

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier XX.XX/ACCESS.XX.Doi Number

A Multi-Information Fusion Correlation Filters Tracker

JUNNAN WANG¹, ZHENHONG JIA¹, HUICHENG LAI², JIE YANG³, AND NIKOLA KASABOV⁴,
Fellow, IEEE

^{1,2} College of Information Science and Engineering, Key Laboratory of Signal Detection and Processing, Xinjiang University, Ürümqi 830046, China

³ Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200400, China

⁴ Knowledge Engineering and Discovery Research Institute, Auckland University of Technology, Auckland 1020, New Zealand

Corresponding author: ZHENHONG JIA (jzh@xju.edu.cn).

This work was supported in part by the National Science Foundation of China under Grant U1803261, in part by the International Science and Technology Cooperation Project of the Ministry of Education of the People's Republic of China under Grant DICE 2016–2196.

ABSTRACT In recent years, trackers based on correlation filters have attracted more and more attention due to the impressive tracking accuracy and real-time performance. However, in real scenarios, the tracking results are often been interfered with by the occlusion, illumination variation, appearance variation and background clutter. In order to find a tracker with better tracking performances, this paper proposed a multi-information fusion correlation filter tracker, which uses channel and spatial reliabilities and time regularization information on samples for filter training, and which not only extends the target search areas but also has a stronger ability to track the targets with significant appearance variations. Thus, results from extensive experiments conducted on OTB100, VOT2016, TC128, and UAV123 data sets show that our tracker with only directional gradient histogram (HOG) and color name (CN) features, performs favorably against the state-of-the-art trackers in terms of tracking precision, tracking success rate, tracking accuracy, and A-R rank.

INDEX TERMS Object tracking, correlation filter, channel reliability, spatial reliability, time regularization.

I. INTRODUCTION

The problem of target tracking has received a significant contributions in recent times due to the rapid developments of artificial intelligence technologies [1], [2], [3]. Target tracking refers to the continuous search of the target position and scale in subsequent video frames, given the target position and scale in the first frame. Currently, target tracking technology has made substantial progress as a result of improved computer hardware performances and the introduction of new target tracking algorithms. However, as numerous adverse factors occur in real scenarios, such as target occlusion, scale variation, illumination variation, background variation, and appearance variation, etc., it is still a major challenge for a tracker to achieve high-precision, high success rate, and reasonable robustness.

Correlation filter trackers train classifiers by minimizing errors. Thus, by extracting the target information and correlating with correlation filters, a group of target-possible response values is obtained, and the position with the highest response value is taken as the center of the target. Furthermore, in order to release the computational

burden, the Fast Fourier transform is often used to transform the loss function of the tracker into the frequency domain.

Various studies have argued that, in recent times, correlation filter-based trackers are widely used in the field of tracking because correlation filter-based trackers have more efficient computational capacity and more robustness than other trackers [3],[4],[5]. Bolme et al. [6] proposed a minimum output sum of squared error filter (MOSSE) tracker, which trains classifiers with gray-scale features of the target in the initial frame, and the correlation filters are used for target tracking for the first time. Henriques et al. [7] on their part added dense sampling and kernel trick based on MOSSE, and the dense sampling reduced the redundancy of training samples by shifting the image vector with a cyclic matrix. Furthermore, based on circulant structure tracking with kernels (CSK), the directional gradient histogram (HOG) feature, which is more robust to the change of illumination was introduced to kernelized correlation filter (KCF) [8], with the aim of achieving a better tracking performance. Unlike the Gauss kernel function used by KCF, with the help of linear kernel that has advantages in

multi-channel feature merging, discriminant correlation filters (DCF) gain faster-tracking speed.

At present, many scholars are still studying and improving DCF-based tracking models. In their study, Danneljan et al. [9] proposed spatial regularization DCF (SRDCF) algorithm by imposing a spatial penalty on DCF coefficients to resolve the boundary effect caused by cyclically shifted samples. Li et al. [10] also suggested spatiotemporal regularized correlation filter (STRCF) by introducing time regularization into SRDCF that deals with the boundary effect without lost of efficiency, and at the same time, having stronger ability to deal with the targets with large appearance variations and occlusion.

In introducing channel and spatial reliabilities into DCF, Lukežić et al. [11] were of the view that a discriminative correlation filter with channel and spatial reliabilities (CSR-DCF), effectively usually track the irregular shaped targets. Additionally, in an adaptive spatial regularization correlation filter proposed by Dai et al. [12], two correlation filter (CF) models with complex features are used to locate the target and whereas the other filter correlations with shallow features are used to estimate the target scale. Consequently, to solve the drawbacks of the DCFs-based trackers that some negative effects are produced by the generated samples and the response map is vulnerable to noise interference, Yuan et al. [51] proposed a target-focusing convolutional regression model for visual object tracking. Therefore, to enhance the robustness of deep regression trackers to complicated situations, (e.g., occlusion, background clutter, and deformation), Yuan et al. [52] further proposed an adaptive structural convolutional filter model. Considering that the appearance model is easily disturbed by noise in the tracking algorithms with a single feature, Yuan et al. [53] proposed a multiple feature fused model into a correlation filter framework for object tracking.

Although many improvements have been made to the correlation filters and good target tracking results have been achieved, the currently available correlation filter trackers still fail to completely solve the boundary effect caused by the cyclically shifted samples used training correlation filters.

To solve the above-mentioned problems, this paper suggested a multi-information fusion correlation filters tracker to boot the tracking performance and robustness of the tracker, in which the channel and spatial reliabilities and time regularization information of samples are used together to train correlation filters for the first time, whereas only one or two of them are used in previous correlation filter trackers. Furthermore, in the proposed tracker, the spatial reliability is used to adjust the filter to the areas suitable for target tracking, which is effective in overcoming the limitation of the tracking target rectangle; the channel reliability is used to weight the response of each feature channel to emphasize its contribution in target location to better located the target; and the time regularization is helpful at some extent in dealing with the boundary effects and

improving the tracking robustness to the target with large appearance variations and much occlusion. At the same time, the alternating direction multiplier method is used in this paper to solve our object function to improve the time performance of our tracker. Extensive experiments conducted on four universal data sets OTB100 [13], VOT2016 [14], TC128 [54], and UAV123 [2] with multiple attributes video sequences suggest that our tracker performs favorably against many state-of-the-art trackers in terms of precision rate, success rate, tracking accuracy, A-R rank, pixel error, and overlap rate.

The main contributions of this paper are listed as follows:

1) By combining the time regularization with the channel and spatial reliabilities of samples for the first time, a multi-information fusion correlation filters tracker was proposed in this study. Thus, the use of the spatial reliability allows searching the target in larger areas, and the channel reliability is helpful in better locating the target; meanwhile, the time regularization can alleviate the influence of boundary effects caused by cyclically shifted samples and improve the tracking robustness to the target with large appearance variations and much occlusion.

2) The alternating direction multiplier method (ADMM) [22] was used to solve the filter and the Lagrange operator iteratively after the augmented Lagrange equation of our tracker being decomposed into sub equations related to the filters, and it significantly reduces the computational complexity of our tracker.

The remainder of this paper is structured as follows: Section II reviewed literature in relation to object tracking; the contents of the proposed tracker are then presented in section III; in Section IV, we evaluated the proposed tracker on publicly available data sets and finally, section IV concludes the study.

II. RELATED WORK

This section reviewed literature on object tracking with specific focus on correlation filter-based tracker and non-correlation filter tracker.

A. CORRELATION FILTER-BASED TRACKER

In recent years, the CF-based tracker has widely been used in the field of visual tracking. Considering that the tracking will be adversely affected by the target surrounding environment, in the learning stage, Mueller et al. [16] added context information to the CF for the first time, which significantly improved the tracking performance. In view of these, the CF filters are controlled by the salient regions on the feature map, which leads to model degradation, Sun et al. [17] introduced a local response consistency regular term to emphasize the equal contribution of different regions.

In a related study to solve the problem of target tracking drift and even failure caused by background clutters or target appearance variations, Li et al. [18] proposed train correlation filters with background patches selected by affinity propagation to maximize the edge between foreground and background, while at the same time using a

multi-level target scale variation supervision mechanism to adjust the target scale. The interdependence between the different features of the tracking target and the spatial constraints between each part was used by Zhang et al. [19] to examine multiple correlation filters to give full play to the advantages of correlation and particle filters to effectively track targets with scale variations and occlusion. Liu et al. [20] on the other hand proposed a template matching via mutual buddy similarity and memory filtering tracker to match targets with reciprocal k-nearest neighbors in complex situations, and the representative and reliable results are selected to learn different types of templates in memory filtering scheme.

To improve the feature extraction, sample training process, and tracking performance of the traditional kernel correlation filter, Yang et al. [21] proposed a joint correlation filter tracker with multi-feature and scale adaptation, which consists of two parts: position and scale correlation filter trackers. Zhang et al. [22] proposed a scale adaptive tight correlation filter to solve the problems that the background interference caused by large scale training samples and target representation errors caused by the appearance variations of training samples.

To solve the problem that the part-based tracker has a poor tracking performance on partially occluded targets, which is caused by the target overall appearance ignorance, Ruan et al. [23] integrated a part-based strategy into the CF framework and proposed a multi-part correlation tracker with triangle structure constraints (MCTTC) by constructing multiple CFs with the global and local appearance of the target. Wang et al. [24] proposed an effective framework for multi-threaded analysis, in which multiple experts were constructed with discriminant correlation filters, and the most appropriate expert was selected to track the target.

Recently, it has become the choice of many researchers to combine CF models with deep features for target tracking. The convolutional neural network (CNN), CN, and HOG features of the target were used by Danelljan et al. [25] to generate the feature maps to train CF models. A CF model with single-scale robust deep features was used by Dai et al. [12] to locate the target accurately. Danelljan et al. [26] used continuous convolution filters to combine feature maps with different spatial resolutions. Li et al. [10] combined the output of the conv3 layer in the VGG-M network with the HOG and CN features to train correlation filters. Sun et al. [17] used the deep features which were the outputs of conv1 from VGG-M and conv4-3 of VGG-16, HOG and CN features to train correlation filters [27].

He et al. [28] combined autocorrelation and cross-correlation with convolution neural networks to represent the target features to jointly exert the advantages of CF and CNN, and obtained excellent tracking performance. To enhance the recognition and tracking abilities of correlation filters to the occluded and deformed targets, Pu et al. [29] constructed a spatial reliability map with deep features by using convolution neural networks and introduced time regularization to train DCFs.

B. NON-CORRELATION FILTER TRACKER

Although several studies suggest the preference for CF-based trackers, some non-correlation filter trackers have also shown good tracking efficiency. In their study, Li et al. [30] proposed a multi-stream deep similarity learning network to learn a strictly offline similarity comparison model, which could still effectively identify the target even if it is interfered with background clutters and appearance variations. Bhat et al. [31] on the other hand proposed a particle filter target tracking algorithm based on multi-feature fusions, in which the color distribution in the particle filter framework is robust to the target with scale variation and partial occlusion and the KAZE (a Japanese word that means wind) feature of the target structure are used to track the target.

In a related study to design a target tracking model with effective online observation and model updating capabilities, Huang et al. [32] proposed representing the target features with the combination of direction gradient change and color histograms, while the single hidden layer feed-forward neural network and recursive orthogonal least-squares algorithm are used as target observation models. Aiming at the high time complexity of the Siamese trackers when used to estimate the scale and angle of the tracking target, Lee [33] proposed a single shot Siamese network that could estimate the size and angle of the target with a single search area. Li et al. [34] proposed a lightweight particle filter tracking method that not only retains the robust tracking ability of particle filters, but also reduces the time cost in sampling with the use of correlation filters.

Inspired by anchor free detectors, Chen et al. [35] opined that Siamese box adaptive target tracking network, composed of Siamese network backbone and multiple boxes adaptive heads, could be appropriately used to resolve the problem of accurate estimation to the target scale and aspect ratio by transforming tracking into classification regression. Danelljan et al. [36] in their contribution argued that probability regression formula for target tracking can model the label noise caused by incorrect annotations and ambiguities, thus leading to an improved the tracking performance. Also, Xu et al. [37] in analyzing the unique characteristics of the target tracking problem, suggested a set of practical target state estimation criteria by designing a full convolution Siamese ++ tracker (SiamFC++) consisting of classification and target state estimation branch (G1), no fuzzy classification score (G2), no prior knowledge tracking (G3) and estimation quality score (G4).

Considering the situation where full convolution Siamese network based on template matching cannot capture the time variation information of the target and background clutter, Li et al. [38] proposed a gradient guidance network to update the template of the current frame with the discrimination information of the gradient. In addition, Li et al. [39] put forward a tracking algorithm, in which two complementary trackers run in parallel, and between these two trackers, the Bayesian tracker (B-tracker) with adaptive learning rate solves the problem of target appearance variations; the S-tracker, which is the tracker with an improved incremental

subspace learning method, solves the problem of target occlusion.

To deal with the problem of illumination variation and occlusion in visual tracking, Li et al. [40] suggested to use only the bright pixels to compare the similarity between candidate and training samples, and to update the model with an online strategy after getting a new target. Moreover, Li et al. [41] in resolving the problem of performance decrease of a single classifier tracking whenever the target is occluded, suggested that a group of related classifiers should first be derived with the combination of particle filters and sample sets, and then a classifier query mechanism should be established to select the appropriate classifier to track the target in the next frame.

III. OUR PROPOSED TRACKER

In this section, we discuss the following in turn: correlation filters (CF) [42], spatially regularized discriminative correlation filters (SRDCF) [9], spatial-temporal regularized correlation filters (STRCF) [10] and discriminative correlation filter with channel and spatial reliability (CSRDCF) [11]. Finally, the correlation filter tracker with channel and spatial reliabilities and time regularization proposed in this paper is discussed.

A. CORRELATION FILTERS

Given a set of N_d channel features $f=\{f_d\}_{d=1:N_d}$, and corresponding target templates (filters) $h=\{h_d\}_{d=1:N_d}$, where $f_d \in \mathbb{R}^{dw \times dh}$, $h_d \in \mathbb{R}^{dw \times dh}$. By minimizing the sum of the square differences between the channel correlation output and the expected output (ground truth) $g \in \mathbb{R}^{dw \times dh}$, the optimal filter is obtained in the learning stage.

$$\arg \min_h \sum_{d=1}^{N_d} \|f_d * h_d - g\|^2 + \lambda \sum_{d=1}^{N_d} \|h_d\|^2, \quad (1)$$

$$= \arg \min_h \sum_{d=1}^{N_d} (\|h_d^H \text{diag}(\hat{f}_d) - \hat{g}_d\|^2 + \lambda \|\hat{h}_d\|^2)$$

where the operator $\hat{a} = \text{VEC}(f[a])$ is the Fourier transform of the column vector $a \in \mathbb{R}^{D \times 1}$, $D = dw \times dh$, $\text{diag}(a)$ is a $D \times D$ diagonal matrix composed of a , $(\cdot)^H$ is a Hermitian transpose, $*$ represents the convolution operator, and λ is a regularization constant.

As the CF model suffers from the unwanted influences of boundary effects caused by circulant shifted samples, which are used for filter learning, its tracking performance is unavoidably degraded.

B. SPATIALLY REGULARIZED DISCRIMINATIVE CORRELATION FILTERS

To mitigate the undesirable boundary effects in the CF model, M. Danelljan et al. [9] proposed spatially regularized discriminative correlation filters (SRDCF) with spatial constraints. In SRDCF, a larger size of the image channel feature f_d is taken to retain more real information of the target, and then punishes the samples far from the target center

through a spatial weight coefficient w . SRDCF is developed by minimizing the following indicators:

$$\arg \min_h \sum_{d=1}^{N_d} \|f_d * h_d - g\|^2 + \lambda \sum_{d=1}^{N_d} \|w * h_d\|^2, \quad (2)$$

where “ \cdot ” denotes the Hadamard product, $*$ stands for the convolution operator, w is the spatial regularization matrix, f_d is the channel feature, and h_d and g are the target template and desired output, respectively.

Although SRDCF [9] can effectively suppress the adverse boundary effects, the spatial regularization on multiple images will destroy the structures of the circulant matrix, resulting in a higher computational burden.

C. SPATIAL-TEMPORAL REGULARIZED CORRELATION FILTERS

By introducing temporal regularization to SRDCF [9], F. Li et al. [10] simplified the SRDCF reduced multiple samples with the spatial-temporal regularized correlation filters (STRCF) into a single sample, and the problem of the large amounts of calculation in the SRDCF [9] model was solved.

$$\arg \min_h \frac{1}{2} \sum_{d=1}^D \|f_d * h_d - g\|^2 + \frac{1}{2} \sum_{d=1}^D \|w * h_d\|^2 + \frac{\mu}{2} \|h - h_{t-1}\|^2, \quad (3)$$

where f_{t-1} denotes the CFs utilized in the $(t-1)$ -th frame, and μ denotes the regularization parameter. Meanwhile, the second term in formula (3) denotes the spatial regularization, and the third term denotes the temporal regularization.

STRCF [10] can adaptively balance the trade-off between aggressive and passive model learning, and has more robust tracking performance in the case of large variations in the appearance of the tracking target.

D. DISCRIMINATIVE CORRELATION FILTER WITH CHANNEL AND SPATIAL RELIABILITY

To alleviate the unwanted boundary effects in the CF model, A. Lukežić et al. [11] introduced a dual variable h_c to the CF model and constrained $h_c - m \odot h = 0$; here, m is the spatial reliability map, which identifies pixels in the filter that should be ignored in learning. The augmented Lagrangian form of CSR-DCF is

$$L(\hat{h}_c, h, \hat{I} | m) = \|\hat{h}_c^H \text{diag}(\hat{f}) - \hat{g}\|^2 + \frac{\lambda}{2} \|h_m\|^2 + [\hat{I}^H (\hat{h}_c - \hat{h}_m) + \overline{\hat{I}^H (\hat{h}_c - \hat{h}_m)}] + \mu \|\hat{h}_c - \hat{h}_m\|^2, \quad (4)$$

where \hat{I} is a complex Lagrange multiplier, $\mu > 0$, and for compact notation, $h_m = m \odot h$ is defined.

At the target positioning stage, the channel reliability is computed as the product of the learning channel reliability $\omega_d = \zeta \max(f_d \times h_d)$ and detection channel reliability $\omega_d^{(det)} = 1 - \min(\rho_{max2}/\rho_{max1}, 1/2)$, where ρ_{max2}/ρ_{max1} is the ratio between the second and the first major modes in the response map.

E. OUR PROPOSED TRACKER

In order to solve the boundary effects caused by cyclically shifted samples used for correlation filters training as much as possible, and obtain better tracking performance for the targets with appearance variation and occlusion, in this paper, we proposed a multi-information fusion correlation filter tracker, in which the channel and spatial reliabilities and time regularization information of samples are used for correlation filter training, and the channel and spatial reliabilities refer to the corresponding contents of CSR-DCF [11].

$m \in [0, 1]^{dw \times dh}$ is the spatial reliability map with elements $m \in [0, 1]$ which indicate the learning reliability of each pixel. In CSR-DCF [11], from the perspective of probability model, Lukežić et al. [11] suggest that the reliable probability of pixel x conditioned on appearance y is

$$p(m=1|y, x) \propto p(y|m=1, x) p(x|m=1) p(m=1), \quad (5)$$

where the first term on the right is the appearance likelihood value, which is computed by the target foreground and background color histograms; the second term is the probability of the high reliability area where the object located, whose value is determined by the distance between the pixel x and the object center; the third term can be regarded as a prior probability, which is determined by the sizes of the extracted foreground and background models. In a Markov random field, the spatial consistency of labeling m is achieved by using (5) as unary terms.

In multi-channel correlation filters, considering the different importance of each channel to filter training, Lukežić et al. [11] suggest that it is necessary to weight the filter h of each channel, and the weight is determined by the product of learning and detection reliabilities. The learning reliability of each channel is determined by the product of a discriminative feature channel f_d and a filter h_d , i.e., $w^d = \zeta \max(f_d * h_d)$; the detection reliability is determined by the ratio between the second and the first major modes in response map, i.e., $w^{(det)} = 1 - \min(\rho_{max2}/\rho_{max1}, 1/2)$; finally, the weight of each channel is the normalized product of $w^d * w^{(det)}$.

The augmented Lagrangian function of our object function is

$$L(h_c, h, \hat{I} | m) = \|h_c^H \text{diag}(\hat{f}) - \hat{g}\|^2 + \frac{\lambda}{2} \|h_m\|^2 + [\hat{I}^H (h_c - h_m) + \hat{I}^H (\hat{h}_c - \hat{h}_m)] + \mu \|h_c - h_m\|^2 + \frac{\gamma}{2} \|h_m - h_m^{t-1}\|^2, \quad (6)$$

where λ and γ are regularization parameters, μ is the constraint penalty factor, and h_c is a dual variable with constraint $h_c - m \odot h = 0$.

Let $h_m = m \odot h$ and $\hat{h}_m = \sqrt{DFM}h$, then, equation (6) can be rewritten as

$$L(\hat{h}_c, h, \hat{I} | m) = \|\hat{h}_c^H \text{diag}(\hat{f}) - \hat{g}\|^2 + \frac{\lambda}{2} \|meh\|^2 + [\hat{I}^H (\hat{h}_c - \sqrt{DFM}h) + \hat{I}^H (\hat{h}_c - \sqrt{DFM}h)] + \mu \|\hat{h}_c - \sqrt{DFM}h\|^2 + \frac{\gamma}{2} \|\sqrt{DFM}h - \sqrt{DFM}h_{t-1}\|^2, \quad (7)$$

where F is an orthogonal matrix composed of Fourier coefficients.

Equation (6) can be iteratively minimized by the alternating direction multiplier method (ADMM) [15]. In each iteration, the following sub-problems are solved:

$$h^{i+1} = \arg \min_h L(\hat{h}_c^{i+1}, h, \hat{I}^i | m), \quad (8)$$

$$\hat{h}_c^{i+1} = \arg \min_{\hat{h}_c} L(\hat{h}_c, h^i, \hat{I}^i | m), \quad (9)$$

The closed-form solutions of equations (8) and (9) can be obtained from the partial derivative values of L to \hat{h}_c and \hat{h} equal to 0 in equation (7), respectively.

In equation (7), We set the first, second, third, fourth and fifth items on the right of the equal sign as L_1 , L_2 , L_3 , L_4 and L_5 , respectively. After many derivations, we got the following equation:

$$\frac{\partial L}{\partial h} = \frac{\partial L_1}{\partial h} + \frac{\partial L_2}{\partial h} + \frac{\partial L_3}{\partial h} + \frac{\partial L_4}{\partial h} + \frac{\partial L_5}{\partial h} = \frac{\lambda}{2} Mh - \sqrt{DMF}^H \hat{I} - \mu \sqrt{DMF}^H \hat{h}_c + \mu DMh + \frac{\gamma}{2} DMh - \frac{\gamma}{2} DMh_{t-1} = 0, \quad (10)$$

From equation (10), the closed solution of the equation (8) is got, i.e.,

$$h^{i+1} = m \odot \left[\mathcal{F}^{-1}(\hat{I}^i + \mu \hat{h}_c^{i+1}) + \frac{\gamma}{2} h_i \right] \odot^{-1} \left(\frac{\lambda}{2D} + \mu^i + \frac{\gamma}{2} \right), \quad (11)$$

The partial derivative values of L to \hat{h}_c is shown as the following equation:

$$\frac{\partial L}{\partial \hat{h}_c} = \frac{\partial L_1}{\partial \hat{h}_c} + \frac{\partial L_2}{\partial \hat{h}_c} + \frac{\partial L_3}{\partial \hat{h}_c} + \frac{\partial L_4}{\partial \hat{h}_c} + \frac{\partial L_5}{\partial \hat{h}_c} = \text{diag}(\hat{f})^H, \quad (12)$$

$$\hat{h}_c \text{diag}(\hat{f}) - \hat{g}^H \text{diag}(\hat{f}) + \hat{I} + \mu \hat{h}_c - \mu \sqrt{DFM}h = 0$$

From equation (12), the closed solution of the equation (9) is got, i.e.,

$$\hat{h}_c^{i+1} = \left[\hat{g} \odot \hat{f} + (\mu \hat{h}_m^i - \hat{I}^i) \right] \odot^{-1} (\hat{f} \odot \hat{f} + \mu^i), \quad (13)$$

The Lagrange multiplier \hat{I} and constraint penalty μ are updated as the following equations (14) and (15), respectively:

$$\hat{I}^{i+1} = \hat{I}^i + \mu (\hat{h}_c^{i+1} - h^{i+1}), \quad (14)$$

$$\mu^{i+1} = \beta \mu^i, \quad (15)$$

In Algorithm 1, we give a brief overview of our proposed tracking framework.

Algorithm 1	The Proposed Tracking Algorithm
Input:	
Frame i ; Object ground-truth g , start_frame=1.	
Output:	
Object bounding-box.	

```

1. Input frame  $i$  and object ground-truth  $g_i$ .
2. If  $i = \text{start\_frame}$ 
3.   Extract foreground and background histograms of the object
   ground-truth area in the current frame.
4.   Calculate foreground prior with foreground and background
   histograms.
5.   Calculate spatial reliability map  $\mathbf{m}$  with foreground prior.
6.   Extract HOG and CN features  $f_i$  of object.
7.   Calculate filter  $h_i$  and dual variable  $h^{(\wedge)}c$  by Eqn. 11 and Eqn.
13.
8.   Update Lagrange multiplier  $\hat{l}$  by Eqn. 14. and constraint penalty
 $\mu$  by Eqn. 15.
9.   Calculate response with features  $f_i$  and filter  $h_i$ .
10.  Calculate per-channel learning reliability  $w^d$  with response.
11.  Construct tracker and the object ground-truth  $g_i$  is output as the
   object bounding-box.
12. else
13.  Extract HOG and CN features  $f_i$  of the object in previous frame
   with previous tracker.
14.  Calculate response with features  $f_i$  and previous filter  $h_{i-1}$ .
15.  Find the position of the maximum response  $rc_{max}$ .
16.  Calculate displacement distance of the object center with  $rc_{max}$ 
   and get the new center of the current frame.
17.  Calculate the bounding-box of the object with the new center
   and previous frame tracker.
18.  Extract foreground histogram and background histogram of the
   bounding-box area.
19.  Calculate foreground prior.
20.  Calculate spatial reliability map  $\mathbf{m}$  with foreground prior.
21.  Extract HOG and CN features  $f_i$  of the bounding-box area of the
   current frame.
22.  Calculate new filter  $h_i$  and dual variable  $h^{(\wedge)}c$  by Eqn. 11 and
   Eqn. 13. with  $\mathbf{m}$  and  $f_i$ .
23.  Update Lagrange multiplier  $\hat{l}$  by Eqn. 14. and constraint penalty
 $\mu$  by Eqn. 15.
24.  Calculate per-channel learning reliability  $w^d$  and detection
   reliability  $w_a^{(det)}$ , then get per-channel weight  $w$  with the dot product
   of  $w^d$  and  $w_a^{(det)}$ .
25.  Construct a new tracker.
26.  Calculate the intersection area  $S$  of the object bounding-box
   and its ground-truth.
27.  if  $S < \text{eps} \ \&\& \ \text{true}$ 
28.    frame = frame + 4;
29.    start_frame = frame + 1;
30.  end
31. end

```

IV. EXPERIMENTS

All our experiments are implemented on OTB100 [13], VOT2016 [14], TC128 [54], and UAV123 [2] data sets in MATLAB R2018a on a PC with 3.6 GHZ Intel Core i7 processor and 8 GB RAM. Data was analyzed with precision plots of one-pass evaluation (OPE), success plots of OPE, accuracy rate, A-R (accuracy and robust) rank, overlap rate, and pixel error to compare with some state-of-the-art trackers.

In conducting the experiment, we set the reliability map estimation parameter at $\alpha_{min}=0.05$, the histogram adaptation rate at $\eta_c=0.04$, the correlation filter adaptation rate at $\eta=0.02$, the regularization parameter at $\lambda=0.01$, the step-size parameter at $\gamma=-1$, and the augmented Lagrangian optimization parameters at $\mu=5$ and $\beta=3$. These parameters remain constant in the experiments. More detailed parameters setting information can be found in the code of CSR-DCF algorithm which is the main foundation of our tracker's code.

A. COMPARISON OF TRACKERS

The study compared the proposed tracker with some state-of-the-art trackers, with the aim of fully demonstrating the tracking performance of the proposed tracker. Specifically, the proposed tracker has been compared with the following trackers.

AutoTrack [43] in which the spatially local response map variation was introduced as spatial regularization to make DCF mainly learn the reliable part of the object; ARCF [44] in which background patches are added as negative training samples to expand the target search areas and resolve the boundary effects, where the ARCF_H [44] is the tracker with HOG feature, and the HOG, CN and greyscale features are used in ARCF_HC [44]; HOG feature based BACF [45] in which the negative samples generated by foreground real shifts are obtained through zero padding operation to include larger search areas and many real backgrounds; C-COT [26] in which the deep neural network VGG-net is used to extract the target features, and the feature maps of different resolutions are interpolated into the continuous space domain through an implicit interpolation mode.

CF2 [46] in which HOG feature in the KCF is replaced with deep convolutional features extracted by conv3-4, conv4-4 and conv5-4 layers in VGG-Net; CSR-DCF [11] in which channel and spatial reliabilities are introduced and the standard HOG and CN features are used to train correlation filters; color statistical features based DAT_USABLE [47], in which distractor-aware tracking (DAT) calculates the color histograms of the foreground and background to obtain their color probability models; the CNN, HOG and CN features based ECO [25], which improved the C-COT [26] by reducing the parameters of the DCF.

GFSDCF [48] in which a correlation filter tracking method for joint group feature selection across both channel and spatial dimensions is proposed, and CN, HOG, intensity channels (IC) and CNN features are used; MCCT [24] in which multiple independent DCF-based experts are used to track the target, and each of them is constructed with different combinations of deep and HOG features, and experts in MCCT_H are constructed with different combinations of CN and HOG features; SCT4 [49] in which the decomposition and integration of attention modulations are used to track the target; STAPLE [50], in which two complementary features HOG and COLOR are used to learn the target; STRCF [10] in which HOG and CN features are used, and it can carry out DCF model learning and updating simultaneously, wherein the STRCF_Deep [10] is the STRCF with CNN features.

B. THE OTB100 DATA SET

The OTB100 data set [13] also known as OTB2015 [12], contains various types of tracked targets, and includes 100 fully annotated video sequences with 11 different attributes, such as 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). Accuracy rate and success rate based on precision plot and success plot are the commonly used evaluation indicators for the OTB data set. The precision plot is used to show the percentage of frames, whose tracking bounding box center positioning error is better than the given threshold, to the total number of frames; whereas the success plot is used to show the percentage of frames, whose tracking bounding box overlap rate is greater than the given threshold, to the total number of frames. The center positioning error on the other hand is the average Euclidean distance between the

center of the predicted target bounding box and the center of the artificially labeled ground truth. While the bounding box overlap rate refers to the ratio of the intersection and union between the target bounding box estimated by tracking algorithm and the ground truth. Thus, the study evaluated the tracking performances of all trackers in this paper on OTB100 with precision plots of OPE, success plots of OPE, overlap rate, and pixel error. The results of the experiment are shown below:

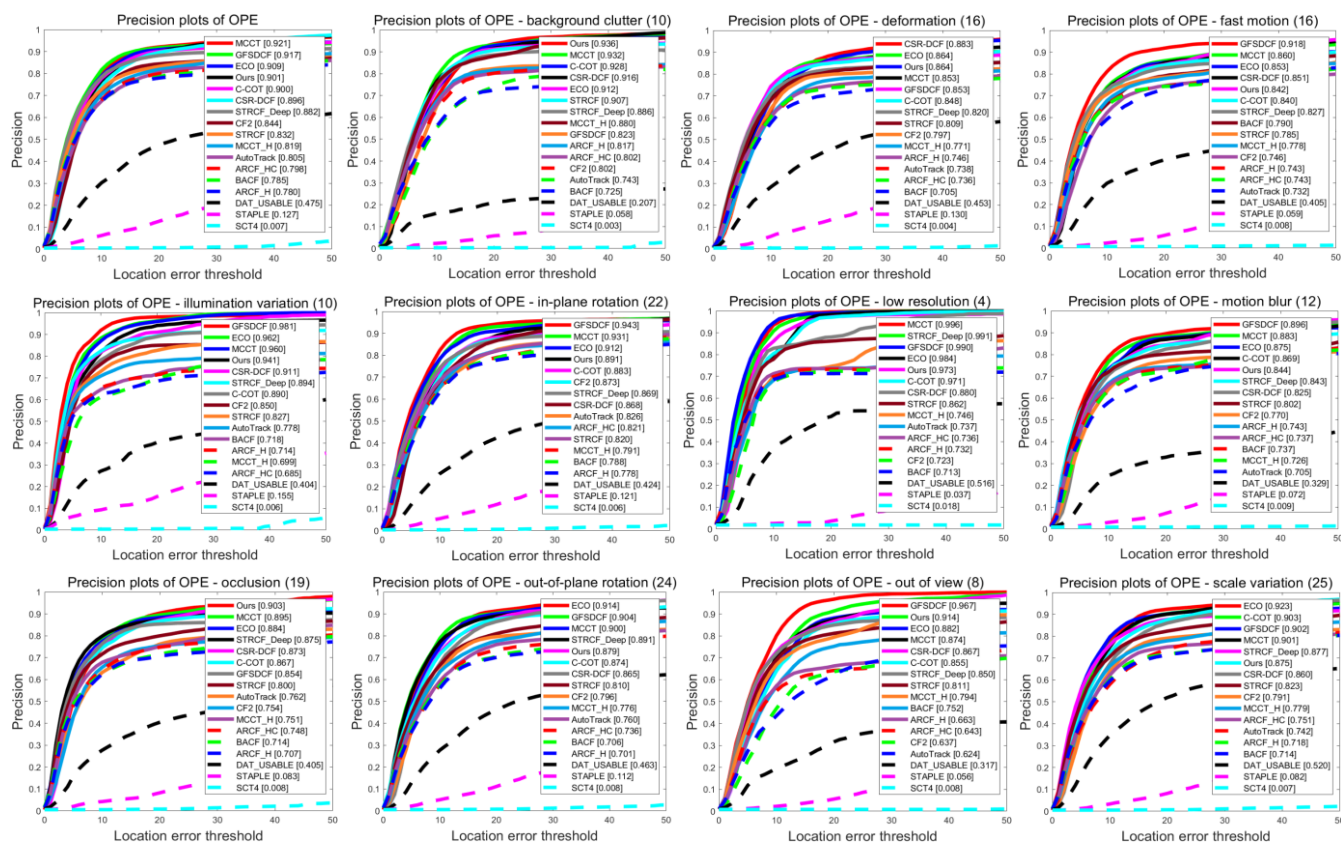
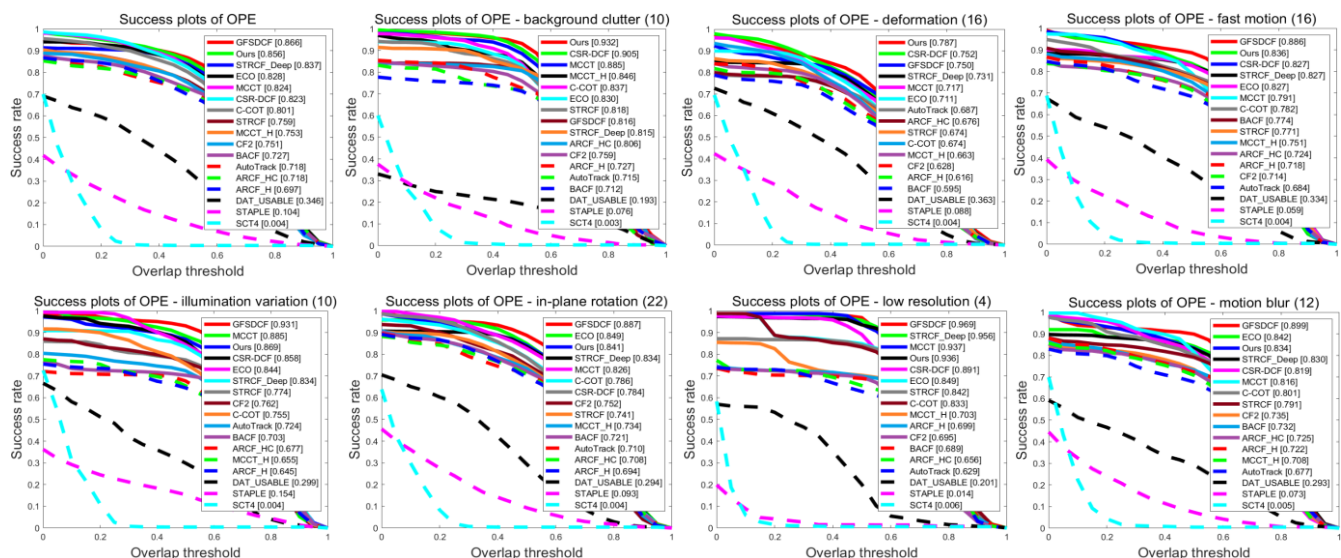


FIGURE 1. Trackers' Precision plots on OTB100



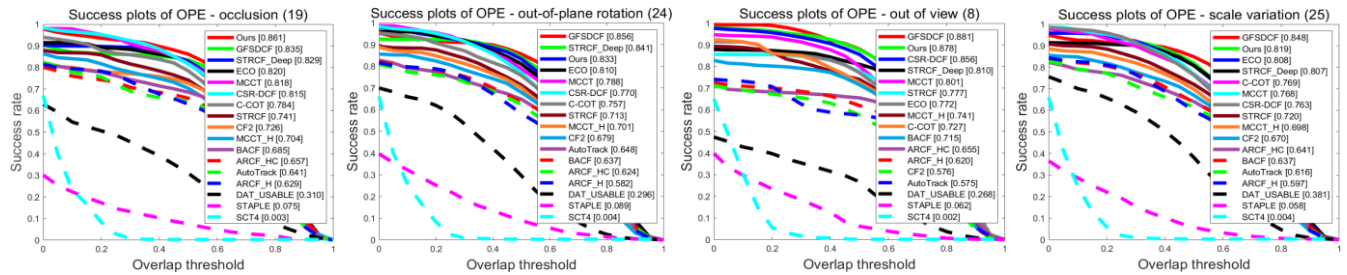


FIGURE 2. Trackers' Success plots on OTB100

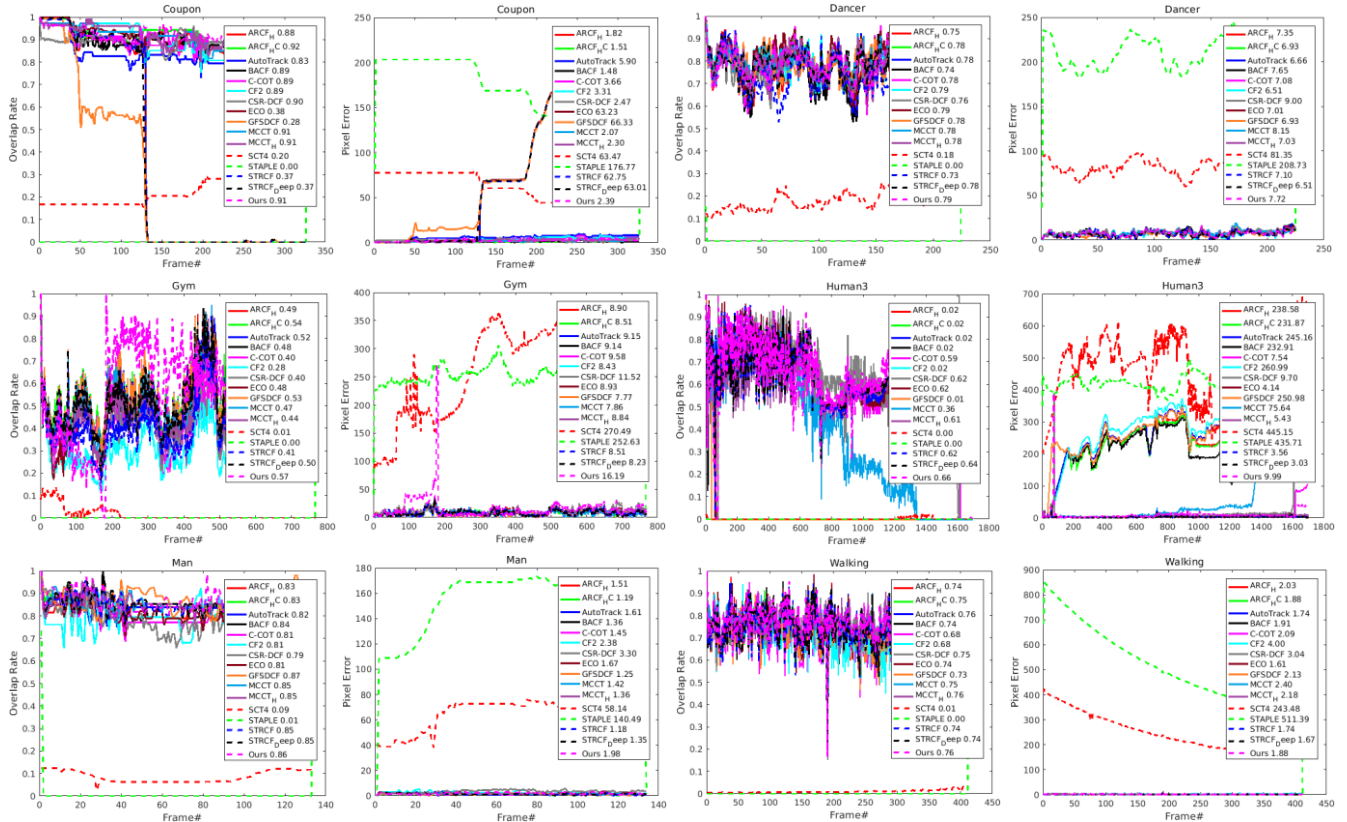
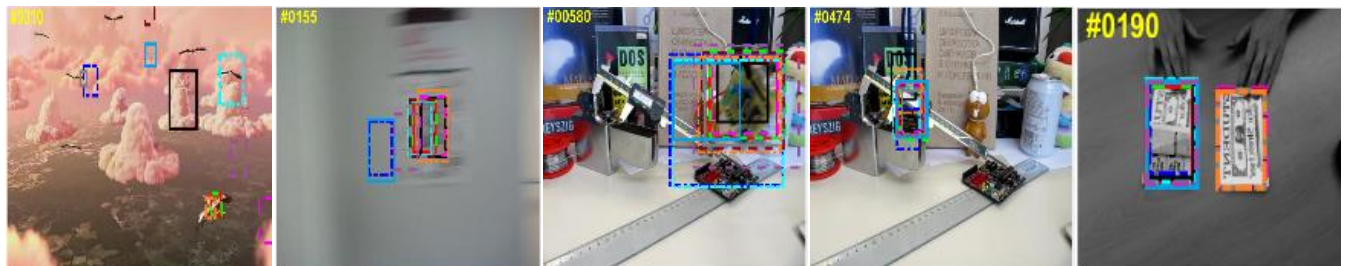


FIGURE 3. Trackers' overlap rates and pixel errors on some OTB100 sequences

TABLE 1.
THE AVERAGE OVERLAP RATE AND PIXEL ERROR VALUES OF SOME TRACKING ALGORITHMS ON OTB100
(THE RED, BLUE AND GREEN NUMBERS DENOTE THE 1ST, 2ND AND 3RD RANK)

Algorithm	AutoTrack	BACF	C-COT	CSR-DCF	ECO	GFSDCF	MCCT	MCCT_H	STRCF	STRCF_Deep	Ours
Overlap Rate(%)	0.63	0.64	0.70	0.66	0.70	0.71	0.69	0.65	0.67	0.70	0.67
Pixel Error(pixel)	26.6	25.9	9.7	12.9	13.7	11.2	10.4	18.3	19.9	14.8	13.0



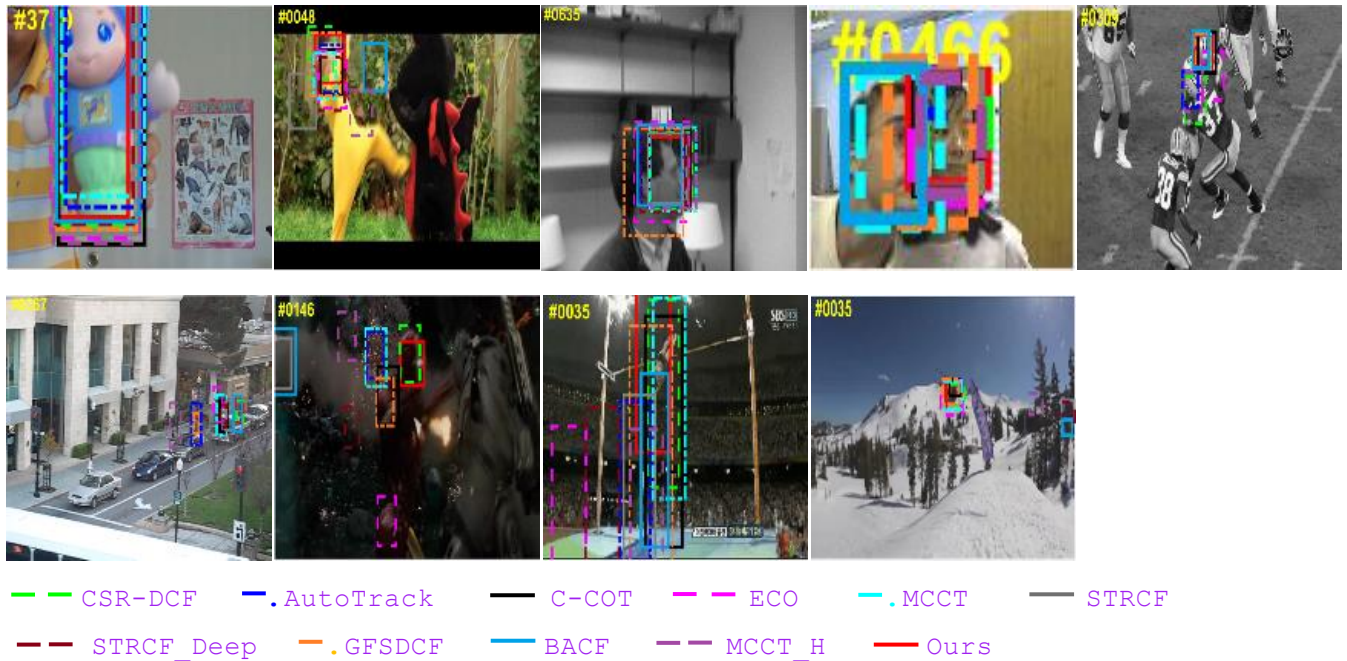


FIGURE 4. Tracking results on some challenging image sequences on OTB100

Figure 1 shows the precision plots of each tracker for video sequences with different attributes on the OTB100 data set [13] under different location error thresholds. It could be seen from Fig. 1 that when tracking targets in OCC and BC video sequences, the tracking precisions of our proposed are the best, which are 3.0% and 2.0% higher than CSR-DCF [11]; for the video sequences with OV, our proposed ranks second only to GFSDCF [48] in terms of tracking precision; for targets in the video sequence with DEF, the tracking precision of our proposed tied with ECO [25] for the second place; and when tracking targets in the video sequences with IV, SV, MB, and FM, although our tracking precisions have fallen to some extent, the worst ranking of our proposed in tracking precisions is sixth, which is still better than some state-of-the-art trackers, such as STRCF [10], BACF [45], AutoTrack [43], and MCCT_H [24].

Figure 2 also shows the success plots of each tracker to video sequences with different attributes in OTB100 data set under different overlap thresholds. Results from Fig. 2, suggest that when tracking targets in video sequences with OCC, DEF and BC attributes, the AUC scores of our proposed all rank first, and 0.026, 0.035 and 0.027 higher than the tackers in the second place, meanwhile, 0.046, 0.035 and 0.027 higher than CSR-DCF [11], respectively; for the video sequences with OV, FM, and SV, our proposed ranks second only to GFSDCF [48] in AUC scores, and 0.020, 0.009 and 0.056 higher than CSR-DCF [11] respectively; when tracking targets in the video sequences with OPR, IPR, IV and MB attributes, the AUC scores of our proposed all ranks third and all higher than CSR-DCF [11]; for the video sequences with LR attributes, the AUC score of our proposed ranks fourth, but still 0.045 higher than CSR-DCF [11]. On the whole, the tracking success rate of our proposed on OTB100 [13] ranks second with an average AUC score of

0.856, which is only 0.010 lower than the first-ranked GFSDCF [48], but 0.033 higher than the sixth-ranked CSR-DCF [11].

Figure 3 shows trackers' overlap rates and pixel errors on some OTB100 sequences. Table 1 shows some trackers' average overlap rates and pixel errors on all OTB100 sequences, here we just list 11 trackers out of 17 with better performances for clearly present the figures, and it can be seen from Fig. 3 and Table 1 that our proposed tracker has better performances in these two metrics with 67% and 13.0 pixels separately.

It could also be observed from Figure 4 that the tracking results of some trackers on some OTB100 [13] frames suggest that the targets are more difficult to track. However, for the targets in sequences with SV, OCC, DEF, MB and BC attributes, results in Fig. 4, show that our proposed tracker has better tracking performances than AutoTrack [43], BACF [45], CSR-DCF [11], MCCT [24] and STRCF [10], etc.

C. THE VOT2016 DATA SET

The VOT2016 data set [14] as used in the 2016 VOT (Visual-Object-Tracking) challenge, contains 60 challenging public image sequences, which include toys, faces, vehicles, animals and many other common target categories, and all with different attributes, such as camera_motion, empty, illu_change, motion_change, occlusion and size change. In addition, the part of the image sequences in the VOT2016 data set are the same as those of the OTB data set. In order to demonstrate the tracking performances of our proposed tracker, we implemented the tracker on VOT2016 [14]. Consequently, the tracking accuracy rate, A-R rank, and AR plot, overlap rate, and pixel error values were measured to evaluate the tracking performances of all trackers.

Table 2 presents tracking accuracy rates and A-R rank values of the targets in the video sequences of the VOT2016 data set [14]. Thus, result from Table 2 show that our tracking accuracy rates to targets all rank first in almost all attributes video sequences, except the size_change, with 3.9%, 3.0%, 6.7%, 1.9% and 16.7% higher than CSR-DCF [11]. Furthermore, results in specific, to tracking targets in video sequences with occlusion, show a tracking accuracy rate of 16.7% higher than CSR-DCF [11], thus, indicating that the time regularization information of sample is really helpful in improving the tracking performances to the occluded target. In addition, results suggest that our tracker ranks second only to GFSDCF [48] in tracking accuracy rate to the targets with size_change, and also 1.6% higher than CSR-DCF [11]. In general, our tracker performs favorably against the state-of-the-art trackers on VOT2016 data set [14] in tracking accuracy rate, and 5.7% higher than CSR-DCF [11]. It can further be observed from Table 2 that the A-R rank of our tracker exceeds the tied second trackers CSR-DCF [11] and GFSDCF [48] by 3.7%.

Table 3 shows some trackers' average overlap rates and pixel errors on all VOT2016 [14] sequences, and it can be seen that our proposed tracker ranks second and third in average overlap rates and pixel errors with 50.2% and 43.19 pixels separately. Figure 5 also show 11 trackers' overlap rates and pixel errors on some VOT2016 [14] sequences,

thus implying that our proposed tracker has a standout overlap rate and pixel error performances.

From Figure 6, the AR plots for mean, camera_motion, empty, illum_change, motion_change, occlusion, and size_change of all trackers used in this paper, suggest that our tracking accuracy rates to the targets in almost all attributes video sequences in VOT2016 [14] all ranked the first, except the size_change, followed by GFSDCF [48] and CSR-DCF [11], which is consistent with all trackers tracking performances in Table 2. It can also be observed that all abscissa values in Fig. 6 indicate that all the tracking robustness values of our tracker to all attributes targets are not as good as other trackers, although they are all higher than 0.8. As a result, most of the trackers are at the expenses of tracking accuracy, and only the tracker proposed in this paper can maintain the balance of tracking accuracy and robustness at the same time accurately and robustly track the targets.

Figure 7 shows the object tracking results of some trackers on some VOT2016 [14] challenging frames. Thus, the tracking results in frames with camera_motion, illum_change, occlusion and size_change as displayed in Figure 7, suggest that our proposed outperforms some state-of-the-art trackers, such as CSR-DCF [11], STRCF [10], AutoTrack [43] and STRCF_Deep [10].

TABLE 2.
THE TRACKING ACCURACY RATE AND A-R RANK VALUES OF SOME TRACKING ALGORITHMS
(THE RED, BLUE AND GREEN NUMBERS DENOTE THE 1ST, 2ND AND 3RD RANK)

Algorithm	camera_motion	empty	illum_change	motion_change	occlusion	Size_change	Mean	A-R rank
ARCF_H	0.306	0.351	0.382	0.239	0.266	0.252	0.299	0.294
ARCF_HC	0.419	0.352	0.330	0.228	0.241	0.239	0.302	0.312
AutoTrack	0.309	0.353	0.342	0.210	0.267	0.209	0.282	0.282
BACF	0.377	0.306	0.325	0.238	0.201	0.235	0.280	0.287
C-COT	0.396	0.397	0.327	0.314	0.250	0.328	0.335	0.352
CF2	0.370	0.299	0.366	0.265	0.276	0.250	0.304	0.300
CSR-DCF	0.467	0.454	0.457	0.433	0.371	0.387	0.428	0.439
DAT-USABLE	0.348	0.306	0.237	0.222	0.189	0.224	0.254	0.272
ECO	0.378	0.404	0.441	0.329	0.278	0.354	0.364	0.364
GFSDCF	0.503	0.441	0.465	0.420	0.304	0.439	0.429	0.439
MCCT	0.498	0.428	0.362	0.367	0.274	0.364	0.382	0.405
MCCT-H	0.413	0.414	0.270	0.291	0.293	0.304	0.331	0.354
SCT4	0.026	0.047	0.070	0.036	0.078	0.035	0.049	0.041
STAPLE	0.401	0.418	0.402	0.319	0.282	0.316	0.356	0.364
STRCF	0.359	0.359	0.370	0.256	0.286	0.267	0.316	0.316
STRCF_Deep	0.379	0.400	0.396	0.295	0.299	0.309	0.346	0.349
Ours	0.506	0.484	0.524	0.452	0.538	0.403	0.485	0.476

TABLE 3.
THE AVERAGE OVERLAP RATE AND PIXEL ERROR VALUES OF SOME TRACKING ALGORITHMS ON VOT2016
(THE RED, BLUE AND GREEN NUMBERS DENOTE THE 1ST, 2ND AND 3RD RANK)

Algorithm	AutoTrack	BACF	C-COT	CSR-DCF	ECO	GFSDCF	MCCT	MCCT_H	STRCF	STRCF_Deep	Ours
-----------	-----------	------	-------	---------	-----	--------	------	--------	-------	------------	------

Overlap Rate(%)	0.355	0.364	0.390	0.496	0.414	0.503	0.450	0.376	0.402	0.423	0.502
Pixel Error(pixel)	88.33	71.14	75.76	25.99	67.03	40.98	58.36	87.58	63.90	65.79	43.19

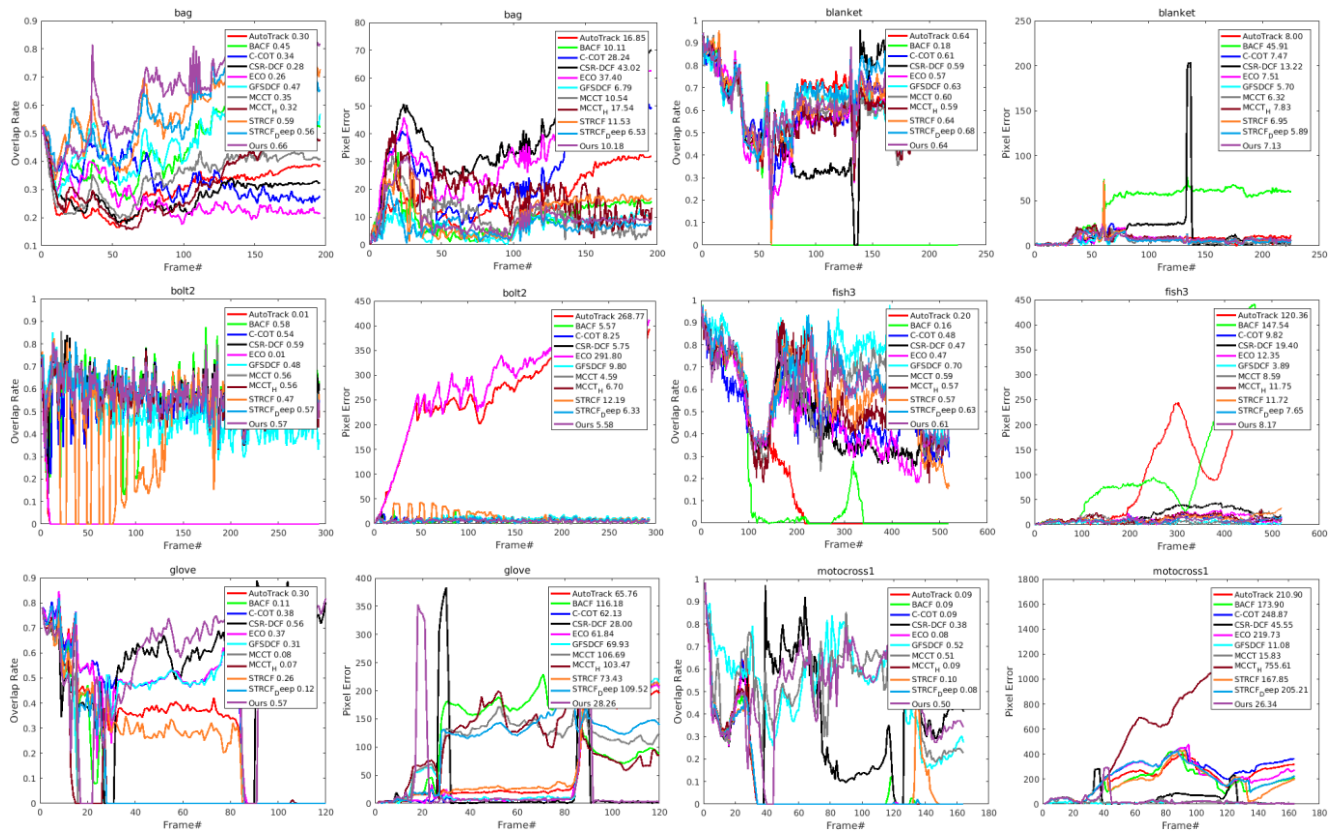


FIGURE 5. Trackers' overlap rates and pixel errors on some VOT2016 sequences

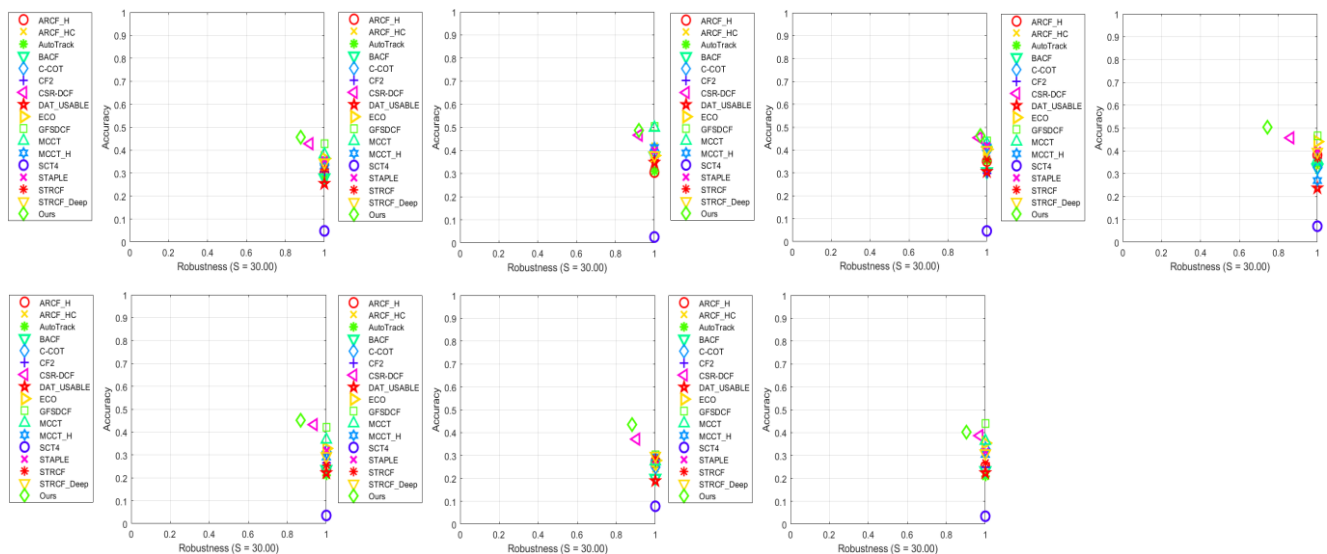


FIGURE 6. AR plot of each tracker on VOT2016 data set (figures, from top to bottom and left to right, are AR plots for mean, camera_motion, empty, illum_change, motion_change, occlusion and size_change)

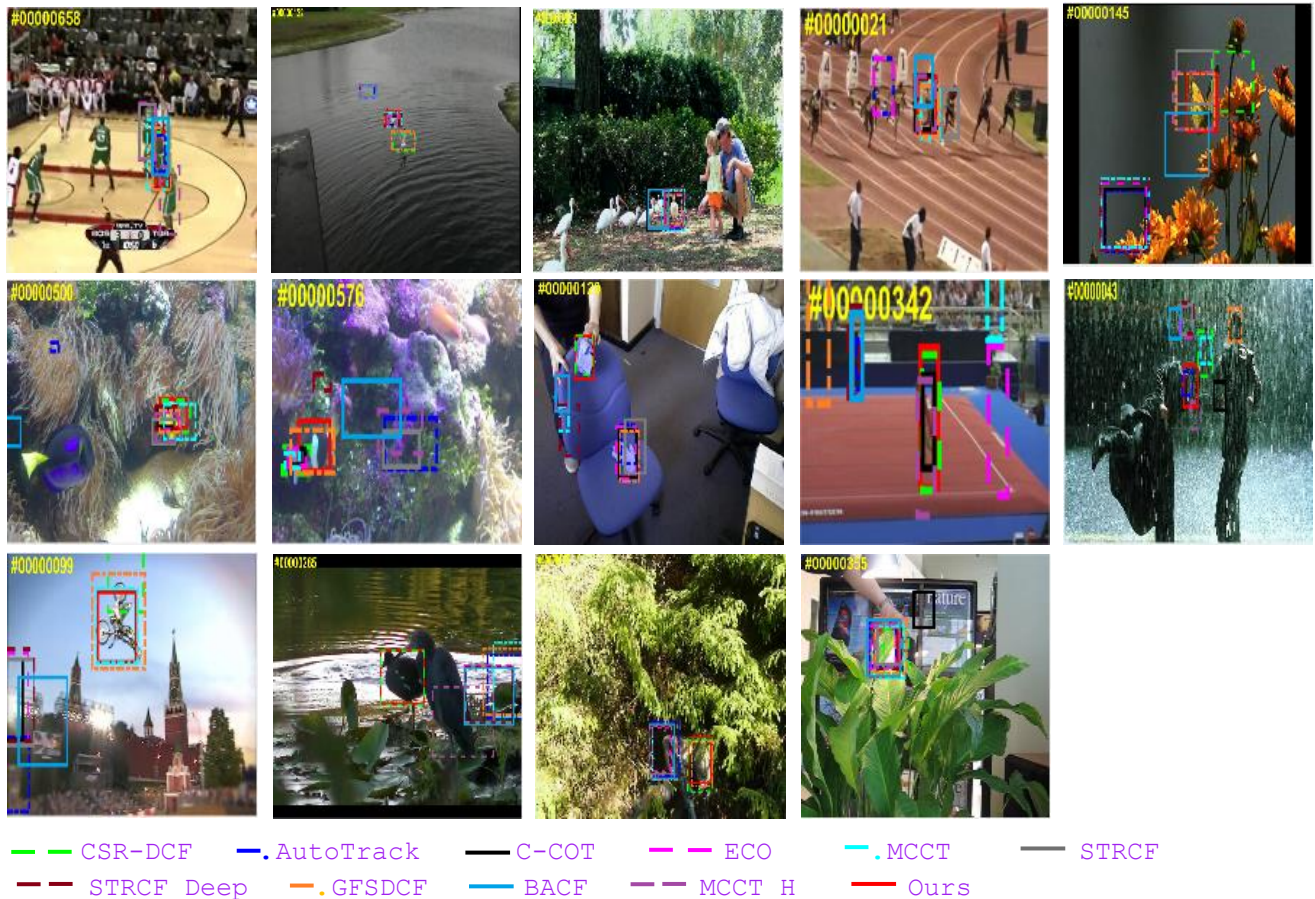


FIGURE 7. Tracking results on some challenging image sequences on VOT2016

D. THE TC128 DATA SET

The TC128 [54] collected 78 new visually challenging videos on the basis of OTB50 data set, contains 128 color sequences with ground truth and challenge factor annotations, such as IV, SV, OCC, DEF, MB, FM, and IPR, etc. The targets to be tracked in these 128 sequences are diverse, including pedestrians, basketballs, ships, cars, cups, animals, toys, fish, kites, and airplanes.

Figure 8 shows the precision plots of each tracker for video sequences with different challenge factor annotations under different location error thresholds. As can be seen from Fig. 8 that when tracking targets in IV, OCC, and OV video sequences, the tracking precisions of our proposed are the best and with margin 0.8%, 0.4% and 4.6% to CSR-DCF [11]; our proposed tracker ranks second in BC, MB, and SV video sequences; and when tracking targets in the video sequences with FM, IPR, and OPR, our proposed tracker ranks third; for the targets in sequences with DEF and LR, although the tracking precisions of our proposed have fallen to some extent, it still ranks the forth and is better than some state-of-the-art trackers. On average, the tracking precision of our tracker ranks first.

Figure 9 shows the success plots of each tracker. Results from Fig. 9, suggest that when tracking targets in BC, MB,

DEF, IV, IPR, LR, OCC, and OV video sequences, the AUC scores of our tracker all rank first; for the video sequences with FM, OPR, and SV, our tracker ranks second only to GFSDCF [48] in AUC scores, and 0.04, 0.087 and 0.098 higher than CSR-DCF [11] respectively. On the whole, the tracking success rate of our proposed tracker on TC128 ranks second with an average AUC score 0.785 and a margin 0.013 to the second GFSDCF [48] and 0.096 to the eighth CSR-DCF [11].

Figure 10 shows the overlap rates and pixel errors of some state-of-the-art trackers on some TC128 video sequences. It could be seen from Fig. 10 that our proposed tracker have higher overlap rates and lower pixel errors. Table 4 shows the average overlap rates and pixel errors on all TC128 sequences, and we can see that our proposed tracker all ranks first in average overlap rates and pixel errors with 64.0% and 20.82 pixels separately.

Figure 11 shows some trackers tracking results in some TC128 frames. The tracking results suggest that for the targets in sequences with BC, MB, DEF, IV, IPR, LR, OCC, our proposed tracker all tracked the targets more accurately and outperforms some state-of-the-art trackers, such as CSR-DCF [11], STRCF_Deep [10], and GFSDCF [48], etc.

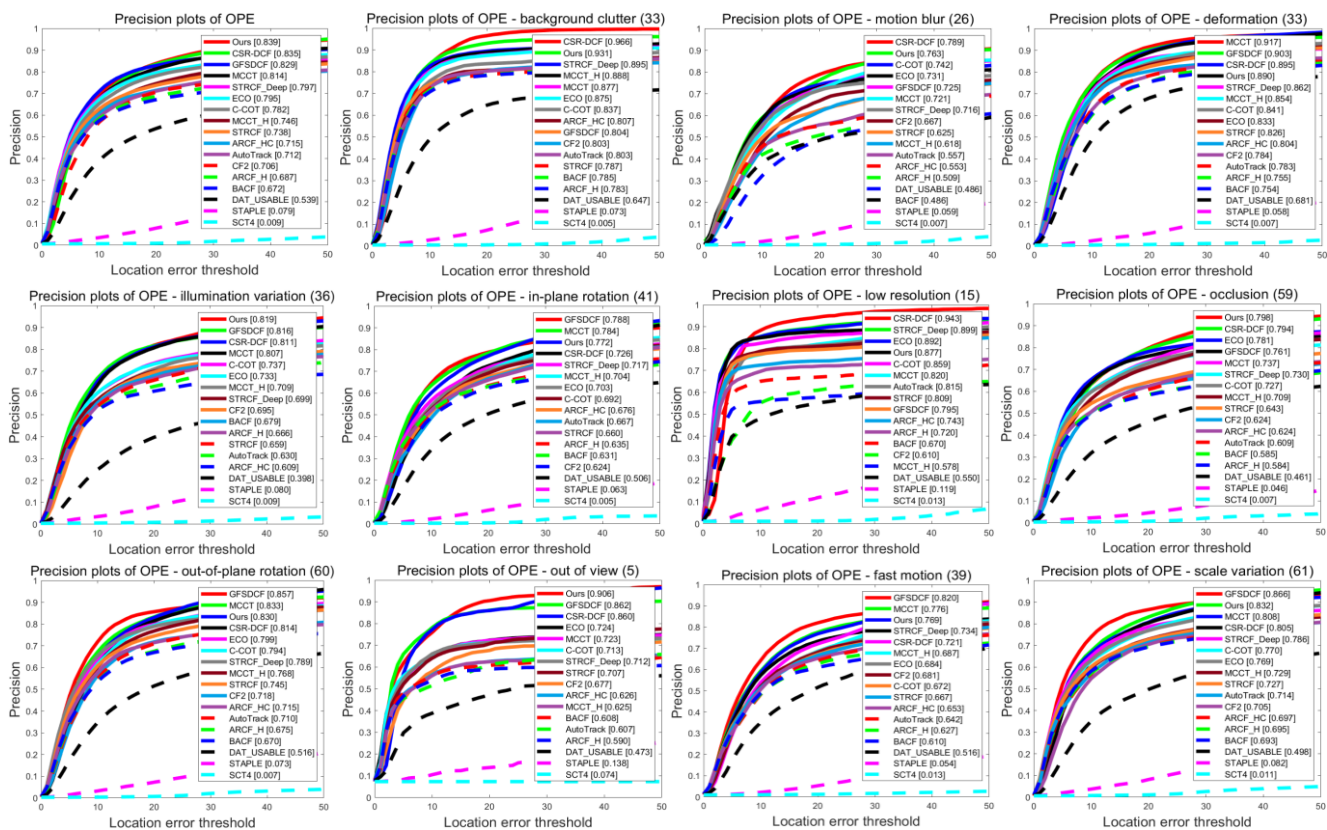


FIGURE 8. Trackers' Precision plots on TC128

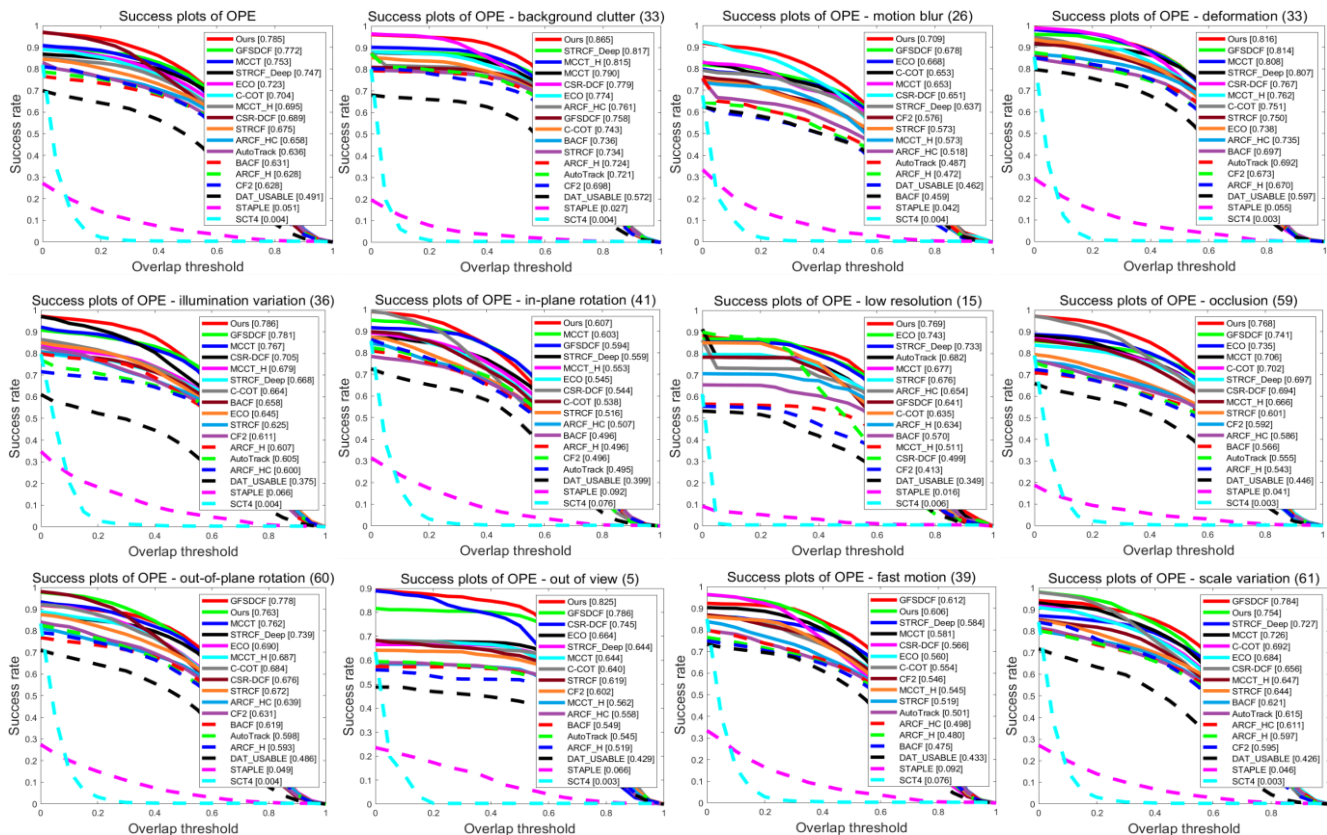


FIGURE 9. Trackers' Success plots on TC128

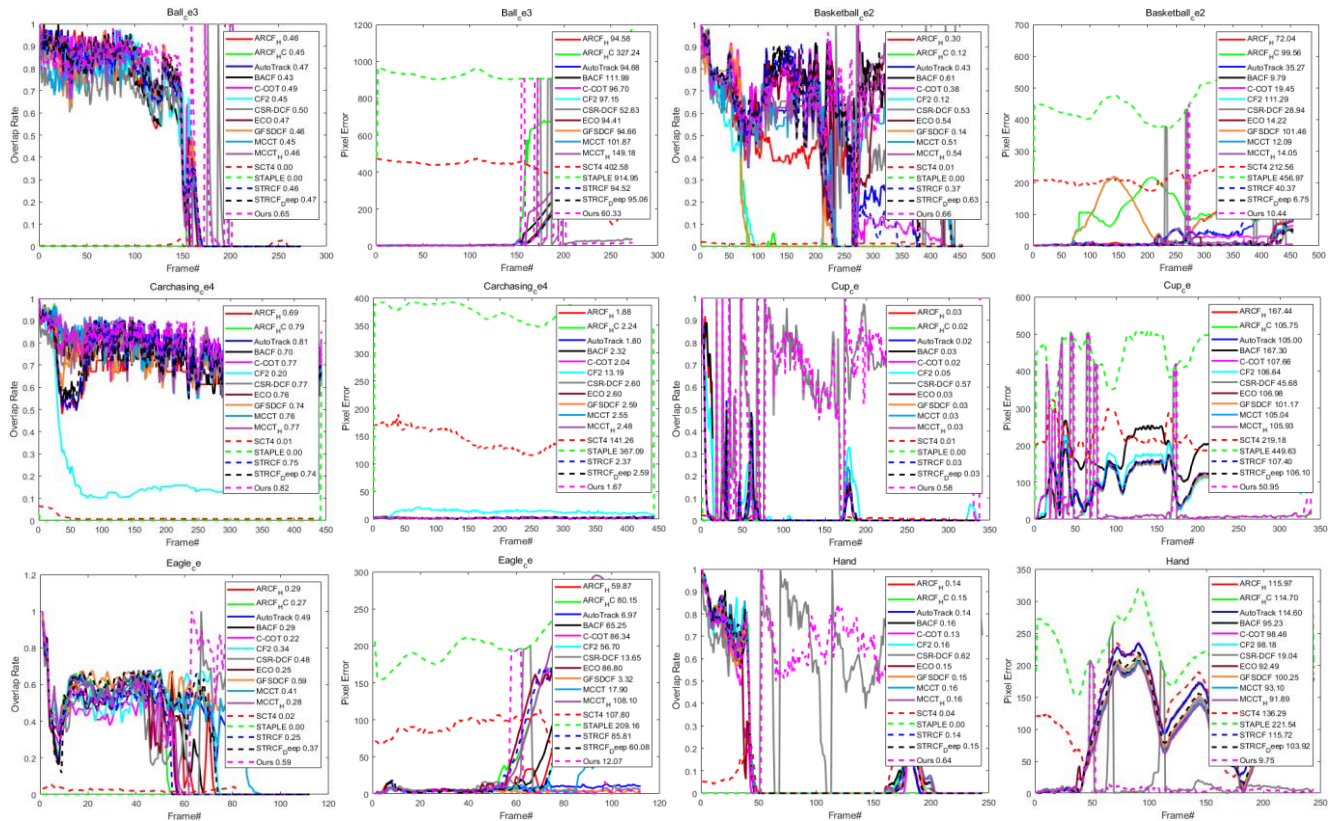


FIGURE 10. Trackers' overlap rates and pixel errors on some TC128 sequences

TABLE 4.
THE AVERAGE OVERLAP RATE AND PIXEL ERROR VALUES OF SOME TRACKING ALGORITHMS ON TC128
(THE RED, BLUE AND GREEN NUMBERS DENOTE THE 1ST, 2ND AND 3RD RANK)

Algorithm	AutoTrack	BACF	C-COT	CSR-DCF	ECO	GFSDCF	MCCT	MCCT_H	STRCF	STRCF_Deep	Ours
Overlap Rate(%)	0.525	0.510	0.579	0.587	0.594	0.609	0.600	0.565	0.557	0.599	0.640
Pixel Error(pixel)	46.90	56.26	30.04	21.19	27.16	25.87	25.17	40.59	43.32	32.09	20.82



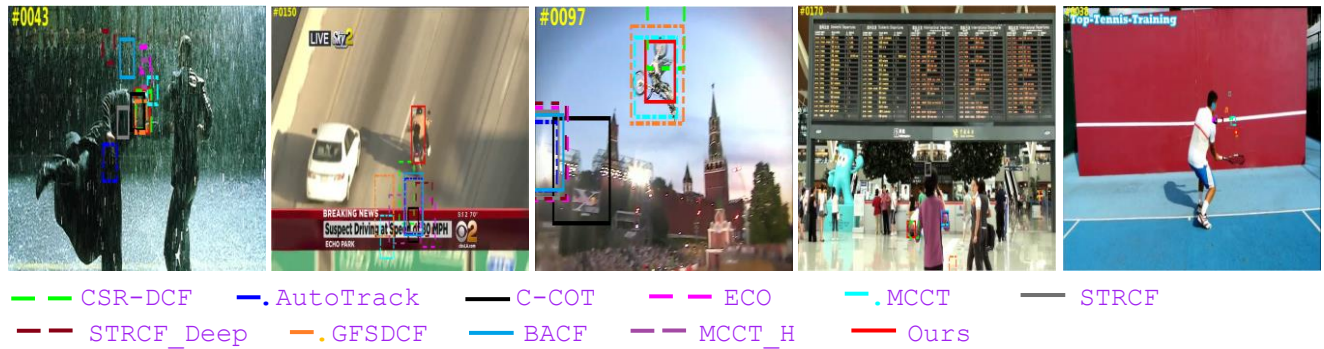


FIGURE 11. Tracking results on some challenging image sequences on TC128

E. THE UAV123 DATA SET

The UAV123 data set [2] obtained by unmanned aerial vehicles (UAV) at low altitude, contains 123 challenging videos with ground truth and 12 kinds of challenge factor annotations, such as IV, SV, Partial Occlusion (POC), Full Occlusion (FOC), Out-of-View (OV), FM, Camera Motion (CM), Similar Object (SOB), Aspect Ratio Change (ARC), Viewpoint Change (VC), BC, and LR. The videos of UAV123 data set are basically shot from top altitude to bottom, and some videos in this data set are shot in real scenes, while others are constructed in virtual environments. Since each video contains more video frames, thus, it is often used to evaluate the long-term tracking performance of a object tracker. The main targets in the data set are pedestrians, ships, airplanes, and cars, in addition, many small targets are contained in it. These factors put forward high performance requirements for the tracker to be tested.

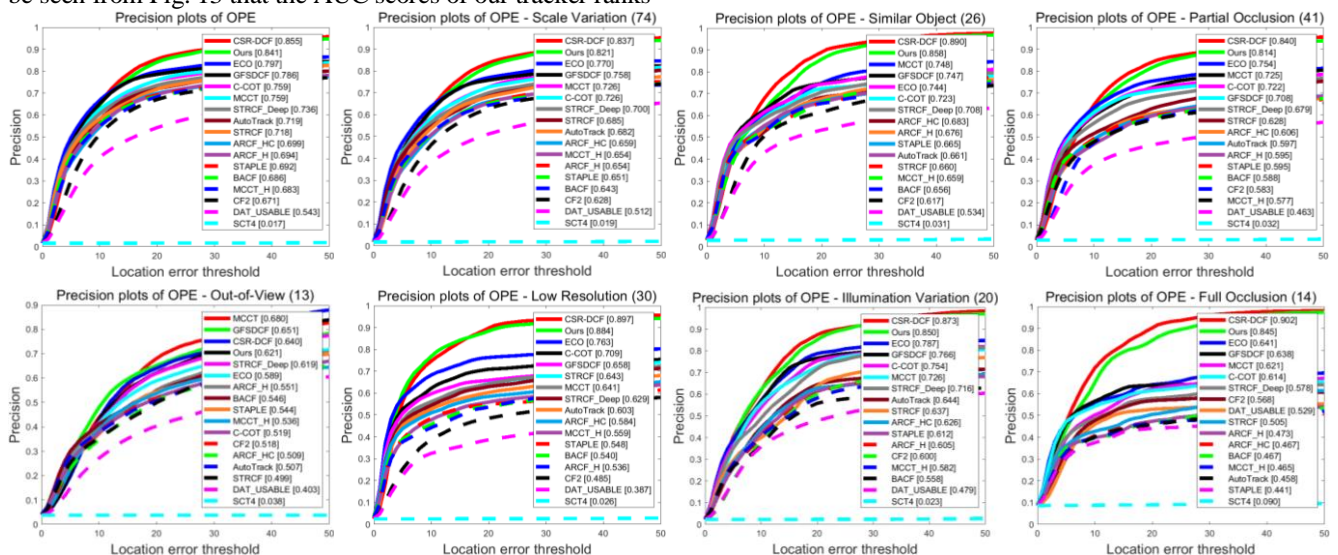
Figure 12 shows the precision plots of each tracker. It could be seen from Fig. 12 that when tracking the targets in SV, SOB, POC, LR, IV, FOC, FM, CM, BC, ARC, VC, and OV video sequences, our tracker ranks second only to GFSDCF [48] and CSR-DCF [11]; when tracking the targets in OV video sequences, although our tracker just rank forth, it is still better than most state-of-art trackers. On average, our tracking precision of ranks the second with 84.1%.

Figure 13 shows the success plots of each tracker. As can be seen from Fig. 13 that the AUC scores of our tracker ranks

first when tracking the targets in BC video sequences with a margin 0.04 to CSR-DCF [11]; when tracking the targets with LR and FOC annotations, our tracker ranks second; for the video sequences with with SOB, SV, OV, IV, FM and CM annotations, our tracker ranks third; though the success plots of our tracker to the targets with POC, ARC, and VC annotations are poor, the worst ranking of our tracker is fifth, which is still better than trackers like CSR-DCF [11] and STRCF [10], etc. On the whole, the tracking success rate of our tracker on UAV123 ranks forth with an average AUC score of 0.644 with a margin 0.048 to CSR-DCF [11].

Figure 14 shows the overlap rates and pixel errors of some state-of-the-art trackers on some UAV123 video sequences. It could be seen from Fig. 14 that our proposed tracker have higher overlap rates and lower pixel errors. Table 5 shows the average overlap rates and pixel errors on all UAV123 sequences, and we can see that the average overlap rate of our proposed tracker ranks forth, and our pixel error ranks second. Results from Fig. 14 and Table 5 suggest that our tracker have better tracking performance on UAV123 data set.

Figure 15 shows the tracking results of some trackers in some challenging UAV123 frames. It could also be observed from Fig. 15 that for the targets in sequences with BC, LR, FOC, etc., our proposed tracker all successfully and accurately tracked the targets, which fully demonstrated its outstanding object tracking abilities.



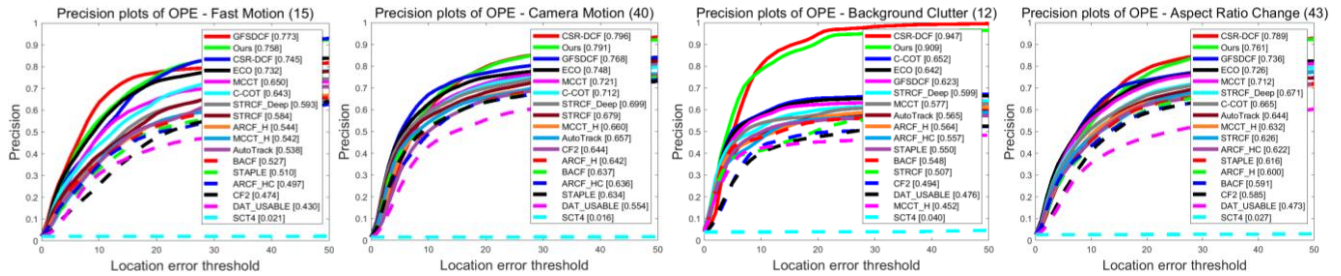


FIGURE 12. Trackers' Precision plots on UAV123

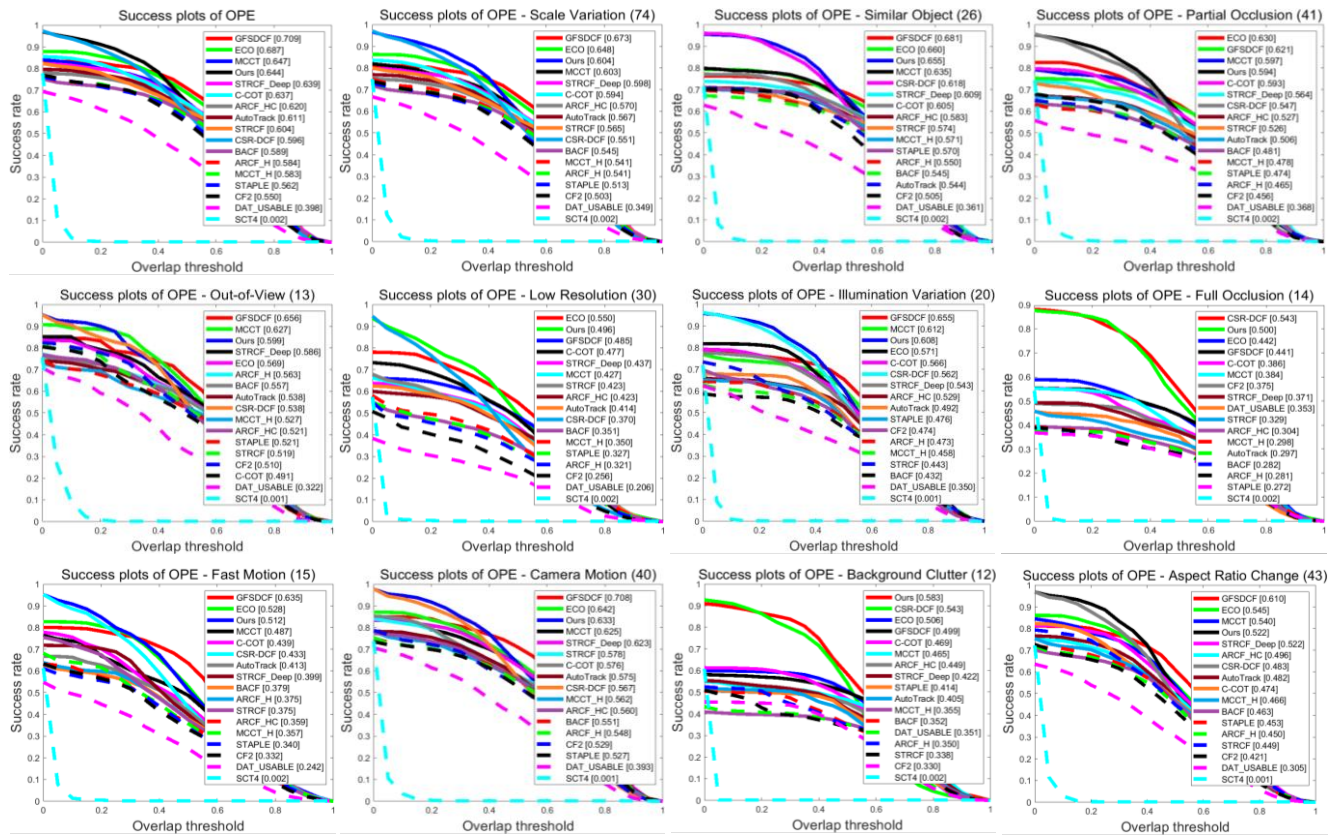
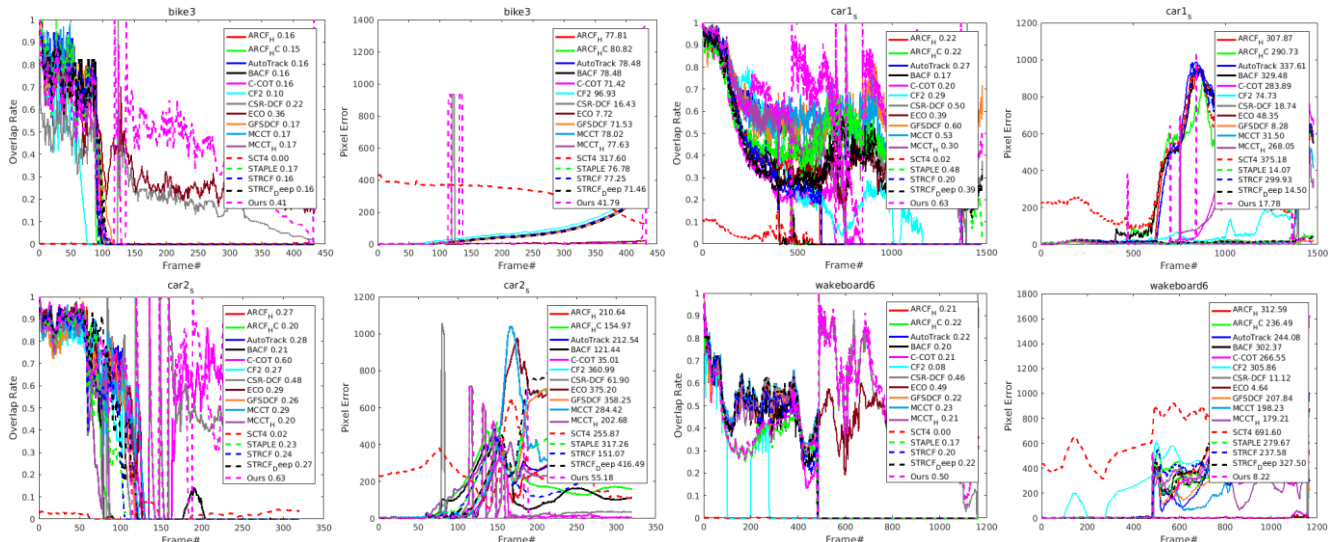


FIGURE 13. Trackers' Success plots on UAV123



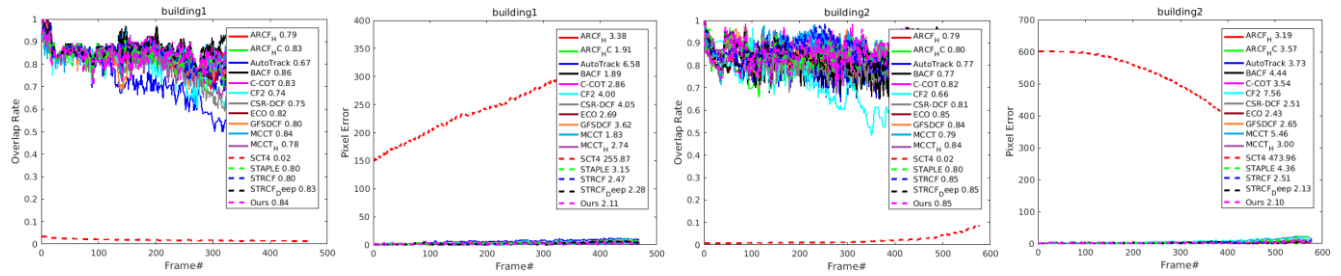


FIGURE 14. Trackers' overlap rates and pixel errors on some UAV123 sequences

TABLE 5.
THE AVERAGE OVERLAP RATE AND PIXEL ERROR VALUES OF SOME TRACKING ALGORITHMS ON UAV123
(THE RED, BLUE AND GREEN NUMBERS DENOTE THE 1ST, 2ND AND 3RD RANK)

Algorithm	AutoTrack	BACF	C-COT	CSR-DCF	ECO	GFSDCF	MCCT	MCCT_H	STRCF	STRCF_Deep	Ours
Overlap Rate(%)	0.570	0.547	0.582	0.551	0.619	0.607	0.572	0.535	0.573	0.592	0.591
Pixel Error(pixel)	46.77	57.61	34.98	16.16	27.28	36.75	31.06	60.36	37.87	43.62	18.28

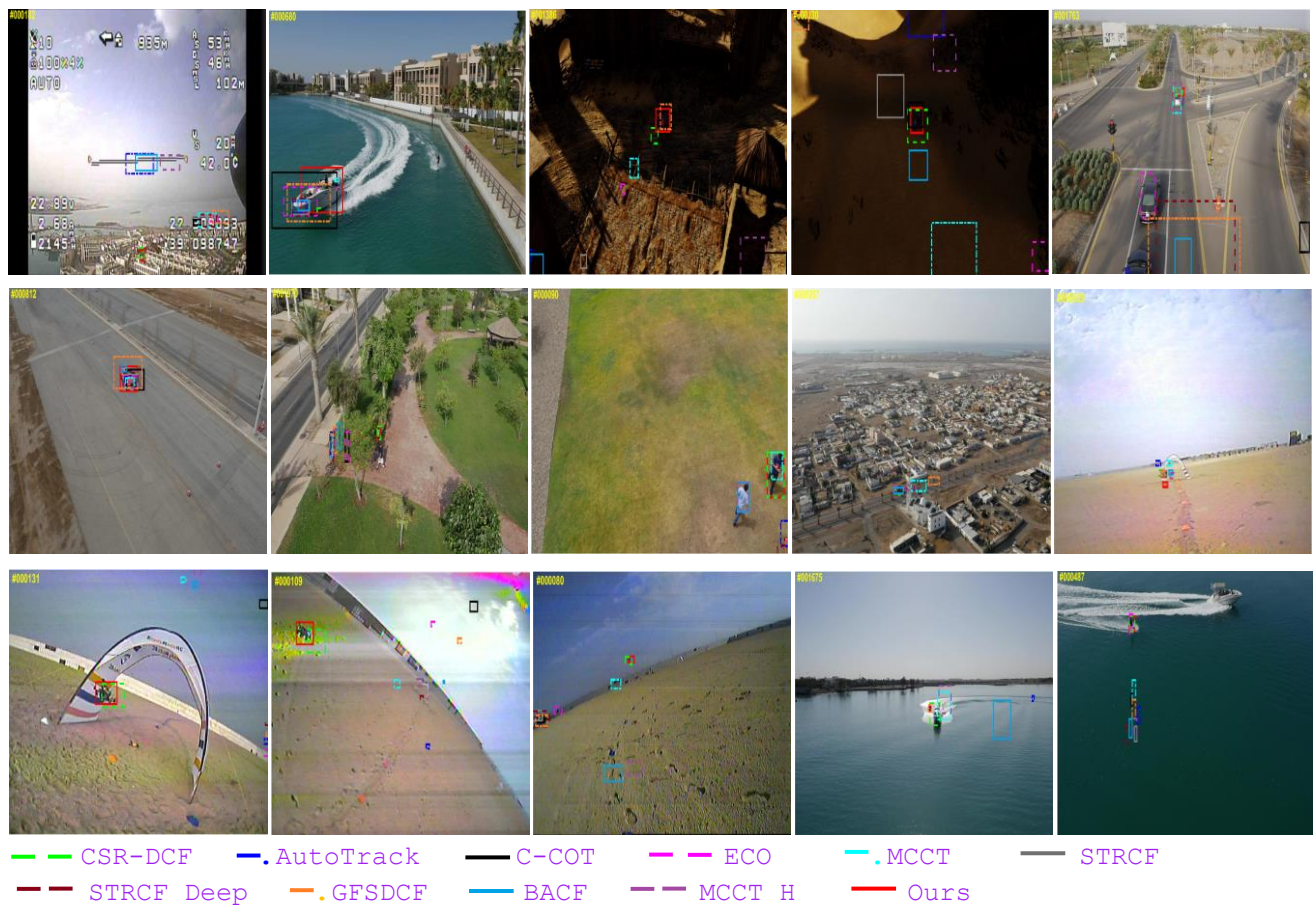


FIGURE 15. Tracking results on some challenging image sequences on UAV123

V. EXPERIMENTAL ANALYSIS

Experimental results on the OTB100 [13], VOT2016 [14], TC128 [54], and UAV123 [2] data sets, suggests that the proposed correlation filters tracker with spatial and channel reliabilities and time regularization can effectively solve the boundary effects by making full use of the spatial, channel and temporal information of samples. Especially for the

targets with significant appearance variations, the tracking performances of our proposed tracker are better than some state-of-the-art trackers, such as GFSDCF [48], CSR-DCF [11], MCCT [24] and C-COT [26].

Moreover, the experimental results demonstrated that the tracking precision, success rate, overlap rate, and pixel error of our proposed are better than CSR-DCF [11] in most video sequences in the OTB100 [13], VOT2016 [14], TC128 [54],

and UAV123 [2] data sets, thus, suggesting that the time regularization information of samples is effective in alleviating the boundary effects and improving the tracking performance of correlation filters. It can further be inferred that, it is important to combine the time regularization information with channel and spatial reliabilities of samples to train correlation filters with more target tracking ability.

VI. CONCLUSION

In this paper, by introducing the time regularization information of samples into the correlation filters with sample spatial and channel reliabilities for the first time to train correlation filter to alleviate the boundary effects and improve the object tracking ability and robustness, we proposed a multi-information fusion correlation filters tracker, that is different from the common correlation filter trackers. In addition, the use of alternating direction multiplier method (ADMM) to solve the objective function of the proposed tracker reduces its time complexity. The paper further demonstrated that our proposed tracker with HOG and CN features performs favorably against some state-of-the-art trackers, such as STRCF, CSR-DCF, MCCT, and AutoTrack in terms of tracking precision, success rate, tracking accuracy, A-R rank, overlap rate, as well as pixel errors with extensive experiments on OTB100, VOT2016, TC128, and UAV123 data sets. In particular, it should be noted that our proposed tracker has better tracking performances for targets with more significant appearance variations. Finally, it is thus suggested that future studies introduce deep features into the proposed tracker to further improve its object tracking performance.

REFERENCES

- [1] M. Kristan, J. Matas, A. Leonardis, T. Vojir, R. Pflugfelder, G. Fernandez, G. Nebehay, F. Porikli, and L. Cehovin, "A Novel Performance Evaluation Methodology for Single-Target Trackers," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 38, no. 11, pp. 2137-2155, Nov. 2016.
- [2] M. Mueller, N. Smith, and B. Ghanem, "A Benchmark and Simulator for UAV Tracking," in *Eur. Conf. Comput. Vis.*, Sep. 2016, pp. 445-461.
- [3] M. Kristan, and J. Matas, "The Visual Object Tracking VOT2015 Challenge Results," in *IEEE Int. Conf. Comput. Vis. Workshop*, Santiago, Dec. 2015, pp. 564-586.
- [4] T. Xu, Z. Feng, X. 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, 2019.
- [5] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-Speed Tracking with Kernelized Correlation Filters," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 37, no. 3, pp. 583-596, Mar. 2015.
- [6] D. Bolme, J. Beveridge, B. Draper, and Y. Lui, "Visual object tracking using adaptive correlation filters," in *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, San Francisco, CA, 2010, pp. 2544-2550.
- [7] J. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the Circulant Structure of Tracking-by-detection with Kernels," in *Eur. Conf. Comput. Vis.*, Berlin, Heidelberg, Oct. 2012, pp. 702-715.
- [8] J. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-Speed Tracking with Kernelized Correlation Filters," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 37, no. 3, pp. 583-596, 2015.
- [9] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg, "Learning Spatially Regularized Correlation Filters for Visual Tracking," in *IEEE Int. Conf. Comput. Vis.*, Santiago, 2015, pp. 4310-4318.
- [10] F. Li, C. Tian, W. Zuo, L. Zhang, and M. Yang, "Learning Spatial-Temporal Regularized Correlation Filters for Visual Tracking," in *IEEE Conf. Comput. Vis. Pattern Anal.*, Salt Lake City, UT, 2018, pp. 4904-4913.
- [11] A. Lukežić, T. Vojir, L. C. Zajc, J. Matas, and M. Kristan, "Discriminative Correlation Filter with Channel and Spatial Reliability," in *IEEE Conf. Comput. Vis. Pattern Anal.*, Honolulu, HI, 2017, pp. 4847-4856.
- [12] K. Dai, D. Wang, H. Lu, C. Sun, and J. Li, "Visual Tracking via Adaptive Spatially Regularized Correlation Filters," in *IEEE Conf. Comput. Vis. Pattern Anal.*, Long Beach, CA, USA, 2019, pp. 4665-4674.
- [13] Y. Wu, J. Lim, and M. Yang, "Object Tracking Benchmark," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 37, no. 9, pp. 1834-1848, Sept. 2015.
- [14] M. Kristan, Aleš Leonardis, J. Matas, M. Felsberg, and Z. Chi, "The Visual Object Tracking VOT2016 Challenge Results," in *Eur. Conf. Comput. Vis. Workshops*, Amsterdam, the Netherlands, Oct. 2016, pp. 777-823.
- [15] 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, 2011.
- [16] M. Mueller, N. Smith, and B. Ghanem, "Context-Aware Correlation Filter Tracking," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, 2017, pp. 1387-1395.
- [17] C. Sun, D. Wang, H. Lu, and M. Yang, "Correlation Tracking via Joint Discrimination and Reliability Learning," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, 2018, pp. 489-497.
- [18] J. Li, X. Zhou, S. Chan, and S. Chen, "Robust Object Tracking via Large Margin and Scale-Adaptive Correlation Filter," *IEEE Access*, vol. 6, pp. 12642-12655, 2018.
- [19] T. Zhang, C. Xu, and M. Yang, "Learning multi-task correlation particle filters for visual tracking," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 41, no. 2, pp. 365-378, 2019.
- [20] F. Liu, C. Gong, X. Huang, T. Zhou, J. Yang, and D. Tao, "Robust Visual Tracking Revisited: From Correlation Filter to Template Matching," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2777-2790, 2018.
- [21] X. Yang, H. Zhang, L. Yang, C. Yang, and P. Liu, "A Joint Multi-Feature and Scale-Adaptive Correlation Filter Tracker," *IEEE Access*, vol. 1, pp. 99, 2018.
- [22] S. Zhang, W. Lu, W. Xing, and L. Zhang, "Learning Scale-Adaptive Tight Correlation Filter for Object Tracking," *IEEE Trans. Cybern.*, vol. 50, no. 1, pp. 270-283, 2020.
- [23] W. Ruan, J. Chen, Y. Wu, J. Wang, C. Liang, R. Hu, and J. Jiang, "Multi-Correlation Filters With Triangle-Structure Constraints for Object Tracking," *IEEE Trans. Multimed.*, vol. 21, no. 5, pp. 1122-1134, May 2019.
- [24] N. Wang, W. Zhou, Q. Tian, R. Hong, M. Wang, and H. Li, "Multi-cue Correlation Filters for Robust Visual Tracking," in *IEEE Conf. Comput. Vision Pattern Recognit.*, Salt Lake City, UT, 2018, pp. 4844-4853.
- [25] M. Danelljan, G. Bhat, F. S. Khan, and M. Felsberg, "ECO: Efficient Convolution Operators for Tracking," in *IEEE Conf. Comput. Vision Pattern Anal.*, Honolulu, HI, 2017, pp. 6931-6939.
- [26] M. Danelljan, A. Robinson, F. Khan, and M. Felsberg, "Beyond Correlation Filters: Learning Continuous Convolution Operators for Visual Tracking," in *Eur. Conf. Comput. Vision*, 2016, pp. 472-488.
- [27] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," in *Br. Mach. Vision Conf.*, 2014.
- [28] Z. He, Z. Zhang, and C. Jung, "Fast Fourier Transform Networks for Object Tracking Based on Correlation Filter," *IEEE Access*, vol. 6, pp. 6594-6601, 2018.
- [29] 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.

- [30] K. Li, Y. Kong, and Y. Fu, "Visual Object Tracking via Multi-Stream Deep Similarity Learning Networks," *IEEE Trans. Image Process.*, vol. 29, pp. 3311-3320, 2020.
- [31] P. Bhat, B. Subudhi, T. Veerakumar, V. Laxmi, and M. Gaur, "Multi-Feature Fusion in Particle Filter Framework for Visual Tracking," *IEEE Sensors J.*, vol. 20, no. 5, pp. 2405-2415, 2020.
- [32] Z. Huang, Y. Yu, and M. Xu, "Bidirectional Tracking Scheme for Visual Object Tracking Based on Recursive Orthogonal Least Squares," *IEEE Access*, vol. 7, pp. 159199-159213, 2019.
- [33] D. LEE, "One-Shot Scale and Angle Estimation for Fast Visual Object Tracking," *IEEE Access*, vol. 7, pp. 55477-55484, 2019.
- [34] 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.
- [35] Z. Chen, B. Zhong, G. Li, S. Zhang, and R. Ji, "Siamese Box Adaptive Network for Visual Tracking," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 6668-6677.
- [36] M. Danelljan, L. Gool, and R. Timofte, "Probabilistic Regression for Visual Tracking," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 7183-7192.
- [37] Y. Xu, Z. Wang, Z. Li, Y. Yuan, and G. Yu, "SiamFC++: Towards Robust and Accurate Visual Tracking with Target Estimation Guidelines," in *The Thirty-Fourth AAAI Conf. Artif. Intel.*, New York, USA, 2020, pp. 12549-12556.
- [38] P. Li, B. Chen, W. Ouyang, D. Wang, X. Yang, and H. Lu, "GradNet: Gradient-Guided Network for Visual Object Tracking," in *IEEE Int. Conf. Comput. Vis.*, 2019, pp. 6162-6171.
- [39] K. Li, F. He, H. Yu, and X. Chen, "A Parallel and Robust Object Tracking Approach Synthesizing Adaptive Bayesian Learning and Improved Incremental Subspace Learning," *Front. Comput. Sci.*, vol. 13, no. 5, pp. 1116-1135, 2019.
- [40] K. Li, F. He, and H. Yu, "Robust Visual Tracking based on Convolutional Features with Illumination and Occlusion Handling," *J. Comput. Sci. Tech.*, vol. 33, no. 1, pp. 223-236, 2018.
- [41] K. Li, F. He, H. Yu, and X. Chen, "A Correlative Classifiers Approach based on Particle Filter and Sample Set for Tracking Occluded Target," *Appl. Math.*, vol. 32, no. 3, pp. 294-312, 2017.
- [42] J. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-Speed Tracking with Kernelized Correlation Filters," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 37, no. 3, pp. 583-596, Mar. 2015.
- [43] Y. Li, C. Fu, and Z. Huang, "AutoTrack: Towards High-Performance Visual Tracking for UAV with Automatic Spatio-Temporal Regularization," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 11923-11932.
- [44] Z. Huang, C. Fu, Y. Li, F. Lin, and P. Lu, "Learning Aberrance Repressed Correlation Filters for Real-Time UAV Tracking," in *IEEE Int. Conf. Comput. Vis.*, 2019, pp. 2891-2900.
- [45] H. Galoogahi, A. Fagg, and S. Lucey, "Learning Background-Aware Correlation Filters for Visual Tracking," in *IEEE Int. Conf. Comput. Vis.*, Venice, 2017, pp. 1144-1152.
- [46] C. Ma, J. Huang, X. Yang, and M. Yang, "Hierarchical Convolutional Features for Visual Tracking," in *IEEE Int. Conf. Comput. Vis.*, Santiago, 2015, pp. 3074-3082.
- [47] H. Possegger, T. Mauthner, and H. Bischof, "In defense of color-based model-free tracking," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, 2015, pp. 2113-2120.
- [48] T. Xu, Z. Feng, X. Wu, and K. Josef, "Joint Group Feature Selection and Discriminative Filter Learning for Robust Visual Object Tracking," in *IEEE Int. Conf. Comput. Vis.*, 2019, pp. 7950-7960.
- [49] J. Choi, H. J. Chang, J. Jeong, Y. Demiris, and J. Choi, "Visual Tracking Using Attention Modulated Disintegration and Integration," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, 2016, pp. 4321-4330.
- [50] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, and P. Torr, "Staple: Complementary Learners for Real-Time Tracking," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, 2016, pp. 1401-1409.
- [51] D. Yuan, X. Zhang, J. Liu, and D. Li, "A multiple feature fused model for visual object tracking via correlation filters," *Multimed. Tools Appl.*, vol. 78, no. 19, pp. 27271-27290, Jun. 2019.
- [52] D. Yuan, N. Fan, and Z. He, "Learning target-focusing convolutional regression model for visual object tracking," *Knowl.-Based Syst.*, Vol. 194, no. 22, pp. 105526, Apr. 2020.
- [53] D. Yuan, X. Li, Z. He, Q. Liu, and S. Lu, "Visual object tracking with adaptive structural convolutional network," *Knowl.-Based Syst.*, vol. 194, no. 22, pp. 105554, Apr. 2020.
- [54] 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.



JUNNAN WANG received the B.S. degree in Nanyang Normal University, Henan, China, in 2014. And the M.S. degree Xinjiang University, Xinjiang, China, in 2018. Currently, He is pursuing the Ph.D. degree in Computer Application Technology with Xinjiang University. His research interests include image processing, computer vision.



ZHENHONG JIA received the B.S. degree in Beijing Normal University, Beijing, China, in 1985 and 1987. And the M.S. degree and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 1987 and 1995. Currently, he is a professor at the Autonomous University Key Laboratory of signal and information processing laboratory, Xinjiang University, China. His research interests include digital image processing, photoelectric information detection and sensor.



HUICHENG LAI received the B.S. degree and the M.S. degree in Xinjiang University, Xinjiang, China, in 1986 and 1990. Currently, he is a professor at the College of Information Science and Engineering, Xinjiang University, China. His major research interests include image understanding and recognition, communication technology.



JIE YANG received his PhD from the Department of Computer Science, Hamburg University, Germany, in 1994. Currently, he is a professor at the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, China. His major research interests are object detection and recognition, data fusion and data mining, and medical image processing



NIKOLA K. KASABOV (M'93–SM'98–F'10) received the M.S. degree in computing and electrical engineering and the Ph.D. degree in mathematical sciences from the Technical University of Sofia, Sofia, Bulgaria, in 1971 and 1975, respectively. He is the Director and the Founder of the Knowledge Engineering and Discovery Research Institute (www.kedri.info)

and a Professor of Knowledge Engineering at the School of Computing and Mathematical Sciences at the Auckland University of Technology, Auckland, New Zealand. Before that, he worked as a Professor with the University of Otago, Dunedin, New Zealand, a Senior Lecturer with the University of Essex, Colchester, U.K., and an Associate Professor with the Technical University of Sofia.