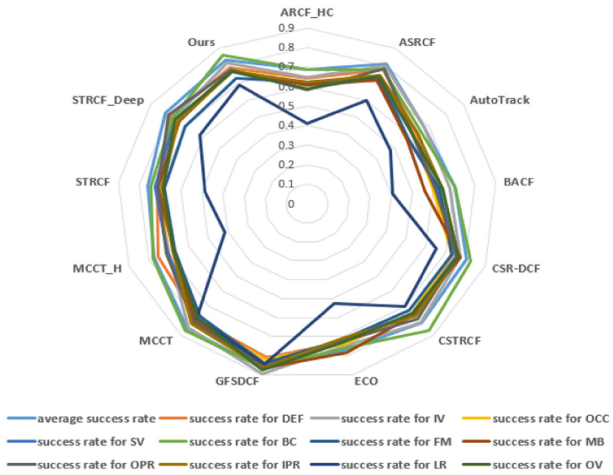

Fig. 4. Radar chart of tracking success rate of trackers on the OTB100.

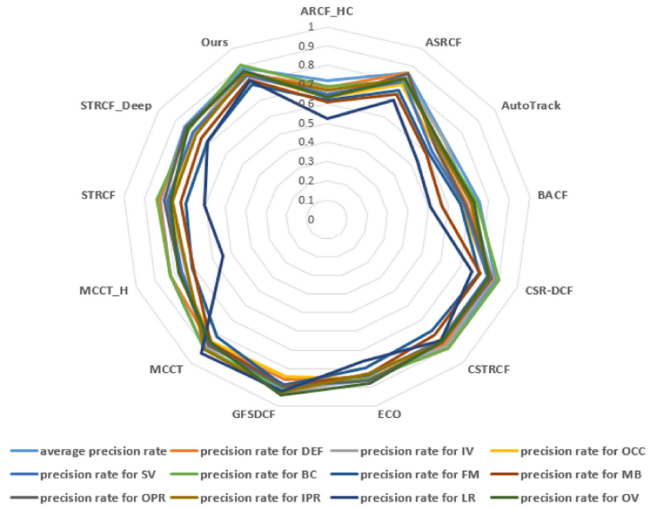

Fig. 5. Radar chart of tracking precision rate of trackers on the OTB100.

written as follows:

$$\hat{g}_{k^{(n)}} = \frac{1}{\mu}\left[\hat{X}_k(n)\hat{y}(n) + \gamma\hat{X}_k(n)\hat{M}_{k-1} - \hat{\zeta} + \mu\hat{f}_k(n)\right] - \frac{\hat{X}k^{(n)}}{\mu b}\left[\hat{S}_{xk}(n)\hat{y}(n) + \gamma\hat{S}_{xk}(n)\hat{M}_{k-1} - \hat{S}\zeta(n) + \mu\hat{S}_{fk}(n)\right] \quad (15)$$

Thus far, the subproblems 1 and 2 are both solved.

3) Parameter update: In this paper, the Lagrange factor ζ is updated as follows:

$$\hat{\zeta}_{k+1}^{j+1} = \zeta_{k+1}^{j} + \mu\left[\hat{g}_{k+1}^{j+1} - \hat{f}_{k+1}^{j+1}\right] \quad (16)$$

where the subscript j and $(j+1)$ denotes the j-th and the $(j+1)$-th iteration respectively.

The object appearance model x is updated as follows:

$$\hat{x}_{k+1} = (1-\eta)\hat{x}_{k+1} + \eta\hat{x}_k \quad (17)$$

where ζ is a appearance model learning rate.

The penalty factor μ is updated according to the following equation:

$$\mu(j+l) = \min\left(\mu_{\max}, \beta\mu^{j}\right) \quad (18)$$

where $\mu_{\max}$ is the maximum value of μ, and β is the scale factor with a constant value.

A brief overview of the proposed tracking model is given in Algorithm 1.

IV. EXPERIMENTS

In this section, the proposed tracking model is evaluated extensively on the data sets of OTB100 [22], VOT2016 [23], TC128 [24], and UAV123 [25], and compared with several state-of-the-art tracking models in terms of the precision plots of OPE (one-pass evaluation), success plots of OPE, accuracy rate, overlap rate between the tracking bounding-box and the ground-truth, pixel error between tracking bounding-box center and the ground-truth center, and tracking speed, etc., where the precision plot is a plot of curves below different center location error thresholds (20 pixels in this paper) and the success plot is a plot of curves below different overlap thresholds (0.5 in this paper). The conventional method to evaluate trackers is to run them throughout a dataset with object initialization, this is referred to be OPE. All experiments were conducted with MATLAB R2018a on a platform equipped with a 4 GHz Intel Core i7 processor and 8 GB RAM.

The values of the related hyperparameters in the proposed tracking model were set as below: the regularized factors $\lambda = 0.001$, $\epsilon = 16$; the aberrance penalty factor $\gamma = 0.71$; the initial value of the penalty factor μ was 1, and $\mu_{\max} = 103$; the bject appearance model learning rate $\eta = 0.0125$; and the scale factor

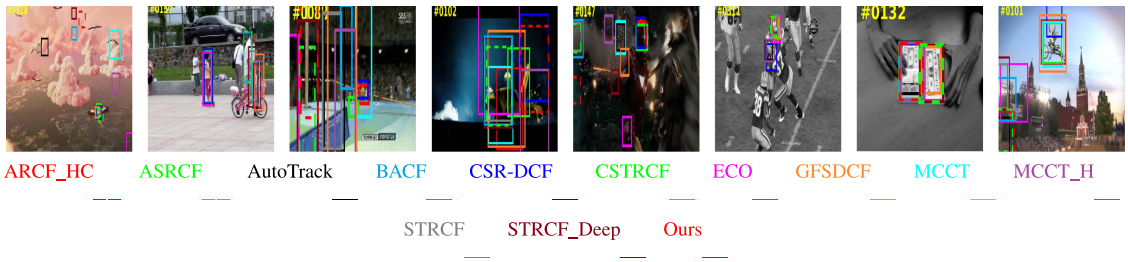

Fig. 6. Tracking results in some image sequences of the OTB100.

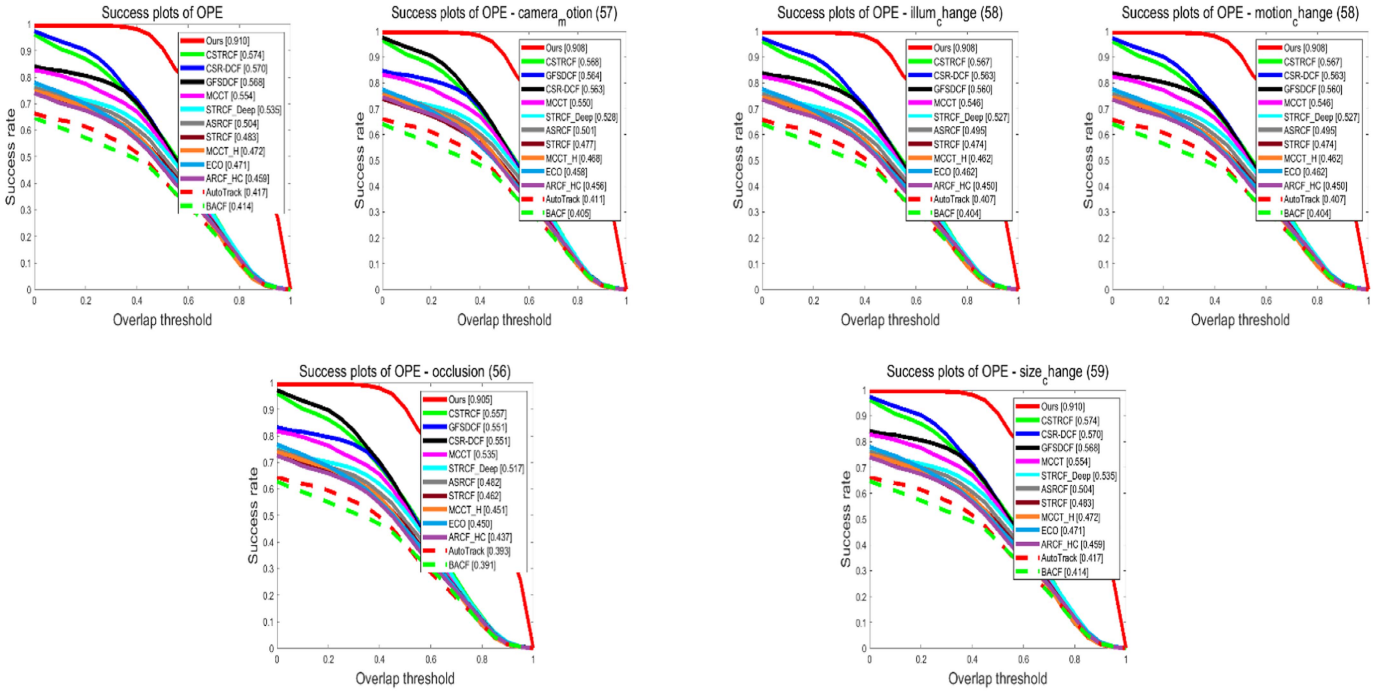

Fig. 7. Success plots on the VOT2016.

$\beta = 10$. All of the mentioned parameters remained unchanged throughout the experiments.

A. Compared Tracking Model

To fully evaluate the performance of the proposed tracking model, extensive experiment results thereof on the OTB100 [22], VOT2016 [23], TC128 [24], and UAV123 [25] data sets were compared with those of several state-of-the-art tracking models [10], [14], [15], [16], [17], [18], [19], [20], [21], [27], [46], [54], which are detailed below.

1) In the HOG feature based BACF [10], by zero padding operation, the negative samples are obtained to include a larger search area and more real backgrounds.
2) In the HOG and CN features based STRCF [14], DCFs learning and updating are done concurrently, and STRCF_Deep [14] is the CNN features based version.
3) In the HOG and CN features based CSR-DCF [15], the reliabilities of channel and spaial are introduced to train CFs.
4) In the CN, HOG, intensity channels (IC) and CNN features based GFSDCF [16], a tracking model for joint group feature selection across both channel and spatial dimensions is established.
5) In MCCT [17], multiple DCF based experts are used to track the objects, and each independent expert is constructed with different combinations of deep and HOG features. The MCCT_H [17] is a MCCT version based on different combinations of CN and HOG features.
6) In ASRCF [18], two CF models are utilized to estimate the object position and scale respectively, and the combination of HOG and deep features is adopted to train the position estimation CF. To train the scale estimation CF, a combination of five scales of HOG features is employed.

Algorithm 1: Object Tracking Based on Correlation Filter.

Input: the tracking object coordinate value in the first frame (r_0, c_0, w_0, h_0) or in the previous frame $(r_{t-1}, c_{t-1}, w_{t-1}, h_{t-1})$;
Output: the tracking object coordinate value in the current frame (r_t, c_t, w_t, h_t);

for frame=1:total_frames
  if frame =1
    Input the tracking object coordinate value (r_0, c_0, w_0, h_0) in the frame;
  else
    Input the tracking object coordinate value $(r_{t-1}, c_{t-1}, w_{t-1}, h_{t-1})$ obtained by our tracking model in the previous frame;
  end
  Calculate the foreground color histogram $hist\_fg$ and background color histogram $hist\_bg$ of the object and its surrounding areas in the current frame in HSV color space;
  Calculate the foreground prior probability f_p of the object;
  Calculate the spatial reliability map $\boldsymbol{m}$ of the object from $hist\_fg$, $hist\_bg$, and f_p;
  Extract the normalized combination feature x_t of HOG and CN features of the object;
  Train the correlation filter f_t according to (11);
  Calculate the auxiliary variable g_t according to (15);
  Calculate the object response $R_t = x_t * f_t$, the position where the maximum response value R_{max} of R_t located is the object center (r_c, c_c) calculated by our tracking model;
  Estimate the object scale (w_t', h_t') using a single-scale spatial correlation filter;
  Output the coordinate value (r_t, c_t, w_t, h_t) of the object in the current frame calculated by the proposed tracking model, where $w_t = r_c - (w_t'/2)$, $h_t = c_c - (h_t'/2)$.
end

7) The number of parameters is significantly reduced in ECO [19] by introducing a factor convolution operator into the objective function of the standard DCF, and the CNN, HOG, and CN features of the objects are adopted to train the CF.
8) In AutoTrack [20], to let DCF predominantly learn the reliable part of the object, the spatially local response map variation is introduced as spatially-regularized.
9) In ARCF_HC [21], the HOG, CN, and grayscale features are employed to train CF. To expand the object search areas, the background patches are used as negative samples.
10) The HOG and CN features of the object are used to train the correlation filter in CSTRCF [46], which improves the object tracking performance by introducing a temporal regularized term into the objective function of CSR-DCF [15].
11) As a two-stage object detection approach for visual object tracking, Siam R-CNN [54] operates in combination with a novel trajectory based dynamic programming algorithm, and can redetect the tracked object after long occlusion and is particularly effective for long-term tracking.
12) Composed of one multi-layer aggregation module and one depthwise correlation layer, SiamRPN++ [27] assembles the hierarchy of connections through the aggregation module thereof to aggregate different levels of representation, and the depthwise correlation layer allows the network to reduce the computation cost and redundant parameters.

B. OTB100 Data Set

The OTB100 data set [22] contains 100 fully annotated video sequences with 11 different attributes, including illumination variation (IV), scale variation (SV), occlusion (OCC), deformation (DEF), motion blur (MB), fast motion (FM), in-plane rotation (IPR), out-of-plane rotation (OPR), out-of-view (OV), background clutters (BC), and low resolution (LR). The proposed tracking model was evaluated on this data set, and the precision and success plots of OPE, overlap rate (the overlap degree between the obtained tracking bounding-box value and the real bounding-box value, and the greater the better), tracking speed, and pixel error (the distance between the obtained tracking bounding-box center and the real bounding-box center, and the smaller the better) were adopted to evaluate the performance thereof with several state-of-the-art tracking models, such as BACF [10], STRCF [14], STRCF_Deep [14],GFSDCF [16], MCCT [17], MCCT_H [17], ASRCF [18], ECO [19], AutoTrack [20], ARCF_HC [21], CSR-DCF [15], CSTRCF [46]. Fig. 2 illustrates the success plots under different overlap thresholds of each tracker on the OTB100 [22], and Fig. 4 shows its corresponding radar performance chart. As shown in the Figs. 2 and 4, when tracking the objects with attributes of DEF, IV, OCC, and SV, the proposed tracking model was superior to AutoTrack [20], STRCF_Deep [14], CSTRCF [46], CSR-DCF [15] and others. For the objects with attributes of BC, IPR, LR, and OV, the proposed model also demonstrated good tracking performance. Overall, the tracking success rate of the proposed model on the OTB100 [22] ranked third with an average score of 0.829, being only 0.043 lower than the first-ranked GFSDCF [16], and 0.022 lower than the second-ranked MCCT [17].

The precision plots under different location error thresholds of each tracker for the tracking objects with different attributes are presented in Figs. 3 and 5 shows its corresponding radar performance chart. Compared with the 12 state-of-the-art trackers, the tracking precision of the proposed tracking model was excellent for the objects with DEF, OCC, and SV attributes, all

TABLE III
THE TRACKING ACCURACY RATE AND A-R RANK VALUES OF TRACKERS ON VOT2016 (THE RED, BLUE, AND GREEN NUMBERS DENOTE THE 1ST, 2ND, AND 3RD RANKS SEPARATELY)

Tracker	camera_motion	illu_change	motion_change	occlusion	size_change	Mean	A-R rank
ARCF_HC	0.42	0.56	0.28	0.23	0.38	0.38	0.38
AutoTrack	0.32	0.53	0.26	0.21	0.33	0.34	0.32
BACF	0.37	0.54	0.26	0.20	0.34	0.34	0.32
CSR-DCF	**0.51**	**0.62**	**0.45**	**0.40**	**0.49**	**0.49**	**0.50**
CSTRCF	0.49	**0.64**	**0.46**	**0.42**	0.48	**0.50**	**0.49**
ECO	0.44	0.59	0.35	0.29	0.46	0.43	0.42
GFSDCF	**0.50**	0.61	0.44	0.32	**0.49**	0.47	0.47
MCCT	0.49	0.58	0.39	0.29	0.46	0.45	0.45
MCCT_H	0.43	0.51	0.34	0.27	0.42	0.41	0.41
STRCF	0.41	0.58	0.32	0.25	0.40	0.40	0.39
STRCF_Deep	0.45	0.59	0.35	0.30	0.42	0.43	0.43
Ours	**0.68**	**0.75**	**0.63**	**0.73**	**0.70**	**0.70**	**0.69**

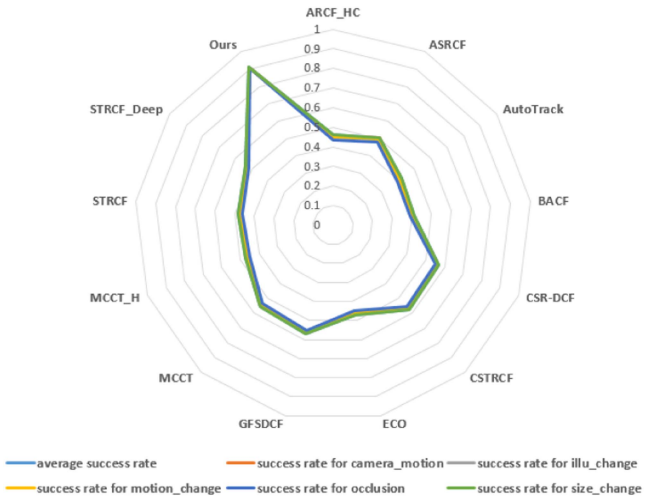

Fig. 8. Radar chart of tracking success rate of trackers on the VOT2016.

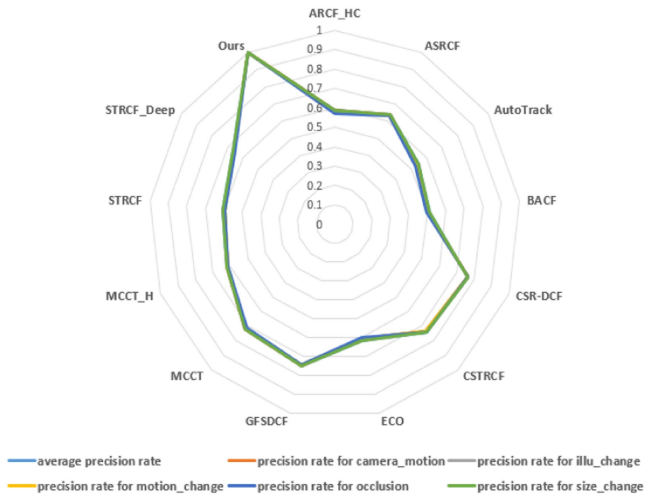

Fig. 9. Radar chart of tracking precision rate of trackers on the VOT2016.

ranking in the top 4. Moreover, the proposed model also exhibited superior tracking accuracy for the objects with attributes of BC, IPR, OV, IV, LR, MB, and OPR. Even for the objects with FM attribute, the precisions of the proposed model fell to some degree, yet the worst ranking thereof was fourth, which was still better than several state-of-the-art trackers, such as AutoTrack [20], STRCF_Deep [14], and CSTRCF [46], etc.

From Figs. 2, 3, 4, and 5, an observation can be made that the proposed tracking model in this paper was outstanding in terms of tracking success rate and tracking precision for the objects in the image sequences of the OTB100 [22] with attributes DEF, OCC, and SV. At the same time, its tracking performance for the objects with other attributes was also worthy of affirmation.

Table I displays the bounding-box average overlap rate between the tracking results and the ground-truth, in addition to the pixel errors between the center of the two aforementioned bounding-boxes of the trackers on the OTB100 [22]. An observation can be made that the proposed tracking model not only ranked first in the average overlap rate, but also had a small pixel error. Table II shows the tracking speed statistics of each tracker on the OTB100 [22]. Although the proposed tracker was generally inferior to GFSDCF [16] and MCCT [17] in success and precision plots, it can be seen from Table II that the tracking speed of these two trackers was far lower than that of the proposed. The proposed tracker had more practical application value due to its certain real-time tracking performance.

The object tracking results of several state-of-the-art trackers and the proposed tracking model in several OTB100 [22] frames are reported in Fig. 6, in which an observation can be made that the proposed tracking model could still track the SV, OCC, DEF, IPR, LR, BC, and other attributes objects accurately, in the case

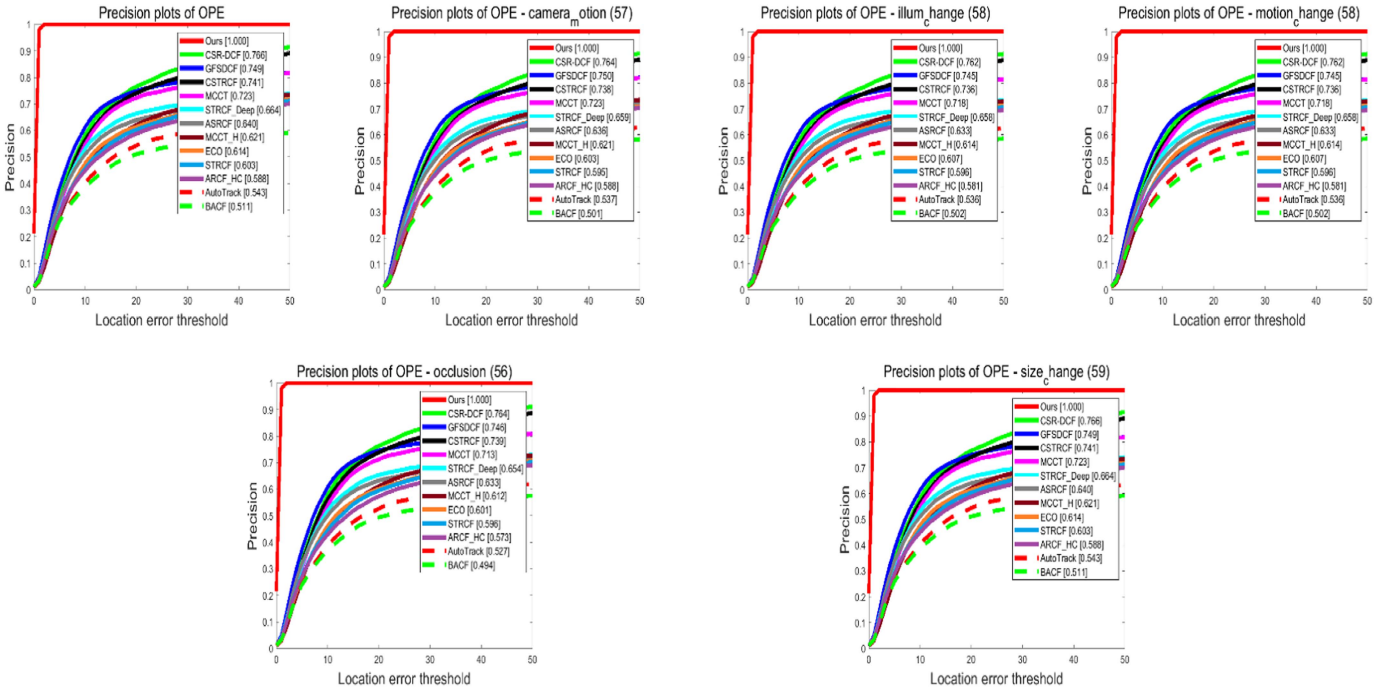

Fig. 10. Precision plots on the VOT2016.

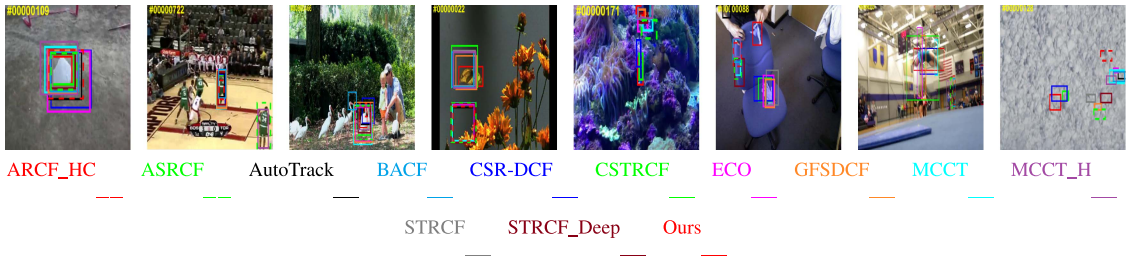

Fig. 11. Object tracking results in some image sequences of the VOT2016.

of GFSDCF [16], AutoTrack [20], BACF [10], CSR-DCF [15], MCCT [17], and STRCF_Deep [14] all tracking failure.

C. VOT2016 Data Set

60 challenging and publicly available video sequences are contained in the VOT2016 data set [23], in which the kinds of the tracking objects are various, including toys, faces, vehicles, animals, etc., which are labeled with camera_motion, illu_change, motion_change, occlusion, and size_change. In this section, the experiment of the proposed tracking model on the VOT2016 [23] is discussed, to evaluate the tracking performance, and the tracking accuracy rate, A-R rank, success plots, precision plots, overlap rate, tracking speed, and pixel errors are used.

The tracking accuracy and A-R rank values of each tracker in the video sequences of the VOT2016 data set [23] are shown in Table III. As can be seen from the data in Table III, the proposed tracking model could still track the objects with camera_motion, illu_change, motion_change, occlusion, and size_change attributes accurately. Better yet, the average tracking accuracy of the proposed tracker was 2% higher than that of the second CSTRCF [46].

The success plots of OPE of the trackers on the VOT2016 data set [23] are shown in Fig. 7 and 8 shows its corresponding radar performance chart. It can be seen from Figs. 7 and 8 that the proposed tracking model had the highest tracking success rates, which were all over 90%, for all objects with occlusion, size change, and other attributes in the VOT2016 data set [23], compared with all the listed state-of-the-art trackers in this paper.

Fig. 10 illustrates the tracking precision plots of each tracker on the VOT2016 data set [23], and Fig. 9 shows its corresponding radar performance chart. As can be seen from Figs. 9 and 10 that the tracking accuracy values of the proposed tracking model were all 1, no matter what attributes the object was, indicating that once the object was successfully tracked, the proposed tracking model could always track the object in the VOT2016 data set [23] completely correctly.

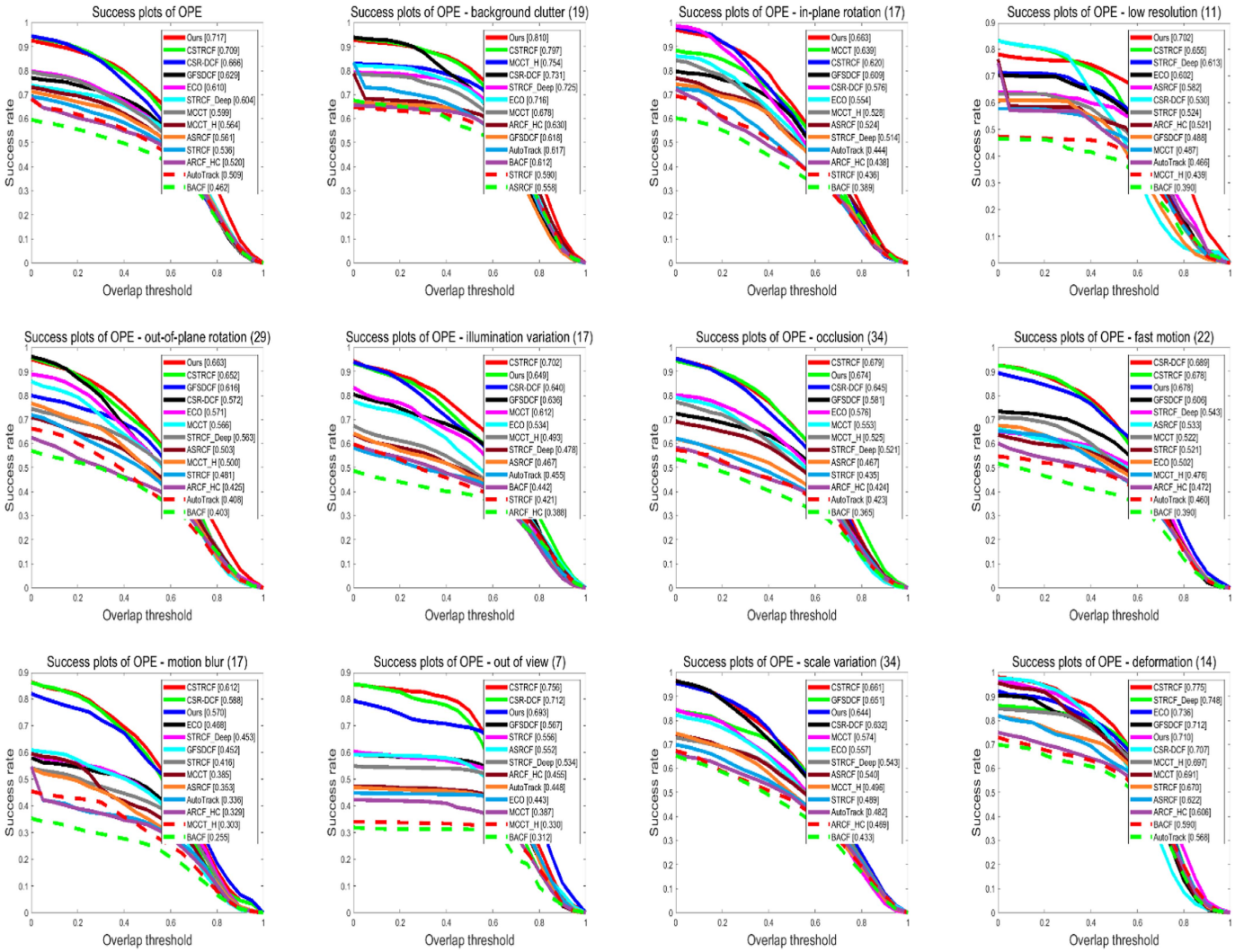

Fig. 12. Success plots on the TC128.

TABLE IV
THE AVERAGE OVERLAP RATE AND PIXEL ERROR VALUES OF TRACKERS ON THE VOT2016 (THE RED, BLUE, AND GREEN NUMBERS DENOTE THE **1ST**, **2ND**, AND **3RD** RANKS SEPARATELY)

Tracker	ARCF_HC	ASRCF	AutoTrack	BACF	CSR-DCF	CSTRCF	ECO	GFSDCF	MCCT	MCCT_H	STRCF	STRCF_Deep	Ours
Overlap Rate (%)	0.412	0.427	0.373	0.358	**0.524**	**0.521**	0.426	0.491	0.472	0.418	0.410	0.453	**0.790**
Pixel Error (pixel)	75.0	73.5	90.6	94.3	163.1	**28.4**	72.8	211.0	**52.1**	76.6	73.2	65.3	**0.5**

TABLE V
THE TRACKING SPEED OF TRACKERS ON THE VOT2016 (THE RED, BLUE, AND GREEN NUMBERS DENOTE THE **1ST**, **2ND**, AND **3RD** RANKS SEPARATELY)

Tracker	ARCF_HC	ASRCF	AutoTrack	BACF	CSR-DCF	CSTRCF	ECO	GFSDCF	MCCT	MCCT_H	STRCF	STRCF_Deep	Ours
Tracking Speed (fps)	12.4	1.3	**20.9**	**15.7**	8.2	9.7	1.9	0.3	0.9	**18.9**	12.6	2.3	**10.1**

The average overlap rate and pixel errors of the trackers on the VOT2016 data set [23] are presented in Table IV, in which the proposed tracking model ranked first in terms of average overlap rate and pixel errors, with 79% and a significant 0.5 pixels, respectively, which were considerably better than the other trackers.

Table V shows the average tracking speed statistics of each tracker on the VOT2016 [23]. According to the data in the table, we can see that the object tracking speed of the proposed tracker in the VOT2016 [23] was better than some state-of-the-art trackers, such as CSR-DCF [15], CSTRCF [46], GFSDCF [16], MCCT [17], and ECO [19], etc., and had more practical

TABLE VI
THE AVERAGE OVERLAP RATE AND PIXEL ERROR VALUES OF TRACKERS ON THE TC128 (THE RED, BLUE, AND GREEN NUMBERS DENOTE THE **1ST**, **2ND**, AND **3RD** RANKS SEPARATELY)

Tracker	ARCF_HC	ASRCF	AutoTrack	BACF	CSR-DCF	CSTRCF	ECO	GFSDCF	MCCT	MCCT_H	STRCF	STRCF_Deep	Ours
Overlap Rate (%)	0.430	0.470	0.430	0.390	**0.570**	**0.590**	0.520	0.510	0.500	0.470	0.450	0.500	**0.620**
Pixel Error (pixel)	79.7	62.7	78.2	88.0	160.8	**26.4**	**41.6**	139.8	44.6	106.8	75.6	55.8	**34.3**

TABLE VII
THE TRACKING SPEED OF SOME TRACKING ALGORITHMS ON THE TC128 (THE RED, BLUE, AND GREEN NUMBERS DENOTE THE **1ST**, **2ND**, AND **3RD** RANKS SEPARATELY)

Tracker	ARCF_HC	ASRCF	AutoTrack	BACF	CSR-DCF	CSTRCF	ECO	GFSDCF	MCCT	MCCT_H	STRCF	STRCF_Deep	Ours
Tracking Speed (fps)	7	1.2	**26.3**	**27.2**	9.6	9.6	4.2	0.3	0.9	**24.5**	16.1	2.3	**18.5**

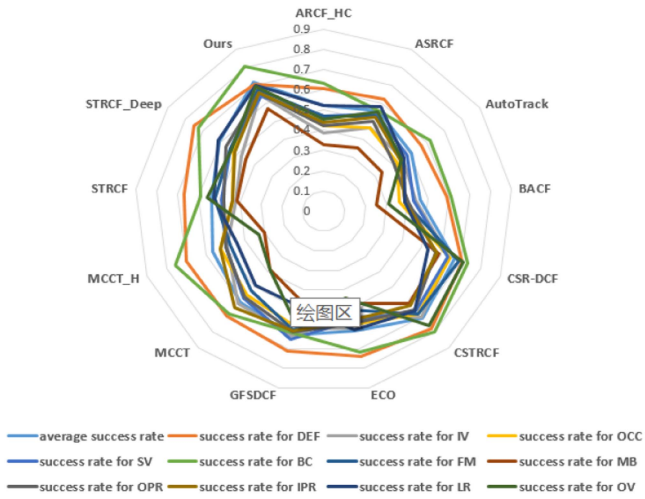

Fig. 13. Radar chart of tracking success rate of trackers on the TC128.

application values than them. Fig. 11 lists the tracking results of each tracker for the tracking objects that are difficult to track in the VOT2016 data set [23]. An observation can be made that the proposed tracking model could track the object accurately no matter what attributes the object was, with a performance far superior than other trackers.

D. TC128 Data Set

There are 128 color sequences in the TC128 data set [24], and the object in each sequence has been labeled with ground-truth and its attributes, including IV, SV, OCC, DEF, MB, FM, and IPR, etc.

The success plots for the objects with different attributes on the TC128 data set [24] are shown in Figs. 12 and 13 shows its corresponding radar performance chart. It can be observed from Figs. 12 and 13 that the proposed tracking model had the highest tracking success rate for the objects with attributes of BC, IPR, LR, and OPR in the TC128 data set [24]. For the objects with IV and OCC attributes, the proposed model was second only to CSTRCF [46], while the proposed model also had a good tracking success rate for the objects of which the attributes were FM, MB, OV and SV. Even for DEF objects, the tracking success rate of the proposed model was ranked fifth, which was still higher than AutoTrack [20], STRCF_Deep [14], ASRCF [18], CSR-DCF [15], and other trackers.

Fig. 14 shows the precision plots of the trackers on the TC128 data set [24] red, and Fig. 15 shows its corresponding radar performance chart. As can be seen from Figs. 14 and 15 that when tracking the IV, IPR, OCC, OPR, and SV objects, the tracking accuracy of the proposed tracking model was better than other trackers. Moreover, for the objects with BC, FM, LR, MB, and OV attributes, the tracking accuracy of the tracking model was far superior to other trackers except CSTRCF [46] and CSR-DCF [15].

The average overlap rate and pixel errors of each tracker on the TC128 data set [24] are listed in Table VI. We can see that the proposed tracking model was superior to other trackers in average overlap rate, ranking first with 62%; the pixel errors of it was second only to CSTRCF [46], with 34.3 pixels, which was also excellent.

Table VII shows the average tracking speed statistics of each tracker on the TC128 [24]. According to the data in the table, we can see that the object tracking speed of the proposed tracker on the TC128 [24] was more better than some state-of-the-art trackers, such as CSR-DCF [15], CSTRCF [46], GFSDCF [16], MCCT [17], and ECO [19], etc.

Fig. 16 shows partial tracking results of several state-of-the-art trackers on the TC128 [24], in which an observation can be made that the proposed could accurately track the objects with IV, OCC, BC, FM, MB, OPR, LR, and other attributes, and sometimes it was the only tracker that can did that, and its tracking performance was better than CSTRCF [46], CSR-DCF [15], GFSDCF [16], and STRCF_Deep [14], etc.

E. UAV123 Data Set

123 challenging video sequences obtained by unmanned aerial vehicles (UAV) are contained in the UAV123 data set [25], in which the object is labeled with ground-truth and some attributes, such as IV, SV, Partial Occlusion (POC), Full Occlusion (FOC), OV, FM, Camera Motion (CM), Similar Object (SOB), Aspect Ratio Change (ARC), Viewpoint Change (VC), BC, and LR.

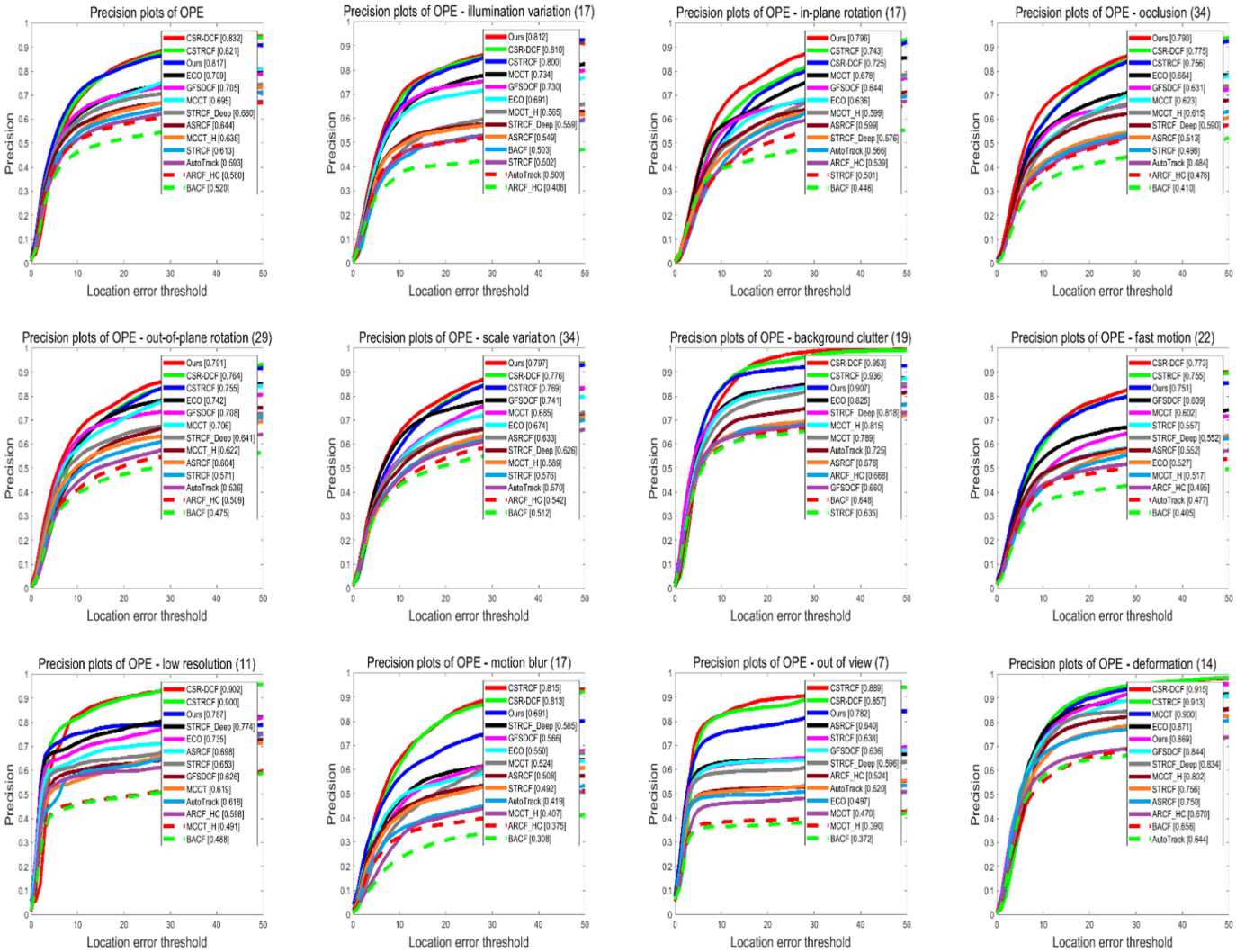

Fig. 14. Precision plots on the TC128.

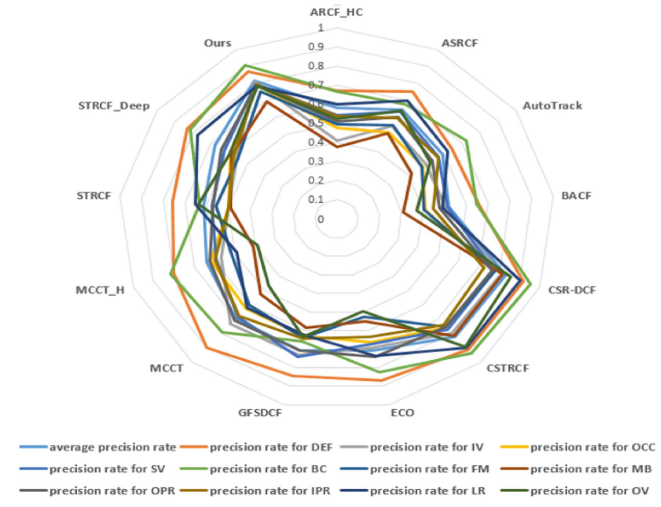

Fig. 15. Radar chart of tracking precision rate of trackers on the TC128.

Fig. 17 shows the success plots of each tracker for different attributes objects in the UAV123 [25], and Fig. 19 shows its corresponding radar performance chart. As can be seen from Figs. 17 and 19 that the proposed tracking model had the highest tracking success rate for the BC objects. In addition, for the objects with attributes of ARC, IV, LR, OV, POC, SV, SOB, and VC, the tracking success rates of the proposed tracking model were second only to CSTRCF [46], SiamRCNN [54], and SiamRPN++ [27]. Even for the objects with CM and FM attributes, the tracking success rates of the proposed tracking model were ranked fifth, which were better than trackers such as CSR-DCF [15], MCCT [17], ASRCF [18], STRCF_Deep [14], and AutoTrack [20], etc.

Fig. 18 shows the precision plots on the UAV123 data set [25], and Fig. 20 shows its corresponding radar performance chart. It can be seen from Figs. 18 and 20 that the proposed tracking model had the highest tracking precision for the IV objects in the UAV123 data set [25] with 88.9%; for the objects with POC, SV, and SOB attributes, the proposed had better

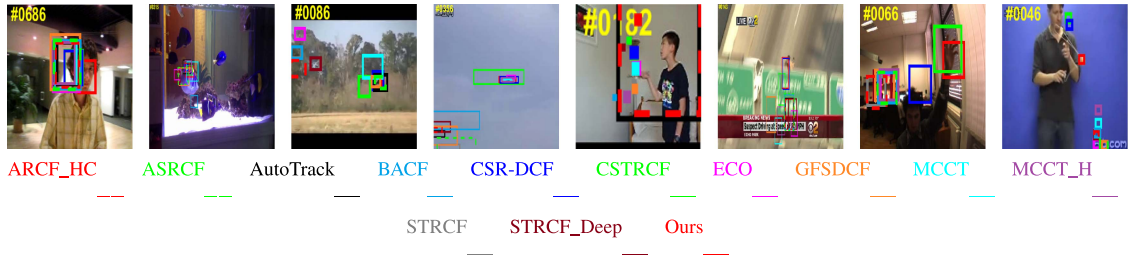

Fig. 16. Object tracking results in some image sequences of the TC128.

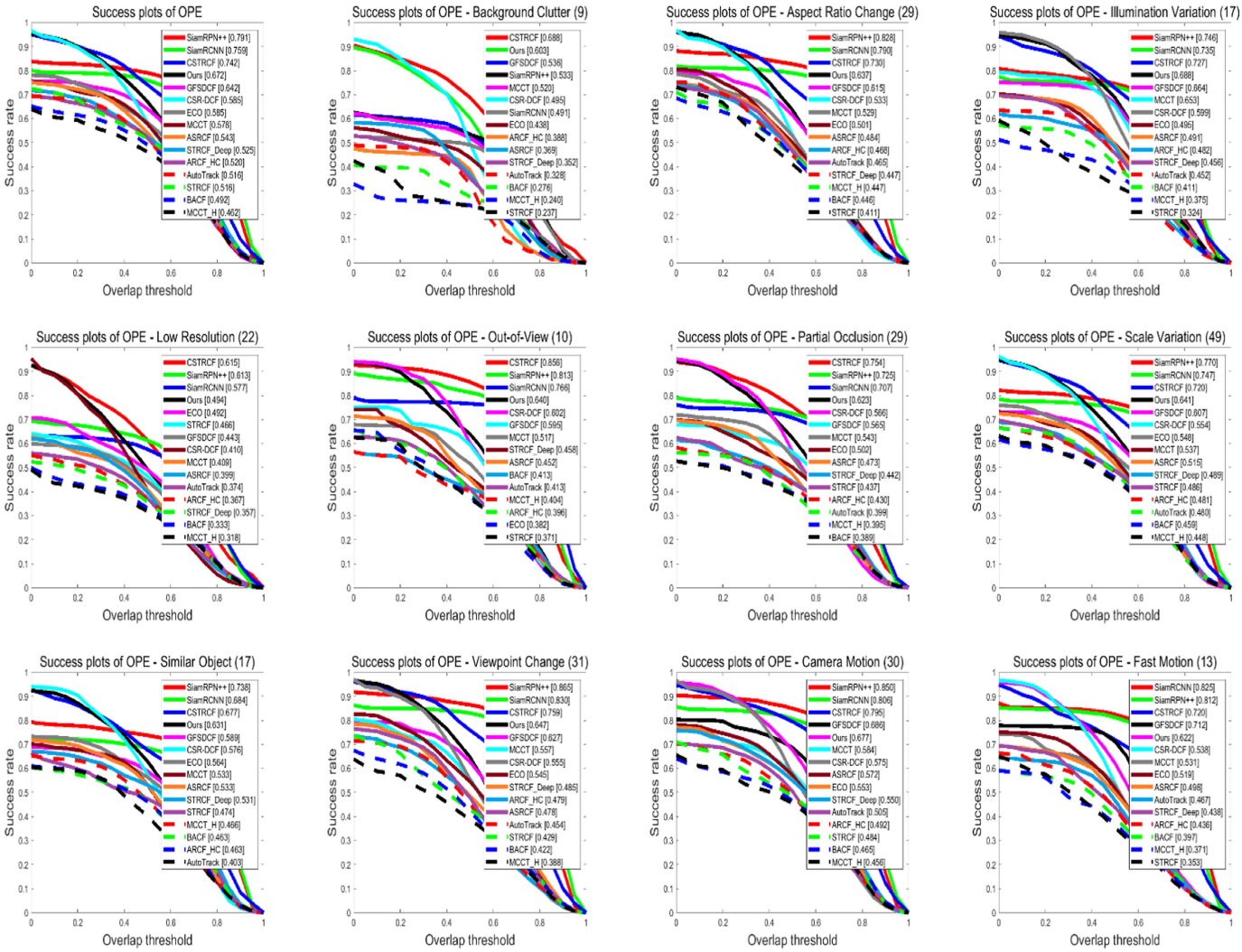

Fig. 17. Success plots on the UAV123.

tracking accuracy than SiamRCNN [54], SiamRPN++ [27], and GFSDCF [16], except for CSTRCF [46] or CSR-DCF [15]; for the objects with BC, ARC, LR, CM, FM, OV, and VC attributes, although the tracking precisions of the proposed tracker ranked fourth, they were higher than most of the trackers, such as CSR-DCF [15], MCCT [17], ASRCF [18], and STRCF_Deep [14], etc.

Table VIII displays the average overlap rate and pixel errors of some trackers on the UAV123 [25]. Although the average overlap rate of the proposed tracking model was not as good as SiamRPN++ [27], SiamRCNN [54], and CSTRCF [46], yet the worst ranking thereof was fourth, which was still better than most trackers. The proposed tracking model ranked first with 23.8 pixel errors.

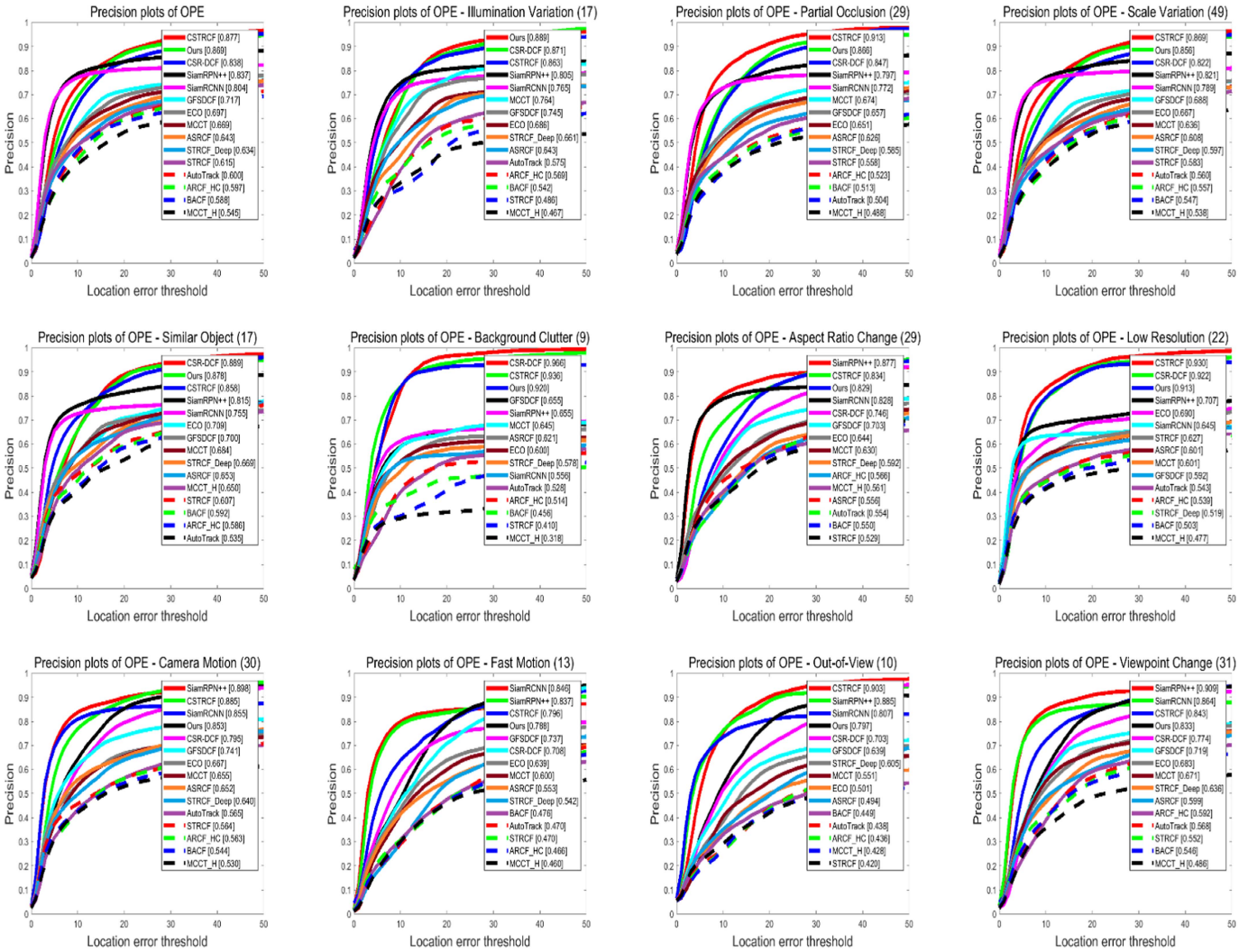

Fig. 18. Precision plots on the UAV123.

TABLE VIII
THE AVERAGE OVERLAP RATE AND PIXEL ERROR VALUES OF TRACKERS ON THE UAV123 (THE RED, BLUE, AND GREEN NUMBERS DENOTE THE **1ST**, **2ND**, AND **3RD** RANKS SEPARATELY)

Tracker	ARCF_HC	ASRCF	AutoTrack	BACF	CSR-DCF	CSTRCF	ECO	GFSDCF	MCCT	MCCT_H	SiamRCNN	SiamRPN++	STRCF	STRCF_Deep	Ours
Overlap Rate (%)	0.470	0.490	0.480	0.470	0.550	**0.610**	0.540	0.530	0.500	0.410	**0.670**	**0.690**	0.500	0.490	0.600
Pixel Error (pixel)	64.4	68.2	78.5	74.4	147.4	83.9	263.7	65.5	54.3	93.3	**52.3**	**31.0**	65.4	76.6	**23.8**

TABLE IX
THE TRACKING SPEED OF SOME TRACKING ALGORITHMS ON THE UAV123 (THE RED, BLUE, AND GREEN NUMBERS DENOTE THE **1ST**, **2ND**, AND **3RD** RANKS SEPARATELY)

Tracker	ARCF_HC	ASRCF	AutoTrack	BACF	CSR-DCF	CSTRCF	ECO	GFSDCF	MCCT	MCCT_H	STRCF	STRCF_Deep	Ours
Tracking Speed (fps)	15.5	1.3	**24.5**	**27.1**	10.6	11	1.9	0.3	0.9	**23.5**	16.4	2.4	**12.2**

Table IX shows the average tracking speed statistics of each tracker on the UAV123 [25]. According to the data in the table, it can be seen that the object tracking speed of the proposed tracker in the UAV123 [25] was more better than some state-of-the-art trackers, such as CSTRCF [46], GFSDCF [16], and MCCT [17], etc., and it had certain real-time tracking abilities.

Fig. 21 shows the tracking results of the trackers on the UAV123 [25]. As can be seen from Fig. 21 that the proposed tracking model had excellent tracking performance for the objects with attributes of IV, SV, POC, SOB, BC, and LR, and in some cases, it even had better tracking performance than SiamRPN++ [27], SiamRCNN [54], and CSTRCF [46].

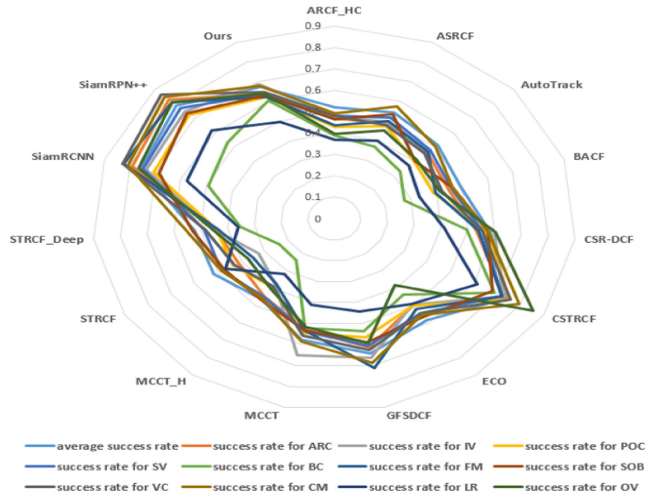

Fig. 19. Radar chart of tracking success rate of trackers on the UAV123.

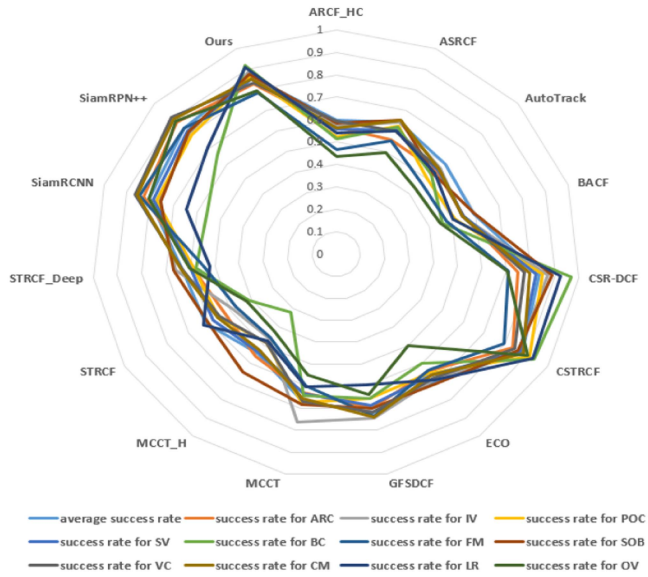

Fig. 20. Radar chart of tracking precision rate of trackers on the UAV123.

F. Ablation Study

To get a comprehensive understanding of the influences of different components in the proposed object tracking models on its tracking performances, ablation study is usually required. Considering that the performance of the three components, i.e., the temporal regularized term $\|f_t - f_{t-1}\|_2^2$, the aberrance repression term $\|\sum_{d=1}^{D}(Bx_{k-1}^d * f_{k-1}^d)[\psi_{p,q}] - \sum_{d=1}^{D} Bx_k^d * f_k^d\|_2^2$, and the spatial reliability map $\mathbf{m}$, used in this paper has been verified in STRCF [14], CSR-DCF [15], and ARCF_HC [21], respectively, and these algorithms have been compared with the proposed in this paper, it is unnecessary to conduct ablation experiments to emphasize the three components, but only show the applications of the three components in relevant algorithms in Table X.

TABLE X
THE USAGE OF THE TEMPORAL REGULARIZED TERM, THE ABERRANCE REPRESSION TERM, AND THE SPATIAL RELIABILITY MAP IN STRCF, CSR-DCF, ARCF_HC, AND THE PROPOSED TRACKING ALGORITHMS

Algorithm / Term	temporal regularized term	aberrance repression term	spatial reliability map
ARCF_HC	×	✔	×
CSR-DCF	×	×	✔
STRCF	✔	×	×
Ours	✔	✔	✔

V. EXPERIMENTAL ANALYSIS

A myriad of experiment results on the OTB100 [22], VOT2016 [23], TC128 [24], and UAV123 [25] data sets reveal that the proposed Spatio-temporal regularized correlation filter with aberrance repression had significantly high tracking performance for the objects with OCC and SV attributes, meanwhile, its tracking ability to the objects with BC, IV, OCC, SV, LR, OV, and IPR attributes was also outstanding. According to the tracking performance of the proposed tracking model in this paper regarding the objects with the aforementioned attributes superior than that of BACF [10], STRCF [14] with the temporal regularized term and spatial regularized matrix, ARCF_HC [21] with the aberrance repression regularized term, and CSR-DCF [15] with the spatial reliability map, etc., a conclusion can easily be drawn that the fusion of the aberrance repression term of response map, spatial reliability map, and temporal regularized term of CFs was considerably beneficial in solving the boundary effects of correlation filters and the response map aberrance of the correlation filters, in addition to improving the object tracking robustness and accuracy of the proposed tracking model.

Meanwhile, from the very unexpected experiment results of the proposed tracking model on the VOT2016 data set [23], we can see that the proposed tracker was extremely effective in tracking the objects in color video sequences with high resolution and without many influence factors, and slightly poor for the objects in gray sequences or sequences with complex influence factors, such as MB, IPR, OPR, and OV, etc.. We believe that it was not only related to the construction of the proposed tracker, but also related to the hand-crafted gradient and CN-based color template features used in this paper. Among them, the HOG-based gradient template feature thinks that the appearance and shape of a local object can be well described by the distribution of the local gradient or edge direction, moreover, since the image is gamma-corrected and the gradient direction is quantified by cell method, it is not sensitive to the geometric or illumination variations; the CN-based color template feature is global, it focuses on the whole image, and has little dependence on the size, direction, and angle of the image itself, and it is not sensitive to the object deformation and fast movement.

However, during the experiment, we found that the proposed tracking model in this paper has a significant deficiency. As shown in Fig. 22(a), when the tracking object is fully occluded,

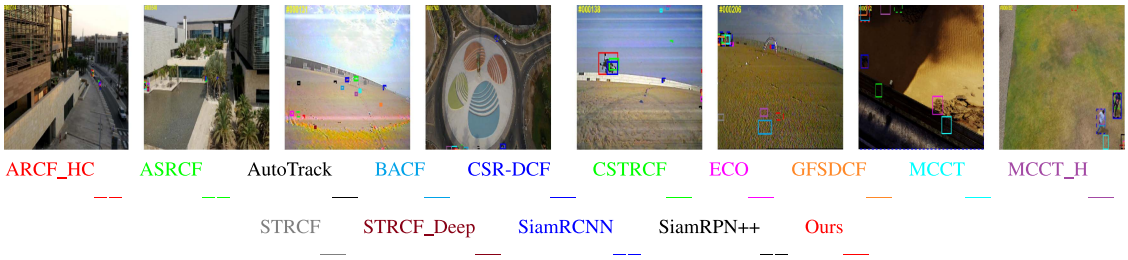

Fig. 21. Object tracking results in some image sequences of the UAV123.

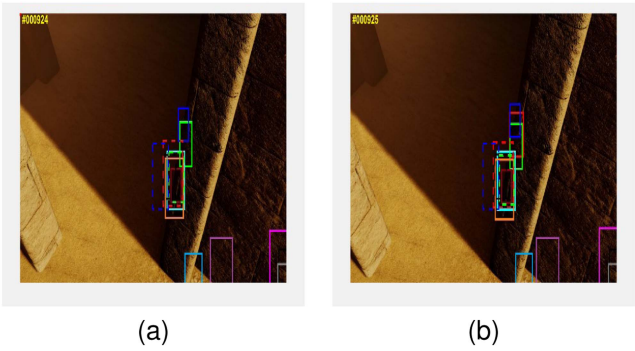

Fig. 22. ARSTCF tracking failed situation.

the proposed tracker fails to track. Although it can successfully track the object again when it appears slightly, as shown in Fig. 22(b), it is undeniable that this is a defect of the proposed tracker. Next, we need to think about how to solve the problem.

Further, from the experiment results on the UAV123 data set [25] presented in Fig. 17, an observation can be made that although the object tracking success rate of the proposed tracking model is higher than most of the trackers for the objects with almost all attributes, SiamRCNN [54] and SiamRPN++ [27] still performed better in tracking success rate and precision, which means that due attention must be paid to the important role of neural networks in the field of object tracking. In our future work, the knowledge of the neural networks is more necessarily to be applied into the research of the object tracking.

VI. Conclusion

In this paper, an object tracking model based on a time-varying Spatio-temporal regularized correlation filter with aberrance repression is proposed, in which the aberrance repression regularized term, spatial reliability map, and temporal regularized term are introduced into the objective function of the standard CF. By limiting the change rate of the response map in the object detection phase, the proposed tracker could obviously repress the aberrance of the response map, and improve the object tracking robustness and accuracy. Additionally, the correlation filter could be adjusted to the regions with high confidence scores to train the filter by using the spatial reliability map, which was conducive to overcoming the excessive object search range of the correlation filter and the limitations of the rectangle hypothesis in the tracking object, and the adverse effects caused by the boundary effect was solved well. The use of the correlation filter temporal regularized term was beneficial in enhancing the tracking ability of the proposed tracking model for the partially occluded object and the object with large appearance variations. A plethora of experiment results indicate that the tracking accuracy and robustness of the proposed tracking model are significantly higher compared with several state-of-the-art trackers.

References

[1] H. Nam and B. Han, "Learning multi-domain convolutional neural networks for visual tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 4293–4302.
[2] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. Torr, "Fully-convolutional siamese networks for object tracking," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 850–865.
[3] Q. Guo, W. Feng, C. Zhou, R. Huang, L. Wan, and S. Wang, "Learning dynamic siamese network for visual object tracking," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 1781–1789.
[4] L. Wang, W. Ouyang, X. Wang, and H. Lu, "Visual tracking with fully convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 3119–3127.
[5] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 702–715.
[6] L. Yang and J. Zhu, "A scale adaptive kernel correlation filter tracker with feature integration," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 254–265.
[7] 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.
[8] M. Kristan et al., "The visual object tracking VOT2015 challenge results," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop*, 2015, pp. 564–586.
[9] 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.*, 2010, pp. 2544–2550.
[10] H. K. Galoogahi, A. Fagg, and S. Lucey, "Learning background-aware correlation filters for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 1144–1152.
[11] M. Mueller, N. Smith, and B. Ghanem, "Context-aware correlation filter tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1387–1395.
[12] 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.
[13] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 4310–4318.
[14] 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.*, 2018, pp. 4904–4913.
[15] A. Lukežic, T. Vojír, L. C. Zajc, J. Matas, and M. Kristan, "Discriminative correlation filter with channel and spatial reliability," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4847–4856.

[16] T. Xu, Z. Feng, X. Wu, and J. Kittler, "Joint Group feature selection and discriminative filter learning for robust visual object tracking," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 7949–7959.

[17] N. Wang, W. Zhou, Q. Tian, R. Hong, M. Wang, and H. Li, "Multi-cue correlation filters for robust visual tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4844–4853.

[18] 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.*, 2019, pp. 4665–4674.

[19] M. Danelljan, G. Bhat, F. S. Khan, and M. Felsberg, "ECO: Efficient convolution operators for tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 6931–6939.

[20] Y. Li, C. Fu, F. Ding, Z. Huang, and G. Lu, "AutoTrack: Towards high-performance visual tracking for UAV with automatic spatio-temporal regularization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 11920–11929.

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

[22] 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.

[23] K. Matej et al., "The visual object tracking VOT2016 challenge results," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 777–823.

[24] 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.

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

[26] B. Wei, H. Chen, Q. Ding, and H. Luo, "SiamOAN: Siamese object-aware network for real-time target tracking," *Neurocomputing*, vol. 471, pp. 161–174, 2022.

[27] B. Li, W. Wu, Q. Wang, F. Zhang, J. Xing, and J. Yan, "SiamRPN : Evolution of siamese visual tracking with very deep networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 4277–4286.

[28] Z. Zhang, Y. Zhang, X. Cheng, and G. Lu, "Siamese network for object tracking with multi-granularity appearance representations," *Pattern Recognit*., vol. 118, 2021, Art. no. 108003.

[29] C. Ramakant, R. Rohit, M. Rohit, S. Upasana, K. Alok, and R. Hiral, "Enhanced the moving object detection and object tracking for traffic surveillance using RBF-FDLNN and CBF algorithm," *Expert Syst. Appl*., vol. 191, 2022, Art. no. 116306.

[30] L. Yu, B. Qiao, H. Zhang, J. Yu, and X. He, "LTST: Long-term segmentation tracker with memory attention network," *Image Vis. Comput*., vol. 119, 2022, Art. no. 104374.

[31] M. Zhu, H. Zhang, J. Zhang, and Li Zhuo, "Multi-level prediction siamese network for real-time UAV visual tracking," *Image Vis. Comput*., vol. 103, 2020, Art. no. 104002.

[32] H. Fu, W. Zhou, X. Wang, and H. Zhang, "Fast and robust visual tracking with hard balanced focal loss and guided domain adaption," *Image Vis. Comput*., vol. 100, 2020, Art. no. 103929.

[33] L. Fu, Y. Ding, Y. Du, B. Zhang, L. Wang, and D. Wang, "SiamMN: Siamese modulation network for visual object tracking," *Multimedia Tools Appl*., vol. 79, pp. 32623–32641, 2020.

[34] Y. Meng, Z. Deng, K. Zhao, Y. Xu, and H. Liu, "Hierarchical correlation siamese network for real-time object tracking," *Appl. Intell*., vol. 51, pp. 3202–3211, 2021.

[35] A. Luke, L. T. Voj, J. Matas, and M. Kristan, "FuCoLoT–A fully-correlational long-term tracker," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 8751–8760.

[36] F. Xie, C. Wang, G. Wang, Y. Cao, W. Yang, and W. Zeng, "Correlation-aware deep tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 8751–8760.

[37] Z. Cao, Z. Huang, L. Pan, S. Zhang, Z. Liu, and C. Fu, "TCTrack: Temporal contexts for aerial tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 14798–14808.

[38] C. Mayer et al., "Transforming model prediction for tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 8731–8740.

[39] J. Ye, C. Fu, G. Zheng, D. Paudel, and G. Chen, "Unsupervised domain adaptation for nighttime aerial tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 8896–8905.

[40] Q. Shen et al., "Unsupervised learning of accurate Siamese tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 8091–8100.

[41] F. Ma et al., "Unified transformer tracker for object tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 8771–8780.

[42] 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.

[43] A. Luke, L. E. Zajc, T. Voj, J. Matas, and M. Kristan, "FuCoLoT–A fully-correlational long-term tracker," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 595–611.

[44] P. Zhang, T. Zhuo, L. Xie, and Y. Zhang, "Deformable object tracking with spatiotemporal segmentation in big vision surveillance," *Neurocomputing*, vol. 204, no. 5 pp. 87–96, Sep. 2016.

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

[46] J. Wang, Z. Jia, H. Lai, J. Yang, and N. K. Kasabov, "A multi-information fusion correlation filters tracker," *IEEE Access*, vol. 8, pp. 162022–162040, 2020.

[47] D. Yuan, X. Zhang, J. Liu, and D. Li, "A multiple feature fused model for visual object tracking via correlation filters," *Multimedia Tools Appl*., vol. 78, pp. 27271–27290, 2019.

[48] J. Liao, C. Qi, and J. Cao, "Temporal constraint background-aware correlation filter with saliency map," *IEEE Trans. Multimedia*, vol. 23, pp. 3346–3361, 2021.

[49] J. P. Sun, E. J. Ding, B. Sun, Z. Y. Liu, and K. L. Zhang, "Adaptive kernel correlation filter tracking algorithm in complex scenes," *IEEE Access*, vol. 8, pp. 208179–208194, 2020.

[50] M. Guan and C. Wen, "Adaptive multi-feature reliability re-determinative correlation filter for visual tracking," *IEEE Trans. Multimedia*, vol. 23, pp. 3841–3852, 2021.

[51] Y. Zheng, X. Liu, X. Cheng, K. Zhang, Y. Wu, and S. Chen, "Multi-task deep dual correlation filters for visual tracking," *IEEE Trans. Image Process*., vol. 29, pp. 9614–9626, 2020.

[52] L. Pu, X. Feng, and Z. Hou, "Learning temporal regularized correlation filter tracker with spatial reliable constraint," *IEEE Access*, vol. 7, pp. 81441–81450, 2019.

[53] S. Li, S. Zhao, B. Cheng, E. Zhao, and J. Chen, "Lightweight particle filter for robust visual tracking," *IEEE Access*, vol. 6, pp. 32310–32320, 2018.

[54] P. Voigtlaender, J. Luiten, P. H. S. Torr, and B. Leibe, "SIAM R-CNN: Visual tracking by re-detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 6577–6587.