A Novel Incremental Multi-Template Update Strategy for Robust Object Tracking

QINGSONG XIE [1,2], **KEWEI LIU** [2], **AN ZHIYONG** [1,3], **LEI WANG** [4], **YE LI** [5], **AND ZHONGLIANG XIANG** [5]

[1] School of Computer Science and Technology, Shandong Technology and Business University, Yantai 264005, China
[2] School of Information and Electronic Engineering, Shandong Technology and Business University, Yantai 264005, China
[3] Shandong Co-Innovation Center of Future Intelligent Computing, Yantai 264005, China
[4] School of Mathematics, Shandong Technology and Business University, Yantai 264005, China
[5] Department of Computer Software Institute, Weifang University of Science and Technology, Weifang 262700, China

Corresponding author: An Zhiyong (azytyut@163.com)

ABSTRACT In the field of correlation filter object tracking, the traditional template-update method easily causes template drift, so it performs poorly in complex scenes. To enhance the robustness of the template, a novel incremental multi-template update strategy is proposed in this paper. We find that reliability varies among all historical filters and that highly reliable filters are key to achieving accurate tracking. The incremental multi-template update strategy combines the local maximum-reliability filter template with the historical filter template incrementally, which is obviously different from the traditional update method. We apply this strategy to two trackers with superior performance. The experimental results of three test benchmarks, including the VOT2016, OTB100 and UAV123 datasets, show that the performance of our trackers is superior to that of the state-of-the-art trackers.

INDEX TERMS Object tracking, correlation filter, multi-template.

I. INTRODUCTION

Object tracking an important area in the field of computer vision, and it is also one of the most difficult areas to study. Object tracking refers to selecting an object in an image or video sequence and then determining the position, movement trajectory, and morphology of the object in subsequent sequences or frames according to the characteristics of the object [1]. Object tracking has been widely used in various fields, such as intelligent security, unmanned driving, smart home, and robot navigation [2]–[6]. With the advent of 5G, the prospects of object-tracking applications will be even more extensive, and the applications of object tracking will certainly bring new surprises.

In recent years, experts have proposed many methods for tracking objects. These tracking methods can be divided into two types: generative methods and discriminative

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

methods [7]. A generative method refers to describing an object to be tracked in a video through an object-representation method in computer vision and then extracting a corresponding object feature from a current frame containing the object to establish an object template [8], [9]. The method then searches the subsequent frames for the area most similar to the object template and gradually iterates to finally achieve the positioning and tracking of the object in the subsequent frames [10]–[17]. A discriminant method refers to applying both the object template and background information to the tracking system. The method sets the feature information of the object as positive samples and the background information as negative samples. Then, machine-learning methods are used for classification training. Next, the discriminant method uses the previously trained classifiers to make classification judgments in subsequent frames and determines the optimal object area [18]–[22]. The discriminative method is more accurate than the generative method because it takes both the background and

foreground into account [21], [23]–[25]. Bolme *et al.* introduced the method of correlation filtering into the discriminant method, which greatly improved the accuracy of the discriminant method [26]. High-speed tracking with Kernelized Correlation Filters (KCF), proposed by Henriques *et al.*, increased the richness of samples by introducing cyclic matrix displacement and improved the real-time performance of the tracking algorithm by using nuclear-related methods for training and detection in the frequency domain [27]. Martin *et al.* proposed efficient convolution operators for tracking (ECO) [28], which greatly improved the robustness and accuracy of object tracking and effectively reduced the amount of computations in the tracking process through convolution operations of factorization, sample space generation by compression, and sparse updating of the template.

Template updates are a critical part of the tracking performance in the correlation filters tracker. Numerous scholars have attempted to update templates [29]–[35]. Niu *et al.* proposed that the template be regularly updated unless no occlusion occurs [36]. However, it is difficult to accurately judge occlusion, so it is easy to omit information in this update method. Wang *et al.* proposed that the CF template be updated unless the HOG and color score are both above the average and that the histograms of the foreground and background as well as the sample templates used for redetection should be updated as well [37]. However, this method has a poor effect when the target and background information are similar. Yuan *et al.* proposed an adaptive structural convolutional filter model (ASCT) to enhance the robustness of deep regression trackers [4]. ASCT can effectively improve the robustness of the tracker through adaptive weighted fusion of local filters. Feng *et al.* proposed using 3 classifiers to simultaneously track the target and update the models of different classifiers according to different environments [38]. However, this update method greatly increases the amount of computations.

The importance of different historical templates has not been considered in previous DCF algorithms [39]–[41]. However, the reliability of object tracking is different at different times in the tracking process, so we believe that the reliabilities of the templates generated in the process of tracking are different. Differently reliable templates have different effects on object tracking. To improve the robustness of the correlation-filtering tracking algorithm, a multi-template adaptive updating strategy is proposed. In this strategy, we treat different historical templates with different reliability. The contributions of this strategy mainly include the following three points:

1) The reliabilities of historical filters are analyzed. We make an important assumption: each historical filter has a different reliability. If this assumption is true, then the reliability of the filter will certainly be reduced when occlusion occurs. Through analysis, we find that high-reliability filter templates are the key to achieving accurate tracking.

2) We propose a novel multi-template update strategy. In contrast to the traditional updating methods, this strategy innovatively merges the local maximum-reliability filter into the tracking-filter template and can greatly enhance the robustness of the template. Furthermore, though memory is added only to a filter algorithm, the tracking performance can be improved.

3) An adaptive-learning rate, which is set by the degree of occlusion of the object, is used to update the tracking-filter template. The degree of occlusion of the object can be calculated by the reliability difference of adjacent frames. The adaptive-learning rate can overcome the template drift to some extent in complex scenes, such as occlusion.

II. RELATED WORK

In this section, we introduce the improved method of the template into the related previous field of filtering target tracking.

The quality of the template determines the performance of the tracker. Previously, modification of the tracker template mainly included template updating and feature fusion. Many trackers have greatly improved the tracking performance by improving the classic tracker template.

First, we introduce related work on template construction. MKCF [42] proposed an approach to fusing the features of different types by means of multiple-kernel learning. MKCF makes full use of the discriminative invariance of the power spectra of various characteristics through multiple-kernel learning, thereby effectively improving the performance of the tracker. MKCFup [43] achieves good performance by optimizing the upper bound of the objective function of the original MKCF and introducing historical samples into the upper bound. Although MKCFup introduces historical samples, it does not analyze their reliability. Introducing samples with low tracking accuracy can easily cause drift of the tracking model. Yuan *et al.* proposed a multiple-feature fused model for visual-object tracking via correlation filters (MFFT) [44]. MFFT uses an adaptive weighting function based on the response map of different characteristics to eliminate the noise of different characteristics. This method makes full use of the positioning information in different features and effectively improves the robustness of tracking and hence the performance of the tracker. Yuan *et al.* proposed a target-focused convolutional regression model for visual object-tracking tasks (TFCR) [6]. TFCR uses the target-focus loss function to effectively reduce the influence of background noise on the response map of the correlation filter tracking image frame and improves the tracking accuracy. The MvCFT [45] tracker performs tracking by fusing multiple view models and selecting highly distinguishable functions. MvCFT effectively improves the robustness of the tracker by making full use of the characteristics of different features.

We now introduce the related work of template updating. Bibi *et al.* proposed Multi-Template Scale-Adaptive Kernelized Correlation Filters (MTSA) [46]. MTSA shows that it is possible to incorporate multiple multiple dimensional templates to compute the optimal filter taps. By reformulating the kernel-correlation problem and by using fixed-point optimization, MTSA demonstrated that using multiple templates improves the performance of a baseline tracker that uses only one template and a heuristic update scheme. Wang *et al.* proposed a preserving multi-cue correlation filter tracker (MCCT) [47]. During the tracking process, MCCT selects the most reliable of many experts as the tracking result and updates the template adaptively according to the robustness of the tracking.

Although previous template-transformation methods have effectively improved the performance of the tracker, these transformation methods ignore the importance of different historical templates. These trackers generate a filter template for each frame during the tracking process, and the reliabilities of these filter templates are different. If these templates are simply linearly integrated into the tracker template, then it is easy to cause pollution of the tracker template. Therefore, we believe that timely fusion of the filter template of the highly reliable historical frame into the current tracking template can enhance the robustness of the tracker when the template is updated in a complex environment.

III. BASELINE APPROACH: ECO

In this work, we adopt the recently introduced Efficient Convolution Operators for Tracking (ECO) as our baseline. The ECO algorithm is an excellent representative of correlation-filter tracking algorithms. The performance of the ECO algorithm is comparable to that of many excellent recent deep-learning tracking algorithms.

The ECO algorithm analyzes previous related filtering and tracking algorithms, summarizes three factors affecting the object-tracking performance and proposes corresponding solutions. These three factors are the template size, the training-set size, and the template update. The corresponding solution of ECO is as follows:

(1) ECO uses the factorized convolution operator method to reduce dimensions by extracting feature subsets.

The ECO algorithm decomposes the filters of D channels into a product of a D * C matrix P and the filters of C channels, where $D \gg C$. In this way, calculating the score according to the filter (f) of the corresponding layer can simplify Equation (1) into Equation (2)

$$S_{\mathrm{f}}\{x\} = f * J\{x\} = \sum_{d=1}^{D} f^{d} * J_{\mathrm{d}}\{x^{d}\}, \quad (1)$$

$$S_{\mathrm{Pf}}\{x\} = Pf * J\{x\} = \sum_{d=1}^{D} \mathrm{p}_{d,c} f^{c} * J_{\mathrm{d}}\{x^{d}\} = f * P^{T} J\{X\}, \quad (2)$$

where $J\{x\}$ is the result of the feature map converted into a continuous domain by interpolation. The matrix P can be learned in the first frame.

Therefore, the objective function is:

$$E(f, P) = ||\hat{z}^{T} P \hat{f} - \hat{y}||_{l^{2}}^{2} + \sum_{c=1}^{C} ||\hat{w} * \hat{f}^{c}||_{l^{2}}^{2} + \lambda ||P||_{F}^{2}, \quad (3)$$

ECO transforms the original D-dimensional filter into a C-dimensional filter through a factorized convolution operator, which greatly reduces the computational complexity.

(2) The ECO uses the method of the generative sample space template to merge similar samples into a component. ECO training samples are then selected from multiple components.

The ECO algorithm classifies the samples by a Gaussian mixture template (GMM) to obtain different components, each of which has a weight and a feature mean. A new sample is given when component fusion is updated, and a new component m is initialized. Here $\pi_{m} = \gamma$, $\mu_{m} = x_{j}$, where γ is the learning rate. If the number of templates exceeds the limit L, then the template with the weight π_{l} below the threshold will be discarded; otherwise, the two closest components k and l will be merged into a common component n. The merging method is as follows in Equations (4) and (5):

$$\pi_{n} = \pi_{k} + \pi_{l}, \quad (4)$$

$$\mu_{n} = \frac{\pi_{k} \mu_{k} + \pi_{l} \mu_{l}}{\pi_{k} + \pi_{l}}, \quad (5)$$

Therefore, the objective function can be simplified as:

$$E(f) = \sum_{l=1}^{L} \pi_{l} ||S_{f}\{\mu_{l}\} - y_{0}||_{L^{2}}^{2} + \sum_{d=1}^{D} ||\omega f^{d}||_{L^{2}}^{2}, \quad (6)$$

ECO uses this method to simplify the training set and guarantee sample diversity.

(3) ECO sets the update interval (Ns) of the template to 6. The update method is:

$$F^{\mathrm{t}} = (1 - \mu_{c}) F^{\mathrm{t-1}} + \mu_{c} FC^{\mathrm{t}}, \quad (7)$$

where F^{t} is the current frame template. $F^{\mathrm{t-1}}$ is the template of the previous frame and μ_{c} is the learning rate of the current frame.

IV. OUR STRATEGY

In this paper, "multi-template" refers to the filter template of the current frame, the tracking template of the previous frame and the filter template with the highest local historical confidence. Among these, the filter template with the largest local history confidence is the incremental template. We believe that the more reliable the tracking template, the higher the accuracy of tracking the target. Moreover, the reliability of the filter template close to the current frame t has a more significant impact on the tracking accuracy.

In this strategy, the location of the object is determined based on the response map of the current frame t and the tracking reliability of the current frame t is calculated based
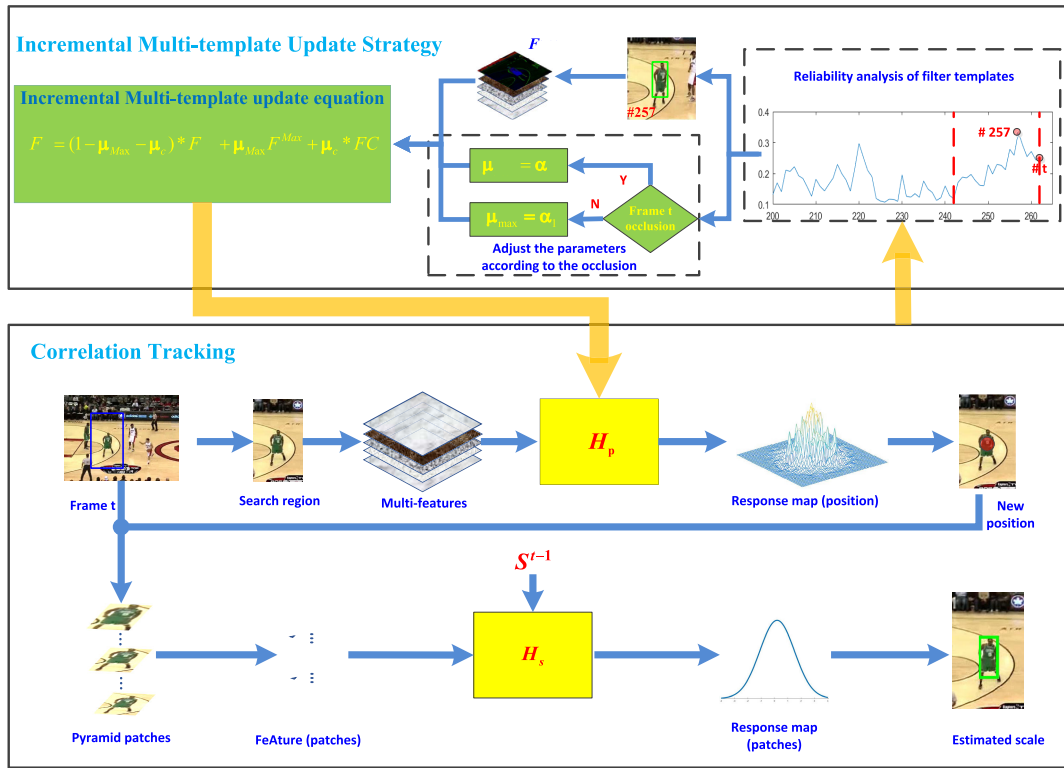
FIGURE 1. Structure of the proposed tracking strategy. This flowchart is divided into two parts. The top is the adaptive multi-template update strategy, including reliability analysis of filter templates, adjusting the parameters according to the occlusion and the adaptive multi-template update equation. The bottom is the baseline correlation filter process.

on the response map. Then, the filter template F^{Max} of the most reliable frame is searched in the range t~t-k-1. We use F^{Max} as the local maximum reliable template in the range of t ~ t-k-1. Concurrently, the possibility of occlusion in the current frame is judged according to the reliability. F^{Max} is fused into the tracking template according to its learning rate μ_{max} at the subsequent template fusion. Among them, the learning rate of F^{Max} is adjusted according to the occlusion condition. The baseline algorithms for this strategy are ECO and ECO-hc. Figure 1 shows the structure of the proposed tracking algorithm.

A. RELIABILITY ANALYSIS OF FILTER TEMPLATES
The formula used to calculate the response map in the discriminative correlation filter (DCF) is:

$$G = F^{t-1} \odot \widehat{w}, \qquad (8)$$

where $\widehat{w}$ is the Fourier transform of w. In Equation (8), w is a feature for extracting the t-th frame according to the search area. F^{t-1} is the tracking template of frame t-1, and F^{t-1} is obtained by fusing the filter template of the historical frame. The tracked object position is determined according to the maximum value of the response map G.

In previous related filtering algorithms, different historical filter templates $FC^1, \cdots, FC^{t-2}$ and FC^{t-1} were not treated differently. In fact, the relative importance of different

historical filter templates obviously varies. The more reliable the filter template is, the more important it is. For example, when the object is affected by the occlusion factor, the current filter template has a smaller benefit on object tracking. However, when the object is affected by occlusion, the reliability of the filter is generally low. Therefore, the weighted peak sidelobe rate (WPSR) is used to measure the reliability of the filter template. The value range of WPSR is [0, 1]. The formula used to calculate WPSR is:

$$WPSR(t) = \max(R_t) \frac{\max(R_t) - mean(\hat{R}_t)}{Std(\hat{R})}, \qquad (9)$$

In Equation (9), R_t is the response map of the t-th frame, $\max(R_t)$ represents its maximum value, and R_t is the response map after removing the center position from the 11×11 pixel range. $mean(\hat{R}_t)$ represents the average of the response, and $Std(\hat{R})$ is the standard deviation of the frame response. The ratio $\frac{\max(R_t) - mean(\hat{R}_t)}{Std(\hat{R})}$ is the traditional peak-to-sidelobe ratio (PSR), and $\max(R_t)$ is equivalent to the weight. Therefore, Equation (9) can be considered to be the weighted peak sidelobe ratio.

Figure 2 shows the reliability of different historical templates for skating2. Because the female dancer in frame 293 is not occluded, the WPSR value reaches a local maximum. The importance of filter templates can be denoted by their reliability. If the value of WPSR(t) is large, then it is indicated
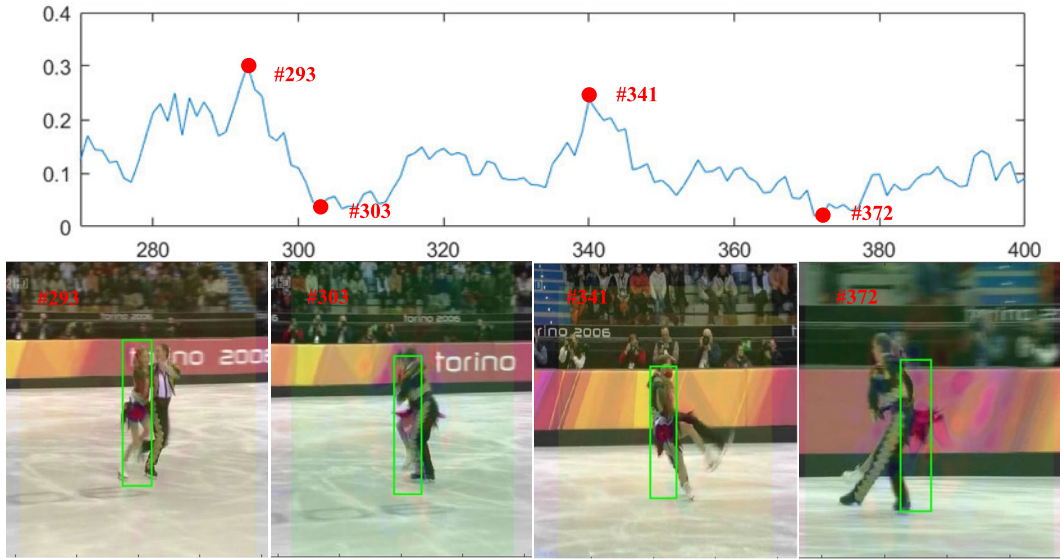
FIGURE 2. Reliability of different historical templates. Frames 293 and 341 clearly show the outline of the object. The objects in frames 303 and 372 are occluded.

that the reliability of filter template is relatively large, and its importance is also relatively large. Conversely, a little WPSR(t) value indicates that the reliability and importance of the filter template is relatively small. Through observation of a large amount of data, it is found that the larger the WPSR(t) is, the more reliable it is in the tracking process.

B. INCREMENTAL MULTI-TEMPLATE UPDATE STRATEGY

In previous correlation-filtering tracking algorithms, some or all the filter templates of historical frames are fused into the tracking template to make better use of the historical information. The template fusion method of previous algorithms is shown in (7). The fusion template in (7) is also a multi-template update. Here, Equation (7) can be expanded into (10):

$$
\begin{aligned}
F^t &= (1 - \mu_c)F^{t-1} + \mu_c FC^t \\
&= (1 - \mu_c)^2 F^{t-2} + (1 - \mu_c)\mu_c FC^{t-1} + \mu_c FC^t \\
&\quad \cdots \cdots \\
&= (1 - \mu_c)^{t-1}F^1 + \cdots (1 - \mu_c)^2 \mu_c FC^{t-2} \\
&\quad \cdots + (1 - \mu_c)\mu_c FC^{t-1} + \mu_c FC^t \\
&= (1 - \mu_c)^{t-1}FC^1 + \mu_c \sum_{k=2}^{t}(1 - \mu_c)^{t-k}FC^k, \quad (10)
\end{aligned}
$$

However, the traditional update method in (10) has obvious shortcomings. This method does not take into account the effectiveness and importance of different historical filtering templates. When the tracking reliability is low, if a filter template with low reliability is still fused into the tracking template in this update method, then the tracking template will be polluted. It is easy to cause tracking failure if the pollution source cannot be eliminated after the tracking template is contaminated.

We propose that updating the template with the local maximum reliability can effectively improve the robustness of the template, especially for complex scenes such as occlusion. This assumption is reasonable because the local highly reliable tracking template is closer to the appearance and characteristics of the real object, and it can resist the negative effects of occlusion. Therefore, a new method is proposed to update the tracking template:

$$ F^t = (1 - \mu_{Max} - \mu_c)F^{t-1} + \mu_{Max}F^{Max} + \mu_c FC^t, \quad (11) $$

This is a method of incremental updates of the template, leading to the incremental fusion of the filter template with the highest local confidence into the tracker template. In (11), F^t is a tracking template for the current frame t. F^{t-1} is a filter template for t-1 frames. F^{Max} is the most reliable filter template in the k-frame range from the t-th frame to frame (t-k-1), that is, the filter template of the frame with the highest WPSR score in this k frame range. FC^t is the filter template for the current frame. μ_{Max} is the learning rate of the filter template F^{Max}. μ_c is the learning rate of the filter template FC^t of the current frame.

To ensure that the fused template will not drift, we have taken the following two measures:

(a) Equation (11) uses an adaptive weighted fusion method. In the process of tracking, we judge whether the target is in a complex environment by calculating the confidence of the current frame. According to the target confidence score, we adaptively adjust the weight of the weighted fusion in (11).

(b) When the incremental template is updated, we choose the filter template with the largest local historical confidence as the increment. This template is closer to the current frame, and its similarity to the current frame is greater than to the historical frame, so its ability to distinguish the target in the current frame is stronger. Additionally, the maximum

confidence in the local historical frame also ensures that the filter template is more reliable.

The physical meaning of this update strategy is that it uses the incremental update template for the local maximum reliability template, which is completely different from the other traditional update templates and has the following advantages:

(a). The incremental update template ensures the richness of the filter template source, which includes not only the historical template F^{t-1} and current template FC^t but also the local maximum reliability template F^{Max}. In fact, the traditional update method in (10) is a special case of our multi-template update strategy.

(b). One only needs to add an extra auxiliary memory space for the local maximum reliability template F^{Max} in our update strategy, which does not cause a significant increase in space complexity and uses less extra memory space.

(c). The incremental update template can enhance the robustness of the filter and overcome template drift to some extent because the filter F^{Max} represents the high reliability filter template.

The traditional template update method fuses only the filter template of the previous frame into the tracker template of the current frame, as shown in (10). If only the filter template with the highest local confidence is merged into the tracker template of the current frame when the template is updated, then the target information will be lost. If the fusion of the filter template with the highest confidence in all the historical frames into the tracker template of the current frame can also improve the performance of the tracker, then it can also be regarded as a special case of our strategy. That is, the template with the highest local confidence is extended to the template with the highest global confidence.

C. THE STRATEGY OF THE ADAPTIVE LENRNING RATE

During the tracking process, if F^{Max} is always fused according to a small learning rate, then the ability of the tracking algorithm to correct errors will be reduced when a template error is encountered. Alternatively, if F^{Max} is always fused according to a large learning rate, then it is easy to cause template drift. To solve this problem, this paper adopts a strategy of adaptively adjusting the learning rate of the local optimal template. This strategy determines the learning rate of F^{Max} according to whether the object is affected by occlusion. In the past, most algorithms used a fixed threshold to determine whether the tracking object was affected by occlusion, which will not be effective when the object's background is complex. Therefore, we propose a new method of discrimination.

In general, when an object is affected by occlusion during object tracking, the reliability of the tracking will decrease sharply. Sometimes, a background change during tracking will also cause the reliability of two adjacent frames to decrease suddenly. To eliminate such interference fluctuations, we determine whether the t-th frame is affected by the object occlusion based on the difference $|WPSR(t) - WPSR(t-2)|$, $|WPSR(t) - WPSR(t-3)|$ of the reliability scores of the current frame t, frame t-2, and frame t-3. If the reliability score $WPSR(t) < \tau_1$ and $|WPSR(t) - WPSR(t-2)| > \tau_2$ of the t-th frame or the reliability score $WPSR(t) < \tau_1$ and $|WPSR(t) - WPSR(t-3)| > \tau_3$ of the t-th frame, then the t-th frame is considered to be affected by object occlusion.

Existing algorithms cannot accurately determine the effect of occlusion on the object. Although our judgment method has high accuracy, it is still far from 100% accurate. Blocked templates can adversely affect tracking, but missing the correct template can easily lead to tracking-template drift. Now, F^{Max} has higher accuracy. This paper considers that the fusion tracking object with an F^{Max} increment can not only ensure the richness of the template but also ensure the accuracy of the tracking template when the object is affected by occlusion. However, if F^{Max} is fused into the trace template every time the template is updated, the template overfitting can easily occur. Therefore, this strategy increases the learning rate of F^{Max} when it is affected by occlusion; otherwise, the F^{Max} learning rate μ_{Max} is decreased. Therefore, the adaptive learning rate can be set

$$\mu_{Max} = \begin{cases} \alpha_0, & \text{if } WPSR(t) < \tau_1 \text{ and } |WPSR(t) \\ & -WPSR(t-2)| > \tau_2 \\ \alpha_0, & \text{if } WPSR(t) < \tau_1 \text{ and } |WPSR(t) \\ & -WPSR(t-3)| > \tau_3 \\ \alpha_1, & \text{others} \end{cases}$$
$$(\text{where } \alpha_0 > \alpha_1), \quad (12)$$

To a great extent, the learning rate μ_{Max} can be adaptive according to the influence of occlusion. The adaptive learning rate can adjust the updating parameters to overcome the occlusion template drift.

V. EXPERIMENT

To accurately analyze the performance of our proposed strategy, we selected the datasets VOT2016 dataset [48], UAV123 [49] and OTB100 dataset [50] in the field of object tracking to test it. The experiments in this paper were all completed with an Intel (R) Core (TM) i7-6700 CPU @ 3.40 GHz running the 64-bit Windows 10 Professional Edition system and MATLAB 2016a.

The VOT2016 dataset consists of 60 videos compiled from a set of more than 300 videos. This article uses the four main evaluation methods in the VOT2016 challenge as quantitative evaluation indicators to evaluate the tracking performance. The four indexes are Expected Average Overlap (EAO), Accuracy (Average Overlap during active tracking), Robustness (failure rate) and Average Overlap (AO). AO is similar to EAO but not exactly the same. EAO will redetect after the object, and the object-prediction boxes do not overlap for 5 frames, while AO does not have such a restart mechanism. The trackers are ranked using the area-under-the-curve (AUC) score in AO. The greater the value of Robustness, the worse the stability in the VOT2016 challenge. The calculation

TABLE 1. Ablation study of the proposed tracker (IMUS) on VOT2016.

	FC^t	$FC^t + F^{Max\_global}$	$FC^t + F^{Max\_local}$
EAO	0.37	0.38	0.38
Accuracy	0.54	0.55	0.55
Robustness	10.83	11.33	10.17
AO	0.44	0.44	0.45

FC^t : The current frame template. F^{Max_global} : Incremental historical of the global maximum confidence template. F^{Max_local} : Incremental historical of the local maximum confidence template

formula of robustness is as follows:

$$\rho_R(i) = \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} F(i,k), \qquad (13)$$

where $F(i,k)$ is the number of failures of the i-th tracker in the k-th repeat. We refer to [48] for details.

UAV123 is a dataset of videos captured by low-altitude drones. UAV123contains 123 video sequences, with frames exceeding 110k. As a dataset of drone photography, UAV123 is dominated by small objects. In this paper, the accuracy plot is the percentage of the number of frames between the actual object position and the predicted object position within different thresholds and the total number of frames. The success plot represents the percentage of the number of frames with the overlapping rate of the tracking result of frame t and the real object within different thresholds of the total number of frames. We refer to [49] for details.

OTB100 is a commonly-used benchmark library in the field of target tracking. OTB100 contains 100 video sequences containing various complex scenes, including Fast Motion, Illumination Variation, Occlusion, etc. The main evaluation indicators of OTB100 are the success rate and precision rate. We refer to [50] for details.

A. IMPLEMENTATION DETAILS

ECO and ECO-HC are our baseline trackers. We used our strategy to improve the benchmark trackers, which were termed the IMUS (Incremental Multi-template Update Strategy) and IMUS-hc (hand-crafted feature version IMUS) tracking algorithms.

In the IMUS, we apply the same feature representation as ECO, a combination of the first (conv-1) and last (conv-5) convolutional layer in the VGG-m [51] network, along with HOG [52] and Color Names (CN) [53]. In this paper, we believe that the range k of the local frame is better within the range of 18 to 20. When testing the VOT2016 dataset, we set k = 18 in the IMUS. The learning rate of the local optimal template F^{Max} in the IMUS was $\alpha_0 = 0.017$, $\alpha_1 = 0$. The learning rate μ_c was set as 0.012 for the IMUS. We normalized the value of WPSR (t). The range of values for WPSR (t) is [0,1]. In Section 3.3 we set the thresholds

for determining occlusion: $\tau_1 = 0.325$, $\tau_2 = 0.125$, $\tau_3 = 0.175$. The other main parameters are the same as for ECO.

In the IMUS-hc, we applied the same feature representation as ECO-hc, along with HOG and Color Names (CN). We found that the value of k of the local frame was better within the range of 18 to 20. When testing the VOT2016 dataset, we set k = 20 in IMUS-hc. The learning rate of the local optimal template F^{Max} in IMUS is $\alpha_0 = 0.018$, $\alpha_1 = 0$. The learning rate $\mu_c = 0.012$ of the current filter template FC^t for the IMUS. In this experiment, the judgment parameters of occlusion are the same as those set in the IMUS. The other main parameters are the same as ECO-hc.

B. INCREMENTAL TEMPLATE ANALYSIS

In this experiment, the choice of the incremental template in the incremental multi-template update strategy has an important impact on the performance of the tracker. In this section, we discuss the choice of the incremental template in the incremental multi-template fusion strategy.

To study the effect of the incremental template on the tracker in the strategy, we selected the global and local optimal templates as the incremental templates. Except for the initial frame, the template with the highest global confidence has higher reliability but is relatively different from the current frame. The template with the highest local confidence is different from that with the highest global confidence. The template with the highest confidence in a small range closer to the current frame has a smaller gap with the current frame and has higher reliability as well.

In Table 1 and Table 2, "FC^t" is the effect of not using the incremental multi-template update strategy. "$FC^t + F^{Max\_global}$" uses the template with the highest confidence among all historical templates except the initial frame as the increment in the incremental multi-template update. "$FC^t + F^{Max\_local}$" is to select the template with the highest confidence among multiple historical templates close to the current frame as the increment in the incremental multi-template update.

It can be observed from Tables I and II that the performance of the tracker is significantly improved after adopting the incremental multi-template fusion strategy. The performance of the tracker is best after the local optimal template is used

TABLE 2. Ablation study of the proposed tracker (IMUS-hc) on VOT2016.

	FC^t	$FC^t + F^{Max\_global}$	$FC^t + F^{Max\_local}$
EAO	0.32	0.32	0.33
Accuracy	0.53	0.54	0.54
Robustness	19.00	16.17	15.50
AO	0.40	0.40	0.41

FC^t : The current frame template. F^{Max_global} : Incremental historical of the global maximum confidence template. F^{Max_local} : Incremental historical of the local maximum confidence template

FIGURE 3. Confidence scores and visualization of selected tracking results. The fluctuation in reliability of frames 60-63 is caused by background changes. The sudden change in the reliability of frames 80-83 is caused by the occlusion of the object.

as the incremental template in the incremental multi-template fusion strategy. The template with the maximum global confidence can improve the performance of the tracker to a certain extent by virtue of its high reliability. The template with the largest local confidence is more similar to the target of the current frame, so it improves the performance of the tracker more significantly. From the above ablation experiment, we can observe that our strategy has a more obvious lifting effect on the tracker using the hand-crafted feature.

C. VOT DATASET EVALUATION

Here, we analyze the effect of our strategy on the VOT2016 dataset. Table 3 shows the performance of our trackers compared to the most advanced trackers.

Compared with the baseline tracker, our tracker is significantly better. Table 3 shows that the EAO score of the IMUS is 0.38. Compared with the baseline tracker (ECO), our strategy achieves a relative gain of 2.7% in EAO. The

accuracy score of the IMUS is 0.55. Compared with the baseline tracker (ECO), our strategy achieved a 1.9% relative gain in accuracy. The IMUS's Robustness score was 10.17. Compared to the baseline tracker (ECO), our strategy has significantly reduced robustness. The IMUS has an AO score of 0.45, which has a clear advantage over the baseline tracker (ECO). Table 3 shows that the EAO score of the IMUS-hc is 0.33. Compared with the baseline tracker (ECO-hc), our strategy achieves a relative gain of 3.1% in EAO. The accuracy score of the IMUS-hc is 0.54. Compared with the baseline tracker (ECO-hc), our strategy achieved a 1.9% relative gain in accuracy. TIMUS-hc's robustness score was 10.17. The robustness of our strategy is significantly improved compared with that of the baseline tracker (ECO-hc). Our strategy achieves a relative robustness gain of 2.7%. The IMUS-hc achieves an AO score of 0.41, which has a clear advantage over the baseline tracker (ECO-hc).

TABLE 3. Comparison of the IMUS, IMUS-hc and state-of-the-art trackers in VOT2016.

	Experiment baseline			Experiment unsupervised
	EAO	Accuracy	Robustness	AO
IMUS	**0.38**	**0.55**	**10.17**	**0.45**
IMUS-hc	0.33	0.54	15.50	0.41
ECO	0.37	0.54	10.83	0.44
ECO-hc	0.32	0.53	19.00	0.40
UDT_plus	0.30	0.52	20.48	0.44
UDT	0.23	0.53	24.83	0.37
SiamFC	0.28	**0.55**	20.50	0.42
DNT	0.28	0.50	16.43	0.43
HCF	0.22	0.44	20.50	0.37
Staple	0.30	**0.55**	21.33	0.39
SRDCF	0.25	0.48	24.33	0.40
DSST	0.18	0.50	36.33	0.33

Compared with state-of-the-art trackers other than the baseline tracker, our tracker has an absolute advantage. These state-of-the-art trackers include: UDT_plus [54], UDT [54], SiamFC [55], DNT [56], HCF [57], Staple [23], SRDCF [22], DSST [25]. As shown in Table 3, among the deep feature trackers that are compared, UDT_plus provides the best results, with an EAO score of 0.30 in the experiment baseline. The EAO value of the tracker of the IMUS using this strategy is improved by 26.7% compared to UDT_plus. Among the compared trackers using hand-crafted features, Staple provides the best results, with an EAO score of 25% in the experiment baseline. The EAO of the tracker of the IMUS-hc with the strategy proposed in this paper is 3% higher than that of Staple. Among the deep feature trackers compared, DNT provides the best results with an AO score of 43% in unsupervised experiments. The AO value of the IMUS is 2% higher than that of DNT. Among the compared trackers using hand-crafted features, SRDCF provides the best results, with an AO score of 40% in unsupervised experiments. The AO of the IMUS-hc is 1% higher than that of SRDDF.

Figure 4 shows the expected overlap scores for the baseline in the VOT2016 dataset in unsupervised experiments. As shown in Figure 4, the expected overlap scores of the IMUS are greater than those of other state-of-the-art trackers. The IMUS-hc has good performance as well.

Figure 5 shows the accuracy and robustness scores of the IMUS, IMUS-hc, and eight state-of-the-art algorithms affected by occlusion in the experiment baseline. As shown in Figure 5, the IMUS and IMUS-hc are more robust than the other algorithms. Figure 6 shows the order of the failures after different factors influence the experiment baseline. Obviously, the ability of the IMUS and IMUS-hc to address occlusion effects exceeds that of other algorithms.

Figure 7 shows examples of unmanned VOT2016 results, which indicate that the IMUS and IMUS-hc algorithms are better than the other state-of-the-art trackers. Because our
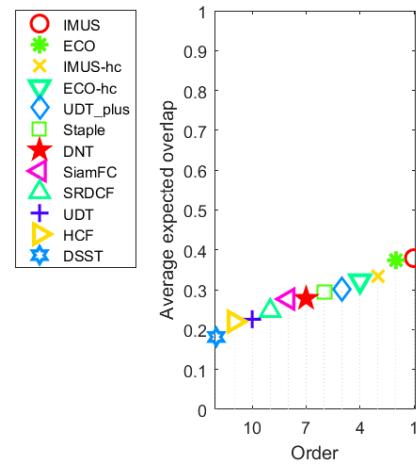

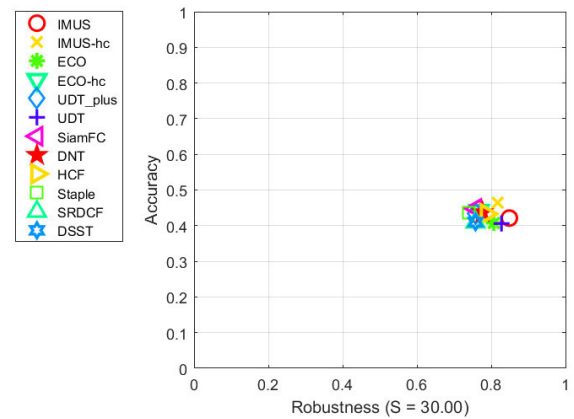**FIGURE 4.** Expected overlap scores for baseline (experiment baseline).

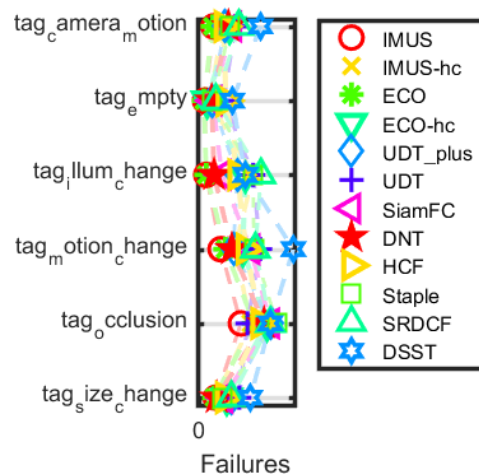**FIGURE 5.** AR plot for tag_occlusion in experiment baseline.

FIGURE 6. Orderings of failures in the experiment baseline.

strategy uses the local maximum reliability filter to update the template, it can overcomes the template drift to a great extent. Therefore, the IMUS and IMUS-hc algorithms can
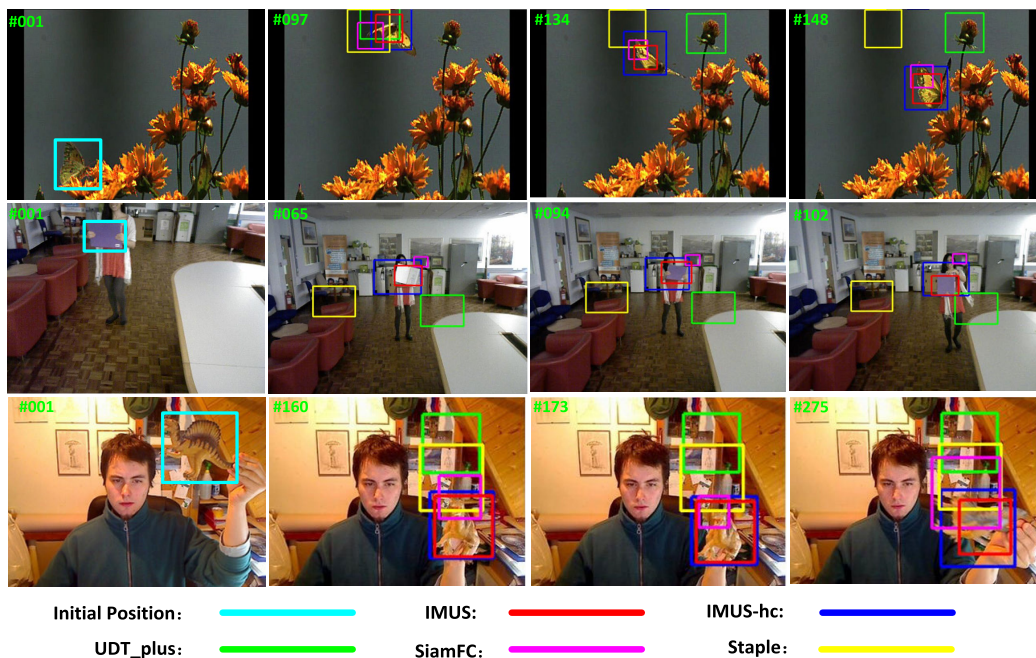
FIGURE 7. Examples of the VOT2016 results. The first, second, third lines show the butterfly, book, and dinosaur image sequences. It is obvious that our trackers perform better than the other state-of-the-art trackers.
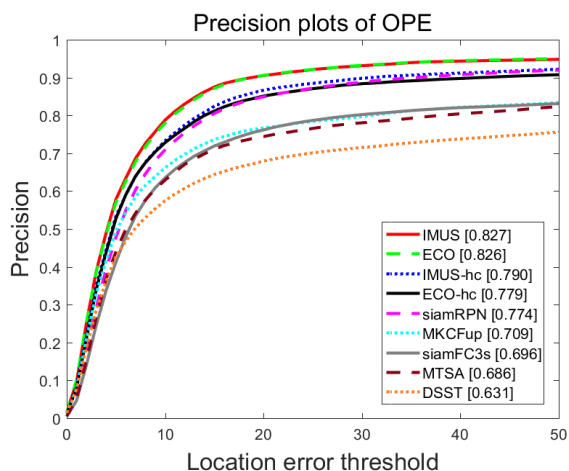

FIGURE 8. Precision plots of the OTB100. Comparison of state-of-the-art trackers with the improved trackers using our strategy.

FIGURE 9. Success plots of the OTB100. Comparison of state-of-the-art trackers with the improved trackers using our strategy.

achieve better tracking performances than other state-of-the-art trackers.

D. OTB DATASET EVALUATION

We compare our IMUS and IMUS-hc trackers with state-of-the-art trackers on the OTB100. The state-of-the-art trackers include ECO, ECO-hc, siamRPN [58], MKCFup [43], siamFC3s [55], MTSA [46], DSST [25].

From Figures 8 and 9, we can observe that our IMUS tracker and IMUS-hc tracker have excellent precision performance. Between them, our strategy improves the IMUS-hc tracker with the hand-crafted feature more obviously. Compared with ECO-hc, our IMUS-hc tracker improved precision by 1.1%, and its success rate greatly improved as well.
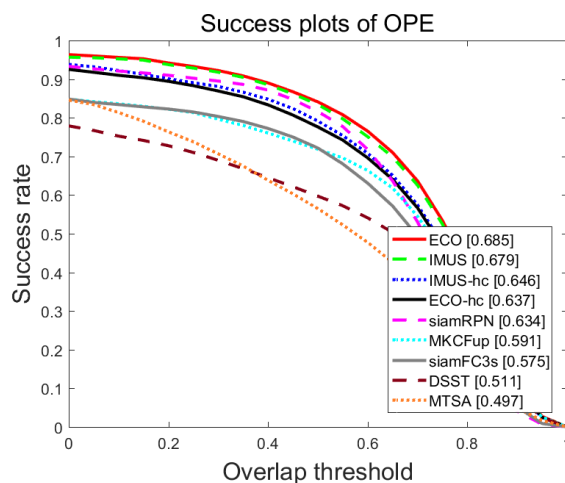
Compared with siamRPN, our tracker IMUS-hc increased precision by 1.6% and the success rate by 1.2%. Compared with MKCFup, our IMUS-hc tracker improved precision by 8.1% and success by 5.5%. Compared with MMTS, our IMUS-hc tracker improved precision by 10.4% and success by 14.0%. From the experimental results, we observe that our strategy improves the performance of the tracker (IMUS) using the hand-crafted feature more significantly.

E. UAV123 DATASET EVALUATION

Here, we analyze the effect of our strategy on the UAV123 [48] dataset.We selected 10 state-of-the-art trackers from the results provided by UAV123 to compare
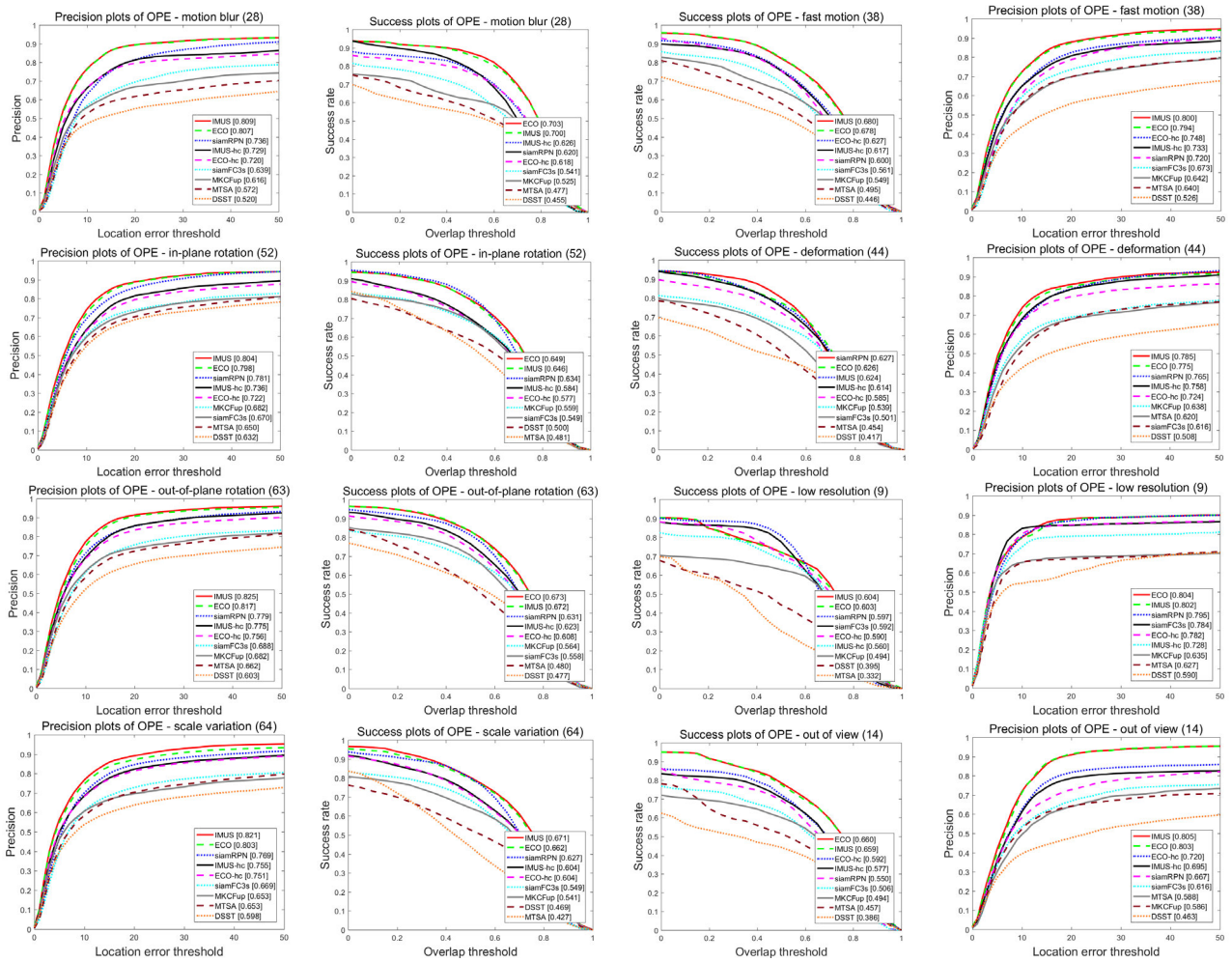
FIGURE 10. Precision and success plots of the proposed tracker compared with state-of-the-art approaches on different attributes of OTB100.
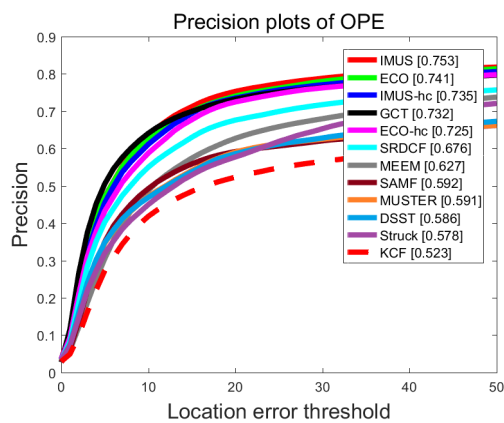

FIGURE 11. Precision plots on the UAV-123. Comparison of the state-of-the-art trackers with the improved trackers using our strategy.
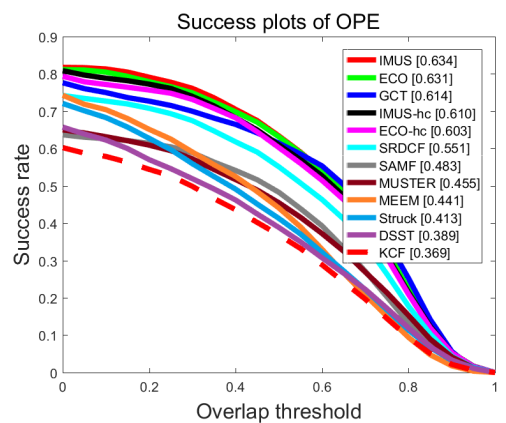

FIGURE 12. Success plots on the UAV-123. Comparison of the state-of-the-art trackers with the improved trackers using our strategy.

with our trackers. The 10 trackers were: GCT [59], ECO, ECO-hc, SRDCF [22], SAMF [60], MUSTER [61], MEEM [62], Struck [63], DSST [25], KCF [27].

As shown in Figures 11 and 12, the precision score of the baseline tracker improved significantly after our policy was

applied to it. The IMUS and IMUS-hc scores are significantly higher than those of the baseline trackers ECO and ECO-hc. As shown in Figure 11, the precision scores of IMUS and IMUS-hc are 0.753 and 0.735, respectively. The precision
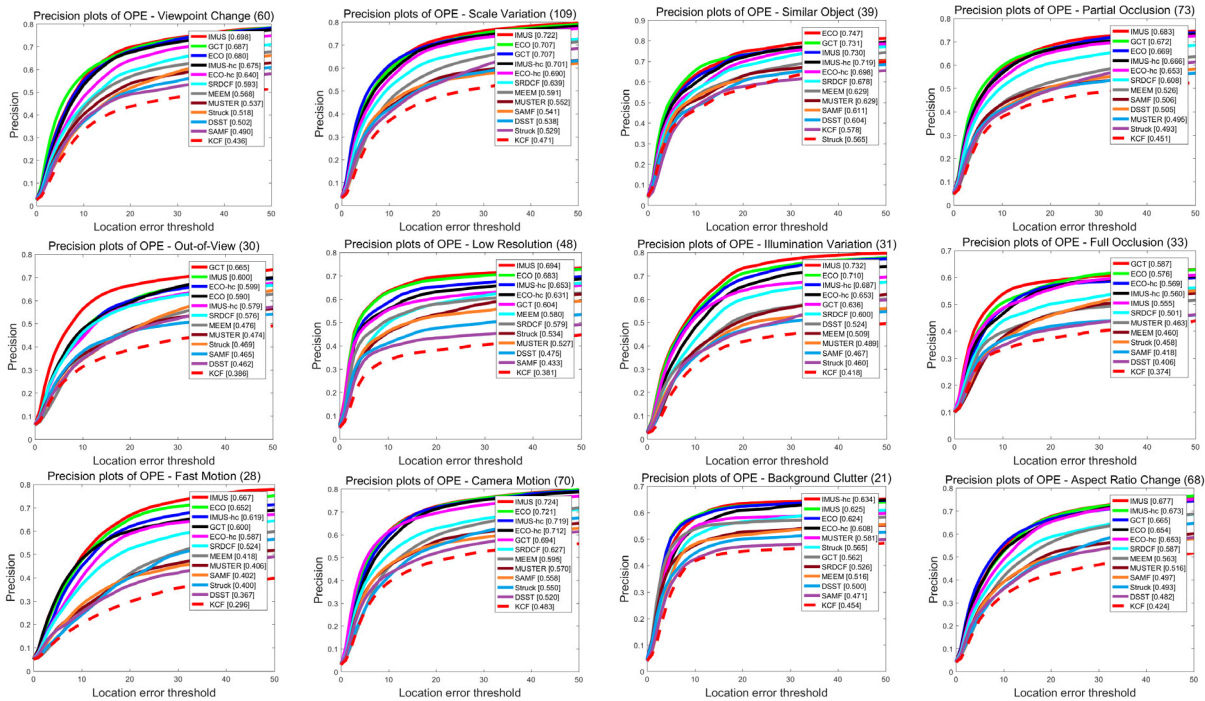
FIGURE 13. Precision plots of the proposed tracker compared with state-of-the-art approaches on different attributes of UAV123.
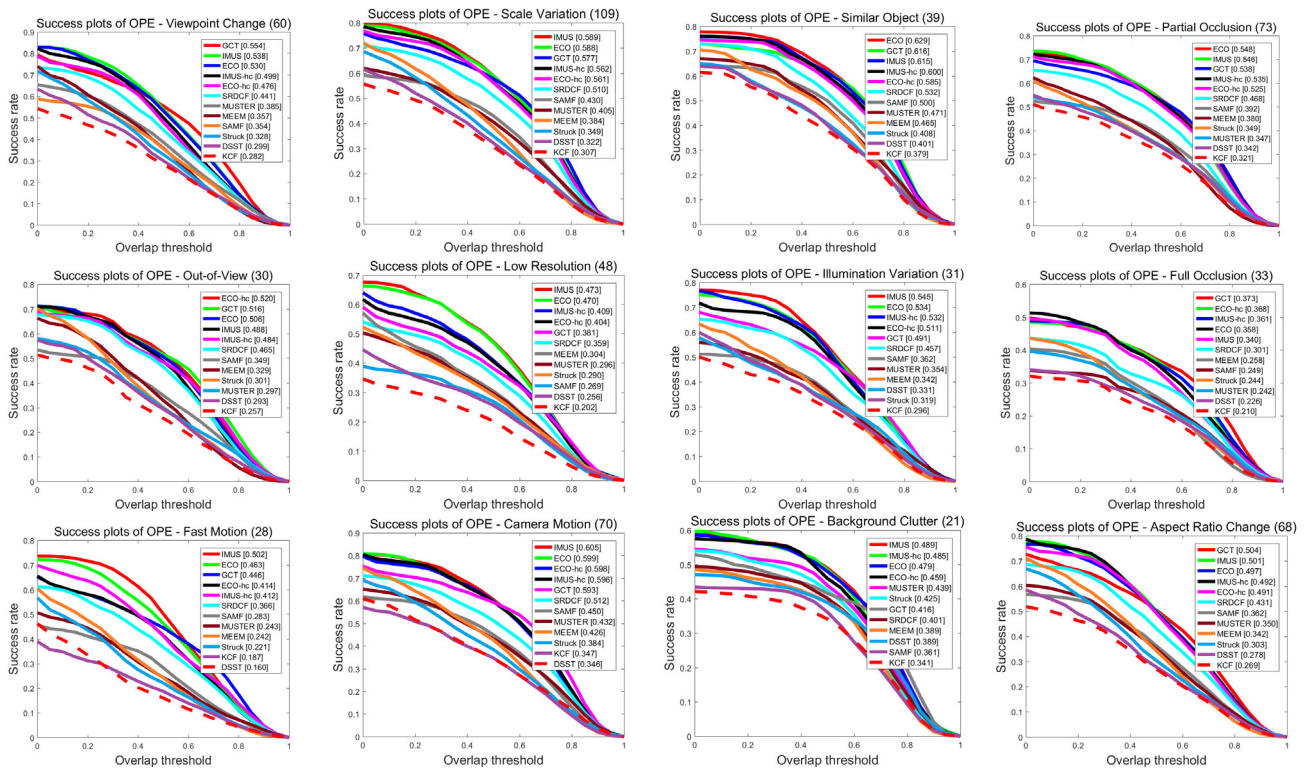

FIGURE 14. Success plots of the proposed tracker compared with state-of-the-art approaches on different attributes of UAV123.

score of the best performing GCT among the eight state-of-the-art trackers, except the baseline tracker, is 0.732. It can be clearly observed that GCT 's precision score has a gap with the IMUS and IMUS-hc. As shown in Figure 12,

the success scores of the IMUS and IMUS-hc are 0.634 and 0.610, respectively. Therefore, there is little difference in performance between GCT and the IMUS-hc. However, the success scores of the IMUS have a clear advantage over

those of GCT. Figure 13 show precision plots of the UAV123 dataset, respectively. We find that the IMUS and IMUS-hc algorithms have advantages in some challenging attributes, such as background clutter, camera motion, fast motion, partial occlusion and illumination variation. In the success plots of Figure 14, the IMUS and IMUS-hc algorithms also achieved good performance in some challenging attributes, such as background clutter, camera motion, partial occlusion, low resolution and illumination variation.

F. DISCUSSION

Our tracker is effective in complex environments. In the correlation filter tracker, the quality of the template determines the tracking effect. In complex environments, the template's impact on the tracker's performance is more pronounced. The template of the related tracker is constantly updated during the tracking process. Therefore, when the target is in a complex environment, it is inevitable to fuse part of the environmental information into the template. Therefore, we fuse the incremental template of the frame with accurate local tracking and little influence of background information into the tracking template. In this way, the effect of complex environments on tracking is effectively eliminated. For example, as shown in Figures 11 and 12, when the target moves rapidly, the background is cluttered, the viewing angle changes, etc., our tracker is significantly better than other state-of-the-art trackers. Of course, in individual cases, our tracking is slightly worse than other advanced trackers, for example, when the goal is beyond the field of vision. We believe that the main reason for this finding is the selectable range of the incremental template, as we showed in the ablation experiment. In different scenes, the targets and backgrounds change at different speeds. In practical applications, the selectable range of the incremental template can be adjusted according to the speed of change of the target and background in the scene. The range we choose in this article can guarantee superior overall performance.

Qualitative and quantitative experiments on the VOT2016, OTB100 and UAV123 dataset show that our strategy more obviously improves the effect of trackers using the hand-crafted feature. We believe that this improvement is due to the high robustness of the depth feature itself, so the performance improvement of the tracker using the depth feature is not particularly obvious. The robustness of the hand-crafted feature is poor, so our strategy improves the effects of trackers using the hand-crafted feature more significantly.

VI. CONCLUSION

We propose a novel incremental multi-Mtemplate update strategy that incrementally fuses the local maximum reliability template into the update of the tracking filter template. This incremental template-update method can enhance the robustness of the tracker and overcome template drift. This strategy also improves the accuracy of object tracking by adjusting the learning rate of the local maximum-reliability template adaptively according to the occlusion state of the

object. Qualitative and quantitative experiments on the VOT, OTB and UAV datasets demonstrate that the novel tracking strategy performs favorably against other state-of-the-art trackers in terms of success rate and precision plots.

REFERENCES

[1] L. Men and X. Yang, "A survey of object tracking algorithms," *Acta Automatica.*, vol. 4, no. 7, pp. 1244–1260, 2019.

[2] A. I. TChikkamagalur, "A survey on moving object detection and tracking techniques," *Int. J. Eng. Comput. Sci.*, vol. 15, pp. 2970–2979, Apr. 2016.

[3] S. Ojha and S. Sakhare, "Image processing techniques for object tracking in video surveillance—A survey," in *Proc. Int. Conf. Pervas. Comput. (ICPC)*, Pune, India, Jan. 2015, pp. 161–166.

[4] Q. Yuan, L. H. Liu, and S. Lu, "Visual object tracking with adaptive structural convolutional network," *Knowl. Based Syst.*, vol. 194, no. 07, pp. 1105–1116, 2020.

[5] Y. Liang, "Multi-task driven feature models for thermal infrared tracking," in *Proc. Assoc. Advance Artif. Intell.*, Feb. 2020, pp. 1036–1045.

[6] D. N. Yuan Fan and Z. He, "Learning target-focusing convolutional regression model for visual object tracking," *Knowl. Based Syst.*, vol. 194, no. 5, pp. 1055–1064, 2020.

[7] S. V. Kothiya and K. B. Mistree, "Notice of removal: A review or real time object tracking in video sequences," in *Proc. Int. Conf. Electr., Electron., Signals, Commun. Optim. (EESCO)*, Visakhapatnam, India, Jan. 2015, pp. 1–9.

[8] R. C. Joshi, M. Joshi, A. G. Singh, and S. Mathur, "Object detection, classification and tracking methods for video surveillance: A review," in *Proc. 4th Int. Conf. Comput. Commun. Autom. (ICCCA)*, Pune, India, Dec. 2018, pp. 1–7.

[9] S. R. Balaji and S. Karthikeyan, "A survey on moving object tracking using image processing," in *Proc. 11th Int. Conf. Intell. Syst. Control (ISCO)*, Pune, India, Jan. 2017, pp. 469–474.

[10] A. S. Montero, J. Lang, and R. Laganiere, "Scalable kernel correlation filter with sparse feature integration," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop (ICCVW)*, Santiago, Chile, Dec. 2015, pp. 587–594.

[11] C. Sun, D. Wang, H. Lu, and M.-H. Yang, "Learning spatial-aware regressions for visual tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake, NV, USA Jun. 2018, pp. 8962–8970.

[12] Z. Liu, Z. Lian, and Y. Li, "A novel adaptive kernel correlation filter tracker with multiple feature integration," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Zurich, Switzerland, Sep. 2017, pp. 254–265.

[13] M. Danelljan, G. Hager, 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.

[14] J. Zhang, Y. Liu, S. Zhang, R. Poppe, and M. Wang, "Light field saliency detection with deep convolutional networks," *IEEE Trans. Image Process.*, vol. 29, no. 1, pp. 4421–4434, Dec. 2020.

[15] F. Zhao, Z. Chen, I. Rekik, S.-W. Lee, and D. Shen, "Diagnosis of autism spectrum disorder using central-moment features from low- and high-order dynamic resting-state functional connectivity networks," *Frontiers Neurosci.*, vol. 14, pp. 233–256, Apr. 2020.

[16] Z. G. A. Hao and Y. Li, "Robust object tracking with scene-adaptive scheme in occlusion," *J. Intell. Fuzzy Syst.*, vol. 36, no. 6, pp. 1–9, 2018.

[17] J. Li, G. Li, and H. Fan, "Image dehazing using residual-based deep CNN," *IEEE Access*, vol. 6, pp. 26831–26842, 2018.

[18] H. K. Galoogahi, A. Fagg, and S. Lucey, "Learning background-aware correlation filters for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1135–1143.

[19] A. Lukezic, T. Vojir, L. C. Zajc, J. Matas, and M. Kristan, "Discriminative correlation filter with channel and spatial reliability," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Venice, Italy, Jul. 2017, pp. 6309–6318.

[20] F. Li, C. Tian, W. Zuo, L. Zhang, and M.-H. Yang, "Learning spatial-temporal regularized correlation filters for visual tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake, LA, USA, Jun. 2018, pp. 4904–4913.
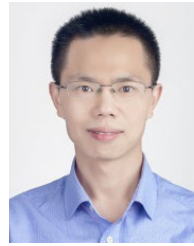
[21] J. Batista, "Exploiting the Circulant Structure of Tracking-by-detection with Kernels," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Florence, Italy, Aug. 2012, pp. 702–715.

[22] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 4310–4318.

[23] 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. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 1401–1409.

[24] R. Khan and M. Felsberg, "Beyond correlation filters: Learning continuous convolution operators for visual tracking," in *Proc. Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands, Aug. 2016, pp. 321–330.

[25] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," in *Proc. Brit. Mach. Vis. Conf.*, Nottingham, England, 2014, pp. 267–277.

[26] D. 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.*, Santiago, Chile, Jun. 2010, pp. 2544–2550.

[27] 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.

[28] M. Danelljan, G. Bhat, F. S. Khan, and M. Felsberg, "ECO: Efficient convolution operators for tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 981–992.

[29] Y. Zhu, J. Wen, L. Zhang, and Y. Wang, "Visual tracking with dynamic model update and results fusion," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Athens, Greece, Oct. 2018, pp. 2685–2689.

[30] D. Wang, W. Fang, W. Chen, and T. Sun, "Model update strategies about object tracking: A state of the art review," *Electronics*, vol. 8, no. 11, pp. 1207–1216, 2019.

[31] X. Tian, "Unifying temporal context and multi-feature with update-pacing Framework for visual tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 6, pp. 1078–1091, Oct. 2019.

[32] Y. Gu, X. Niu, and Y. Qiao, "Robust visual tracking via adaptive occlusion detection," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, Brighton, MA, USA, May 2019, pp. 2242–2246.

[33] H. Wang, X. Zhang, L. Yu, and X. Wang, "Research on mean shift tracking algorithm based on significant features and template updates," in *Proc. IEEE Int. Conf. Mechatronics Autom. (ICMA)*, Changchun, China, Aug. 2018, pp. 168–176.

[34] D. Moujahid, O. Elharrouss, and H. Tairi, "Visual object tracking via the local soft cosine similarity," *Pattern Recognit. Lett.*, vol. 110, pp. 79–85, Jul. 2018.

[35] Y. Qi, S. Zhang, L. Qin, Q. Huang, H. Yao, J. Lim, and M.-H. Yang, "Hedging deep features for visual tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 5, pp. 1116–1130, May 2019.

[36] X. Niu and Y. Qiao, "Context-based occlusion detection for robust visual tracking," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Beijing, China, Sep. 2017, pp. 3655–3659.

[37] N. Wang, W. Zhou, and H. Li, "Reliable re-detection for long-term tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 3, pp. 730–743, Mar. 2019.

[38] D. Feng, F. Gao, X. Wang, G. Wang, and H. Dai, "Cascaded iterative training model and parallel multi-classifiers for visual object tracking," *IEEE Access*, vol. 7, pp. 63099–63112, 2019.

[39] M. Wang, Y. Liu, and Z. Huang, "Large margin object tracking with circulant feature maps," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulum, China, Jul. 2017, pp. 4021–4029.

[40] H. Hu, B. Ma, J. Shen, and L. Shao, "Manifold regularized correlation object tracking," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1786–1795, May 2018.

[41] M. Mueller, N. Smith, and B. Ghanem, "Context-aware correlation filter tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 1396–1404.

[42] M. Tang and J. Feng, "Multi-kernel correlation filter for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 3038–3046.

[43] M. Tang, B. Yu, F. Zhang, and J. Wang, "High-speed tracking with multi-kernel correlation filters," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake, NV, USA, Jun. 2018, pp. 4874–4883.

[44] 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, no. 19, pp. 27271–27290, Oct. 2019.

[45] X. Li, Q. Liu, Z. He, H. Wang, C. Zhang, and W.-S. Chen, "A multi-view model for visual tracking via correlation filters," *Knowl.-Based Syst.*, vol. 113, pp. 88–99, Dec. 2016.

[46] A. Bibi and B. Ghanem, "Multi-template scale-adaptive kernelized correlation filters," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop (ICCVW)*, Santiago, Chile, Dec. 2015, pp. 613–620.

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

[48] M. Kristan, A. Leonardis, J. Matas, M. Felsberg, and R. Pflugfelder, "The visual object tracking vOT2016 challenge results," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Aug. 2016, pp. 777–823.

[49] M. Mueller, N. Smith, and B. Ghanem, "A benchmark and simulator for uav tracking," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Aug. 2016, pp. 445–461.

[50] 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.

[51] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," in *Proc. Brit. Mach. Vis. Conf.*, Nottingham, U.K., 2014, pp. 387–398.

[52] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, San Diego, CA, USA, Apr. 2005, pp. 118–127.

[53] J. van de Weijer, C. Schmid, J. Verbeek, and D. Larlus, "Learning color names for real-world applications," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1512–1523, Jul. 2009.

[54] N. Wang, Y. Song, C. Ma, W. Zhou, W. Liu, and H. Li, "Unsupervised deep tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 1308–1317.

[55] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. S. Torr, "Fully-convolutional siamese networks for object tracking," in *Proc. Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands, 2016, pp. 850–865.

[56] Z. Chi, H. Li, H. Lu, and M.-H. Yang, "Dual deep network for visual tracking," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 2005–2015, Apr. 2017.

[57] C. Ma, J.-B. Huang, X. Yang, and M.-H. Yang, "Hierarchical convolutional features for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3074–3082.

[58] B. Li, J. Yan, W. Wu, Z. Zhu, and X. Hu, "High performance visual tracking with siamese region proposal network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake, NV, USA Jun. 2018, pp. 8971–8980.

[59] J. Gao, T. Zhang, and C. Xu, "Graph convolutional tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 4649–4659.

[60] M. Kristan, R. Pflugfelder, A. Leonardis, J. Matas, and L. ehovin, "The visual object tracking Vot2014 challenge results," in *Proc. Eur. Conf. Comput. Vision(ECCV)*, Zurich, Switzerland, Sep. 2014, pp. 191–217.

[61] X. Hong, "Multi-store tracker (muster): A cognitive psychology inspired approach to object tracking," in *Proc. Comput. Vis. Pattern Recognit.(CVPR)*, Boston, MA, USA, Jun. 2015, pp. 749–758.

[62] J. S. Zhang Ma and S. Sclaroff, "MEEM: Robust tracking via multiple experts using entropy minimization," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Zurich, Switzerland, Sep. 2014, pp. 188–203.

[63] S. Hare, A. Saffari, and P. H. S. Torr, "Struck: Structured output tracking with kernels," in *Proc. Int. Conf. Comput. Vis.*, Barcelona, Spain, Nov. 2011, pp. 263–270.

QINGSONG XIE received the M.S. degree in computer software from Shandong University, Jinan, China, in 1996. He is currently a Professor with the School of Computer Science and Technology, Shandong Technology and Business University, Yantai, China. His research interests include algorithmic analysis and design, machine learning, and image processing.

KEWEI LIU received the bachelor's degree in engineering from Liaocheng University, in 2018. He is currently pursuing the degree with Shandong Technology and Business University. His research interests include visual tracking and data mining.

AN ZHIYONG received the M.S. degree in computer science from the Taiyuan University of Technology, in 2004, and the Ph.D. degree in computer science from Xidian University, in 2008. He is currently an Associate Professor with Shandong Technology and Business University, Yantai, China. His research interests include data visualization and information processing.

LEI WANG received the M.S. degree in information economy from the Dongbei University of Finance and Economics, Dalian, China, in 2008. He is currently an Engineer with the Information Construction Office, Shandong Technology and Business University, Yantai, China. His current research interests include visual tracking, object detection, person re-ID, and hadoop.

YE LI received the M.S. degree from the Ocean University of China, Qingdao, China, in 2015. He is currently an Assistant with the College of Computer Science and Engineering, Weifang University of Science and Technology, Weifang, China. His current research interests include machine learning and computer vision.

ZHONGLIANG XIANG received the Bachelor of Science (B.S.) degree in computer science from Mudanjiang Normal University, in 2003, the Master of Science degree in computer science from the Ocean University of China, in 2010, and the Ph.D. degree in computer science from Dongseo University, in 2015. He is currently an Associate Professor with the Weifang University of Science and Technology, China. His research interests include machine learning, especially in nonparametric Bayesian models and reinforcement learning.

• • •