Received 28 February 2023, accepted 26 March 2023, date of publication 28 March 2023, date of current version 5 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3262726

RESEARCH ARTICLE

Research on Background Learning Correlation Filtering Algorithm With Multi-Feature Fusion

HONGGE REN[1], LEIGANG XING[2], AND TAO SHI[3]
[1]School of Control and Mechanical Engineering, Tianjin Chengjian University, Tianjin 300384, China
[2]College of Electrical Engineering, North China University of Science and Technology, Tangshan 063210, China
[3]School of Electrical Engineering and Automation, Tianjin University of Technology, Tianjin 300384, China

Corresponding author: Tao Shi (st99@email.tjut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61203343, and in part by the Natural Science Foundation of Hebei Province under Grant F2018209289.

ABSTRACT Aiming at the problems of occlusion, drift, and background change in visual image tracking, a background learning correlation filtering algorithm based on multi-feature fusion is proposed. In the framework of correlation filtering, multi-feature fusion, multi-template update, and background learning regularization are used to improve the performance of the filter in the problem of template contamination and object occlusion. The fast directional gradient histogram (FHOG), color feature (CN), and texture feature (ULBP) were extracted, and the feature channels were connected in series. Then the depth features of Conv4-4 and Conv5-4 layers were extracted through the VGG-19 network, and the appearance model of the target was constructed. To reduce the sensitivity of the filter to the sudden change of background, a background learning filter is constructed, and the alternate direction multiplier method (ADMM) is used to speed up the calculation of the filter. In the model update stage, aiming at the problem of pollution of the original template caused by target occlusion, a high-confidence multi-template fusion update strategy is proposed by fusing the template with the highest confidence in the current frame, the previous frame, and the history frame. Finally, the proposed algorithm is tested on OTB50, OTB100, UAV123, and TC128 experimental data sets, and some classical and latest algorithms. The experimental results show that the tracking accuracy and robustness of the correlation filtering algorithm are improved.

INDEX TERMS Target tracking, machine vision, correlation filtering, background learning, more templates.

I. INTRODUCTION

As a basic research work of machine vision, visual image tracking algorithm has always been one of the hot topics in the field of computer vision and machine learning. It has a wide range of application prospects in military guidance [1], human-computer interaction [2], virtual reality [3], and many other aspects. Visual image tracking algorithm models the appearance and motion information of the target by using the features and background information of the video or image sequence, to predict the target motion state and obtain the final position of the target. Visual image tracking algorithms can be divided into two categories from the perspective of model construction: generative algorithms [4] and discriminative algorithms [5]. The generative model algorithm does not consider the background information and achieves the purpose of tracking by matching the established model with the target category, such as the algorithm based on sparse representation and the algorithm based on subspace. The discriminant model transforms the tracking problem into a decision problem between the tracking target and background. With the rapid development of vector machine models [6] and deep learning, the discriminant model has become the main tracking algorithm research model.

Video image tracking technology has made a breakthrough in recent years. However, the undirected motion of the target makes the appearance model of the target and the target background change, and the occlusion between the target and the background makes the target tracking task more difficult. Currently, multiple challenging attributes commonly

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

encountered in target tracking are illumination variation (IV), scale variation (SV), background clutter (BC), out-of-plane rotation (OPR), in-plane rotation (IPR), motion blur (MB), occlusion (OCC), distortion (DEF), out-of-field of view (OV), fast motion (FM), and low resolution (LR). To better overcome these interferences in the process of video image tracking, this paper adds a background learning factor to the algorithm for optimization and proposes a background learning correlation filter.

With the introduction of the artificial neural network model [7], [8], [9], [10], some scholars have optimized and improved the neural network model, and the proposed tracking algorithm has been greatly improved in accuracy and speed. In 2016, Valmadre et al. proposed SiamFC [11], which trains a Siamese network to locate a sample image in a large search area to solve the similarity learning problem. The algorithm calculates the cross-correlation of two inputs through the bilinear layer to achieve a dense and efficient sliding window evaluation. Although the tracking speed of the SiamFC algorithm is fast, the accuracy is lower than some tracking algorithms of related filtering frameworks. In the same year, Tao et al. first proposed the twin network target tracking algorithm SINT [12]. SINT uses a deep learning network to learn a matching function that makes the trained model more generalizable. After training, the model does not need to be updated online or adjusted to any parameters. Although the SINT algorithm reached an advanced level at that time, it was time-consuming to train the model. In 2022, Chen et al. [13] proposed a stable object-tracking algorithm that introduces a spatial weight function to strengthen the feature response recognition between the object and other objects. The algorithm restricts the size of the filter coefficients and has a good effect on overcoming the target occlusion problem. Based on the improvement of the correlation filtering algorithm [14], by extracting the target feature information into the filter for training, it can not only meet the real-time tracking, but the tracking speed is also very superior.

In 2010, Bolme et al. proposed the MOSSE algorithm [15], which is the first application of the correlation filtering framework to the field of target tracking, where the maximum response value is obtained by the correlation operation of the filter. The MOSSE algorithm can maintain better tracking and is faster, reaching up to 669 FPS. Henriques et al. proposed the CSK algorithm [16] with a cyclic structure by adding a regular term to the MOSSE algorithm. The CSK algorithm is simple in principle to further improve the tracking speed of the algorithm, but the window is fixed and the resistance to occlusion is weak. In 2015, Caseiro et al. proposed the KCF algorithm [17] by adding a multi-channel histogram of oriented gradients for feature extraction in the stage of object appearance model construction. The research improvement based on the KCF algorithm is still the research field of many scholars.

In 2014, Danelljan et al. proposed the DSST algorithm [18], which proposed the multi-feature fusion mechanism for the first time, and trained the scale filter and the position filter for scale estimation and object localization respectively. However, the tracking efficiency of the algorithm is low, and the speed is 25FPS on VOT competition. Then, Hager et al. proposed the FDSST algorithm [19]. The FDSST algorithm is accelerated by two techniques of feature downscaling and interpolation. With the research of deep learning, many scholars use deep convolutional layers to extract feature information of the target and add it to the relevant filtering framework. Ma et al. proposed the HCF algorithm [20], which is an improved algorithm of the KCF algorithm, to extract the depth features of Conv3-4, Conv4-4, and Conv5-4 layers by VGG-19 [21] network to replace the original manual features. The depth features used by the HCF algorithm improve the tracking accuracy of the algorithm, but the tracker cannot accurately track the target when it is occluded for a long time.

There are two main types of tracking algorithms. The tracking algorithm based on deep learning has high accuracy, but the tracking speed is slow because of the huge training set of the algorithm. The tracking algorithm based on the discriminative correlation filter framework has strong real-time performance, but the tracking accuracy needs to be improved under the interference of target occlusion and background mutation. In the proposed algorithm, feature fusion and background learning regularization are introduced to improve the utilization of target appearance information and background information. The multi-template update strategy is used to improve the accuracy of the algorithm in the case of occlusion challenges.

Our main contributions are as follows:

1) The appearance information of the target is extracted jointly using three manual features and a deep feature-weighted fusion of two layers of the VGG-19 network to achieve complementary feature information.
2) The use of background learning correlation filters effectively learns background information and enhances the generalizability of the filters when the background changes.
3) Update the target template using a high-confidence multi-template update strategy to overcome the tracking drift problem caused by contaminated templates.

II. RELATE WORKS

The correlation filtering algorithm has strong real-time performance and stable tracking. We use CF as the basic framework of the tracking algorithm. Our method can learn to track background information well. In the following, we will briefly review some workers related to algorithm research.

A. MULTI-FEATURE FUSION

The tracking algorithm uses the method of multi-feature fusion to extract features, which can effectively avoid the singleness of feature information. In 2020, Li et al. [22] dynamically fused two hand-crafted features in a correlation filtering framework, adjusting the weights according to the consistency of the filter prediction results to overcome large changes in the target model in challenging scenarios. In 2021, Zhao et al. [23] proposed an adaptive feature fusion mechanism and applied it to convolutional neural networks to enhance the tracking and discrimination ability of different convolutional layers. In the same year, Elayaperumal and Joo [24] utilized a multi-feature fusion strategy to enhance the target appearance model in unconstrained environments, which combined contextual information and spatial background variations to further improve the accuracy and robustness of the tracker. In 2022, Ke et al. proposed the STCDFF algorithm [25]. STCDFF assigns different weights to features by their discrimination ability and representation ability to the tracked target and makes full use of the advantages of different features to deal with complex appearance changes and background mutation challenges.

B. CORRELATION FILTER TRACKER

The correlation filter has achieved good results as a discriminative tracking method in the field of visual tracking due to its high computational efficiency. There are many ways to improve the correlation filters, such as adding contextual information [26], adding scale filters, and adding constraint terms. In 2017, Galoogahi et al. proposed the BACF algorithm [27], which expands the region sampled by the cyclic matrix in the correlation filter to crop out more valid samples, allowing the filter to learn more background information and thus improve the tracking accuracy. Yuan et al. then improved on BACF and proposed the TRBACF algorithm [28]. The TRBACF algorithm uses a temporal normalization strategy to improve tracking accuracy and robustness through an efficient model that adapts to abrupt changes in the tracking scene. In 2019, Dai et al. proposed the ASRCF [29]. The ASRCF algorithm uses an adaptive mechanism to improve the weight and spatial constraint matrix of the correlation filter, and uses an optimization algorithm to accelerate the filter so that the algorithm can obtain better tracking results.

Adding different regular terms [30], [31] to the correlation filter to construct different loss functions, and then minimizing the sum of these terms by optimization methods, can make the tracker have different performance. In this paper, a background learning correlation filter is proposed by introducing a background learning factor and adding a background learning regularization term to the objective function. Experimental results show that the tracking performance of the filter is improved under the background interference problem.

C. OBJECT MODEL UPDATE

The accurate update of the object templates is the key to overcoming the problem of object scale [32] change and object occlusion [33], [34], [35]. Traditional correlation filtering algorithms use online updating. When the target is occluded, the target template will be contaminated, and the traditional correlation filtering algorithm will extract the feature information from the contaminated template for model updating, which will seriously affect the filter's ability to discriminate the tracking target and lead to tracking failure. In 2017, Wang et al. proposed the LMCF algorithm [36], which uses a high-confidence multimodal forward strategy to detect whether the target template needs to be updated. When the output result of the current frame exceeds the set value, the tracking result is considered correct and the updated model is selected. The LMCF algorithm well solves the problem of template contamination caused by target drift and target occlusion.

In 2020, Wu [37] proposed a tracking algorithm based on a multi-template update strategy by adding a tracker to improve the robustness of the algorithm. Xu et al. proposed LADCF [38]. LADCF uses the mask multimodal detection method to extract more effective target information according to edge information and then compares multiple peaks with the first frame to overcome the interference of background changes and achieve accurate tracking. In 2022, Wang et al. [39] added long-term memory and short-term memory in the subsequent template stage and proposed a long short-term memory update template strategy with a stable tracking effect.

III. ALGORITHM DESIGN

In this section, we will introduce our tracker in detail. The overall framework is shown in Figure 1.

A. ALGORITHM DESCRIPTION

According to the series of theoretical derivation and formula analysis, the proposed tracking algorithm flow is shown in Figure 1:

First, initialize the constructed background learning-related filters and tracking objects to start the input first video frame. Then FHOG features [40], CN features and texture features are extracted from the search region of the input current frame, and then depth features of Conv4-4 and Conv5-4 layers are extracted to jointly construct the appearance model of the target. The filter performs the correlation operation, and the ADMM method is used to optimize the filter, speed up the calculation of the correlation response to obtain the final response score and predict the target position of the next frame as the maximum. After obtaining the predicted target position, in the target model update phase, the template information is updated using a high-confidence multi-template fusion strategy, and after the update is completed, the operation continues for the next frame until all video sequences are completed, ending the tracking of the target.

B. APPEARANCE MODEL CONSTRUCTION

The FHOG feature is an improved feature of HOG. The FHOG features can describe the target edge information in detail and perform well for local shape information and have good robustness to changes in light intensity. CN features describe the color information of the target region of the image with good stability to the geometric deformation of the target, and the input dimension is orthogonally

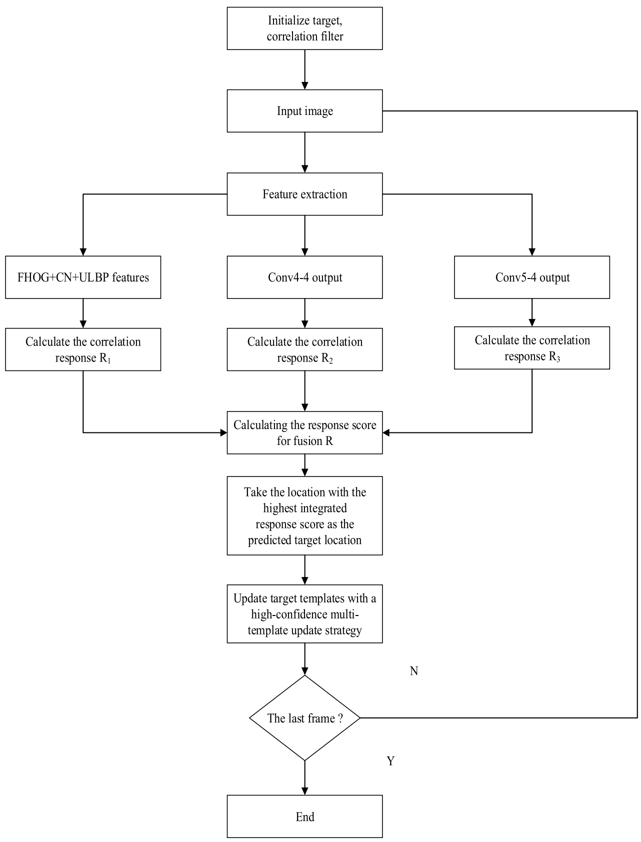


FIGURE 1. **The tracking algorithm flow.**

transformed with principal component analysis (PCA) to reduce from 11 to 2 dimensions to improve the operation speed. The texture features are extracted by the ULBP method, which is the equivalent pattern of LBP, and the number of LBP features is further reduced by record hopping, and the number of patterns is reduced from $2p$ to $p(p-1)+2$, where P represents the number of sampling points in the domain set.

The FHOG features, CN features, and ULBP features [41] are concatenated in feature channels. The total dimension is 34 dimensions, containing 31-dimensional FHOG features, 2-dimensional CN features, and ULBP features that form the 34th dimension. Deep feature extraction is performed using the VGG-19 network. In the VGG-19 deep convolutional network model, each convolutional layer extracts features with different expressiveness to the target. The bottom convolutional features contain more texture and edge information, and the higher-level convolutional features contain more advanced semantic information. To complement the texture information expressed by the manual features and make the algorithm more discriminative to the target, this paper chooses to extract the depth features of Conv4-4 and Conv5-4 layers. The ability of the different convolutional layers of the VGG-19 network to represent the target is shown in Figure 2:

In this paper, the algorithm uses semantic information contained in-depth features (Conv4-4, Conv5-4) combined

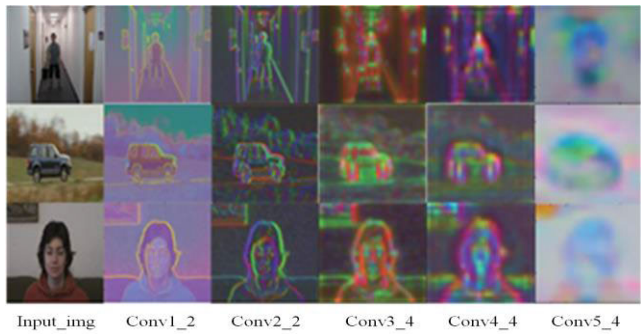


FIGURE 2. **Visualization of the output of the convolutional layer of the VGG-19 network.**

with texture information of manual features (FHOG, CN, ULBP) to achieve effective complementation of texture information and high-level semantic information, train the relevant filters separately, build a powerful target appearance model, and optimize tracking speed while improving tracking accuracy. First extract the FHOG features, CN features, and ULBP features of the target search region for feature channel concatenation, noted as R_1. Then the depth features of the Conv4-4 and Conv5-4 layers are extracted using the VGG-19 network, denoted as R_2, R_3. Learning is performed by filters to obtain multiple response values, and the response values are weighted and fused, and noted as R. The weighting factor formula is:

$$R = H_1R_1 + H_2R_2 + H_3R_3 \tag{1}$$

In the formula, H_1 is the weight of manual features, H_2 is the weight of Conv4-4 features, and H_3 is the weight of Conv5-4 features. After repeated experiments, $H_1 = 0.7$, $H_2 = 1$, $H_3 = 0.5$ were taken.

C. BACKGROUND LEARNING FILTER OPTIMIZATION

For the problem that targets tracking changes in the background, resulting in the tracking drift phenomenon. In this paper, we add a background learning regularization term to the filter inspired by the BACF algorithm to achieve effective learning of the tracking background by the filter and prevent the tracking target from failing due to sudden changes in the background. The objective function constructed in this paper is as follows:

$$\begin{aligned} L(h) = \frac{1}{2}\left\| \sum_{e=1}^{E} Bx^e * h^e - y \right\|_2^2 + \frac{\lambda}{2}\sum_{e=1}^{E}\left\|h^e\right\|_2^2 \\ + \frac{\alpha}{2}\left\| \sum_{e=1}^{E} B(x_p^e - x_{p,t-1}^e) * h^e \right\|_2^2 \end{aligned} \tag{2}$$

In (2), E indicates the total number of channels, $x^e \in R^N$ indicates the feature information extracted from the e th channel, $x_p^e \in R^N$ is the environmental information extracted from the end channel. $B \in R^{M\times N}$ denotes a binary matrix that obtains the middle element of x. $*$ indicates the relevant operation, λ is the regularization parameter. α is the environmental learning factor, which controls the adaptation of the filter to the environment. $t-1$ indicates the previous frame of the video. The third term is the joined background learning regularization term.

Since video continuous tracking is continuous in time, it can be assumed that the background changes are similar in the short term. The proposed filter can maintain robustness to background variations in the latter frame of video tracking by filtering for effective learning of background differences $x_p - x_{p-1}$. $g^d = B^T h^d \in R^N$ is the introduced auxiliary variable, $x_\varepsilon = x_p - x_{p,t-1}$ indicates background differences. For more efficient computation, the objective function is put into the Fourier domain for processing, and the Fourier form of the objective function is:

$$\hat{L}(h,\hat{g}) = \frac{1}{2N}\left\|\sum_{e=1}^{E}\hat{x}^e\odot\hat{g}^e-\hat{y}\right\|_2^2+\frac{\lambda}{2}\sum_{e=1}^{E}\left\|h^e\right\|_2^2 + \frac{\alpha}{2N}\left\|\sum_{e=1}^{E}\hat{x}_\varepsilon^e\odot\hat{g}^e\right\|_2^2 \tag{3}$$

In equation (3), $\odot$ indicates Hadamard product, Λ denotes Fourier transform.

The augmented Lagrangian form of equation (3) is:

$$\begin{aligned}\hat{L}(h,\hat{g},\hat{\zeta}) &= \frac{1}{2N}\left\|\sum_{e=1}^{E}\hat{x}^e\odot\hat{g}^e-\hat{y}\right\|_2^2 \\ &+ \frac{\lambda}{2}\sum_{e=1}^{E}\left\|h^e\right\|_2^2+\frac{\alpha}{2N}\left\|\sum_{e=1}^{E}\hat{x}_\varepsilon^e\odot\hat{g}^e\right\|_2^2 \\ &+ \frac{\mu}{2}\sum_{e=1}^{E}\left\|\hat{g}^e-\sqrt{N}FB^Th^e\right\|_2^2 \\ &+ \sum_{e=1}^{E}(\hat{g}^e-\sqrt{N}FB^Th^e)^T\hat{\zeta}^e\end{aligned} \tag{4}$$

In equation (4), $\hat{\zeta} = \left[\hat{\zeta}^1,\ldots,\hat{\zeta}^E\right] \in R^{N\times E}$ is the Lagrangian vector, μ is the penalty factor.

D. ADMM METHOD TO ACCELERATE THE SOLUTION

Solve the subproblem h^e: Using the ADMM method [42] to transform the previous formulation into 2 subproblems. h^e_{i+1} of the equation function, as follows:

$$\begin{aligned}h^e_{i+1} &= arg\min_{h^e}\frac{\lambda}{2}\left\|h^e\right\|_2^2+\frac{\mu}{2}\left\|\hat{g}-\sqrt{N}FB^Th^e\right\|_2^2 \\ &\times(\hat{g}^e-\sqrt{N}FB^Th^e)^T\hat{\zeta}^e\end{aligned} \tag{5}$$

In equation (5), The iterative solution of $\hat{\zeta}_{i+1}$ is:

$$\hat{\zeta}_{i+1} = \hat{\zeta}_i + \mu(\hat{g}_{i+1}+\hat{h}_{i+1}) \tag{6}$$

In equation (6), i indicates iteration.

Due to the sparsity of h^e, given g and ζ, The optimal solution for this h can be derived as:

$$h^e = (\zeta^e+\mu g^e)/(\lambda/N+\mu) \tag{7}$$

The equation of $\hat{g}_{i+1}$ can be expressed as:

$$\begin{aligned}\hat{g}_{i+1} &= arg\min_{g}\frac{1}{2N}\left\|\sum_{e=1}^{E}\hat{x}^e\odot\hat{g}^e-\hat{y}\right\|_2^2 \\ &+\frac{\alpha}{2N}\left\|\sum_{e=1}^{E}\hat{x}_\varepsilon^e\odot\hat{g}^e\right\|_2^2 \\ &+\frac{\mu}{2}\sum_{e=1}^{E}\left\|\hat{g}^d-\sqrt{N}FB^Th^d\right\|_2^2 \\ &+\sum_{e=1}^{E}(\hat{g}^d-\sqrt{N}FB^Th^d)^T\hat{\zeta}^d\end{aligned} \tag{8}$$

Solve the subproblem $\hat{g}$. In the presence of other variables h and ζ, optimize the solution of the subproblem $\hat{g}$ by decomposing the subproblem $\hat{g}$ into N smaller problems, as follows:

$$\begin{aligned}\hat{g}(n)^* &= \frac{1}{2N}\left\|\hat{x}(n)^T\hat{g}(n)-\hat{y}(n)\right\|_2^2+\frac{\alpha}{2N}\left\|\hat{x}_\varepsilon(n)^T\hat{g}(n)\right\|_2^2 \\ &+(\hat{g}(n)-\hat{h}(n))^T\hat{\zeta}(n)+\frac{\mu}{2}\left\|\hat{g}(n)-\hat{h}(n)\right\|_2^2\end{aligned} \tag{9}$$

To speed up the solution, specify $\hat{x}$ and $\hat{x}_\varepsilon$ as $\hat{x}_0$ and $\hat{x}_1$. The optimal solution for s is written as:

$$\begin{aligned}\hat{g}(n)^* &= \frac{1}{\mu N}(\hat{x}(n)\hat{y}(n)-N\hat{\zeta}_f+\mu N\hat{h}(n))-\frac{\sum_{k=0}^{1}v_k\hat{x}_k(n)}{\mu\theta} \\ &\times(\frac{1}{N}\eta\hat{y}(n)-\sum_{k}^{1}\hat{x}_k(n)^T\hat{\zeta}(n)+\mu\sum_{k=0}^{1}\hat{x}_k(n)^T\hat{h}(n))\end{aligned} \tag{10}$$

In equation (10), $\theta = \mu N + \sum_{k=0}^{1}v_k\hat{x}_k(n)^T\hat{x}_k(n)$, $\hat{v}_0 = 1$, $v_1 = \varepsilon$, $\eta = \sum_{k=0}^{1}v_k\hat{x}_k(n)^T\hat{x}_k(n)$.

In summary, the objective function and the associated filter can be derived.

E. APPEARANCE MODEL UPDATE STRATEGY

When tracking a target, the current template will be contaminated if the target is obscured or if the target drifts. The accuracy of target tracking can be seriously affected after the filter learns the feature information on the contaminated template. In this paper, we propose a high-confidence multi-template fusion strategy to update the target template. Firstly, APCE (average peak-to-correlation energy) [43] is introduced to calculate the confidence level of the template and judge the reliability of the tracking results. First, average peak-correlation energy (APCE) is introduced to calculate the confidence level of the template. The expressions are:

$$APCE = \frac{|F_{\max}-F_{\min}|^2}{mean(\sum_{w,h}(F_{w,h}-F_{\min})^2)} \tag{11}$$

In equation (11), $F_{\max}$ is the maximum value of the response matrix. $F_{\min}$ is the minimum value of the response matrix. $F_{w,h}$ denotes the value of the element in column w and row h of the response matrix.

A higher confidence level of the template means that the target features in the template are more distinct and not affected by occlusion or target drift. The high-confidence multi-template fusion strategy fuses the highest confidence tracking templates from the current frame template, the previous frame template and the historical frames during the template update process. Specific expressions, as follows:

$$F = (1 - \omega_{\max} - \omega_c)F_{t-1} + \omega_{\max}F_{\max} + \omega_c F_t \qquad (12)$$

In (12), $\omega_{\max}$ is the learning rate of the highest confidence tracking template, $F_{\max}$ is the filter template with the highest confidence in the history frame. This high-confidence multi-template fusion update strategy maximizes the retention of historical information on the templates and ensures the diversity and correctness of the target feature sources. Since the tracking template with the highest confidence in the historical frames is introduced, it can avoid the effect of template contamination caused by the introduction of background information and overcome target occlusion and target drift.

IV. EXPERIMENTS

A. IMPLEMENTATION DETAILS

In this paper, we propose a multi-feature fusion background learning correlation filter tracking algorithm, which uses three manual features and two layers of depth features for fusion to construct the appearance model of the target and extract richer target features. The background learning filter is trained and the ADMM method is used to speed up the filter computation, and then the target template is updated using a high-confidence multi-template fusion strategy to overcome target occlusion, target drift, and background mutation. To verify the effectiveness of the tracking algorithm, it is compared with MOSSE_CA [44], KCF, NSAMF [45], SAMF_CA [46], and FSC2F [47]. The experimental platform was simulated in Matlab R2018b with an Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz 1.99GHz.

The performance of the six algorithms is tested on four experimental datasets using OTB50 [48], OTB100 [49], TC128 [50], and UAV123 [51], which contain multiple challenges that may be encountered in target tracking. Algorithm performance is evaluated using tracking success rate and algorithm accuracy. The algorithm proposed in this paper is represented by LGCF. The learning rate ω used in this paper's algorithm is set to 0.0207, the environmental learning factor α is set to 0.2, and the number of ADMM iterations is set to 2 to speed up the filter solution.

B. RESULTS ON OTB50

The OTB50 dataset contains 49 video sequences with 11 different tracking challenge attributes, as shown in Table 1. Thirty of these video sequences were selected, containing all the challenge attributes. The training results of the test algorithm on the OTB50 dataset are shown in Figure 3.

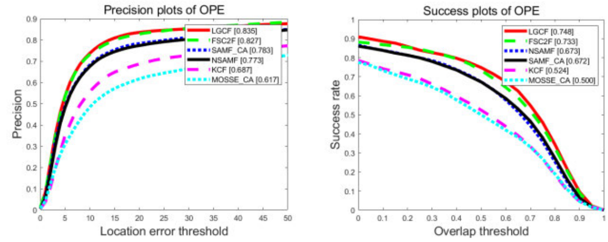


FIGURE 3. LGCF and five trackers on OTB50 success rate graph and accuracy graph.

Where the horizontal coordinates of both the accuracy and success rate plots are their given thresholds and the vertical coordinates are percentages. The accuracy graph reflects the percentage curve of video frames whose center error is less than a given threshold, reflecting the tracking accuracy of the algorithm comparison. The accuracy graph reflects the percentage curve of video frames whose center error is less than a given threshold, reflecting the tracking accuracy of the algorithm comparison.

The accuracy graph shows that the accuracy of the algorithm proposed in this paper is 0.835, which is higher than the other five algorithms compared, indicating the superiority of this algorithm. Compared with the recent new algorithm SAMF_CA, the accuracy of this algorithm is improved by 5.2% and the success rate is improved by 7.6%, which shows the high performance of the proposed algorithm in this paper. The LGCF algorithm runs on the OTB50 dataset with an accuracy of 0.835 and a success rate of 0.748, which is much higher than the benchmark algorithm MOSSE_CA, proving that the optimization of this algorithm is effective in feature extraction and filter regularization.

To show the results of each algorithm running on the 0TB50 dataset more clearly, a table was carried out, as shown in Table 2. The ARCF [52] algorithm is also added to it to compare with various algorithms. The leftmost column of the table shows the name of the algorithm, and the second and third columns show the accuracy and success rates of the algorithm running on the OTB50 dataset. The proposed algorithm in this paper shows a 3.1% improvement in accuracy and a 3.5% improvement in success rate compared with the joined comparison algorithm ARCF, highlighting the high accuracy and success rate of the LGCF algorithm.

C. RESULTS ON OTB100

The OTB100 dataset is the most commonly used evaluation training dataset for existing algorithms. The OTB100 dataset has 98 test videos and is often used to calculate benchmarks. The challenge attributes included are comprehensive, with 38 sequences containing the illumination change (IV) attribute. 64 sequences containing the scale change (SV) attribute, 49 sequences containing the target occluded (OCC) attribute, and 44 sequences containing the target deformation (DEF) attribute. 29 sequences containing the fast motion (FM) attribute, 51 sequences containing the in-plane

TABLE 1. Part of challenging video sequences with 11 challenges.

Sequence	OCC	SV	MB	IV	DEF	BC	FM	IPR	OPR	LR	OV
Basketball	√			√	√	√			√		
Biker	√	√	√				√		√	√	√
Trellis		√		√		√		√	√		
Walking2	√	√								√	
Singer2				√	√	√		√	√		
Jumping			√				√				
Human4	√	√		√	√						
Dudek	√	√			√	√	√	√	√		√
ClifBar	√	√	√			√	√	√			√
Car1		√	√	√		√	√			√	
Subway	√				√	√					
Panda	√	√			√			√	√	√	√

TABLE 2. Performance comparison between LGCF and five tracking algorithms on OTB50. means bigger is better.

Tracker	OTB50	
	Precision ↑	Success ↑
LGCF	0.835	0.748
FSC2F	0.827	0.733
SAMF_CA	0.783	0.672
NSAMF	0.773	0.673
KCF	0.687	0.524
MOSSE_CA	0.617	0.500
ARCF	0.804	0.713

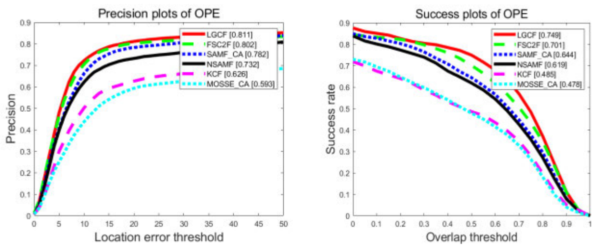


FIGURE 4. LGCF and five trackers on OTB100 success rate graph and accuracy graph.

rotation (IPR) attribute, and 63 sequences containing the out-of-plane rotation (OPR) attribute. 14 sequences containing the background complex (BC) attribute, and 9 sequences containing the low resolution (LR) attribute.

In this paper, 40 test video sequences from the OTB100 dataset are selected for algorithm training, which contains various challenge contexts and can better demonstrate the evaluation performance of each algorithm, and the final algorithm training results are shown in Figure 4:

It can be seen that the LGCF algorithm proposed in this paper improves the accuracy rate by 0.9% and the success

TABLE 3. Performance comparison between LGCF and five tracking algorithms on OTB100. means bigger is better.

Tracker	OTB100	
	Precision↑	Success ↑
LGCF	0.811	0749
FSC2F	0.802	0.701
SAMF_CA	0.782	0.644
NSAMF	0.732	0.619
KCF	0.626	0.485
MOSSE_CA	0.593	0.478
ARCF	0.807	0.732
ECO	0.798	0.725

rate by 4.8% over the second-place FSC2F algorithm in the OTB100 dataset training. It shows that the LGCF algorithm is effective in optimizing the tracking effect. To more clearly represent the training results of the algorithm in the OTB100 data, we plotted Table 3. The ARCF algorithm and the ECO algorithm [53] with good tracking performance in recent years was added to perform a comparison of the test algorithms. In the comparison of the results of our proposed LGCF algorithm and the comparative algorithms NSAMF, FSC2F, SAMF_CA, KCF, and MOSSE_CA selected in this paper.

The LGCF algorithm ranks first in terms of success rate and accuracy. Compared with the ARCF tracking algorithm we added, the accuracy rate improved by 0.4%, and the success rate improved by 1.7%. A comparison with the ECO algorithm showed a 1.3% increase in accuracy and a 2.4% increase in success rate. The results show that the LGCF algorithm tracking is accurate and has a high success rate. The optimization of background learning filters and the proposed high-confidence multi-template fusion strategy perform well for overcoming background interference, target occlusion,

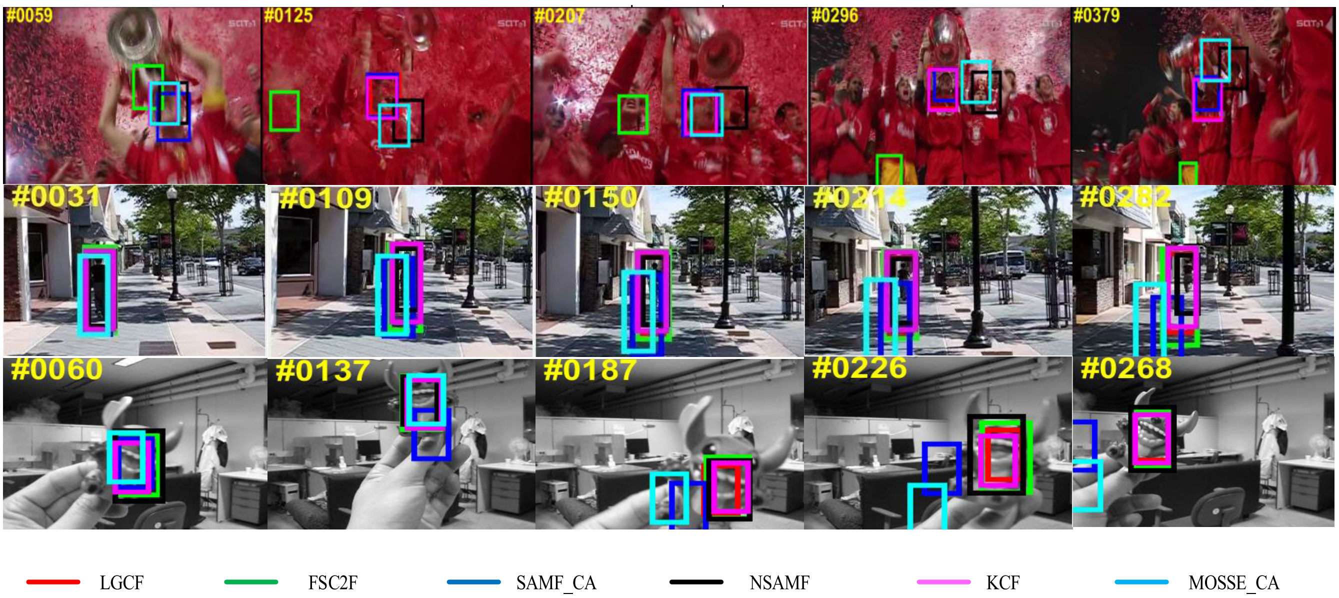


FIGURE 5. (a) soccer sequence (b) Human9 sequence (c) Toy sequence. Compare LGCF and five test algorithms on OTB100.

and target drift problems, which bring tracking inaccuracy and tracking failure.

In this paper, three video challenge sequences, Human9, soccer, and Toy, were selected in the OTB100 dataset, which contained all the challenge attributes. The tracking performance of the proposed algorithm under complex challenge attributes is analyzed by comparing the tracking effect of the LGCF algorithm, NSAMF algorithm, FSC2F algorithm, SAMF_CA algorithm, KCF algorithm, and MOSSE_CA algorithm on these three video challenge sequences, as shown in Figure 5.

In Figure 5(a), the tracked target is an athlete with a trophy in his hand jumping continuously, and the blocking, deformation, and fast movement of the tracked target are reflected. As the target jumps in frame 125, the FSC2F and NASMF algorithms track the offset, while the LGCF algorithm can accurately track the target in the search area. In frame 379, the target jumps as the scene changes. The FS2CF, MOSSE_CA, and NSAMF have failed to track and the tracking frame is well off target. The LGCF algorithm proposed in this paper is stable and accurate in tracking video, which reflects the superiority of the algorithm.

In Figure 5(b), a tracking object is a man and the video sequence contains fast-moving and target-drifting challenge attributes. In 109 frames, the LGCF algorithm and the five comparison algorithms were able to track the target relatively consistently. In 282 frames, with the drift of the target, MOSSE_CA, SAMF_CA, and FSC2F algorithms fail to track, while the rest of the algorithms track accurately, and the LGCF algorithm proposed in this paper has superior performance compared to MOSSE_CA, SAMF_CA and FSC2F algorithms in the target drift challenge problem.

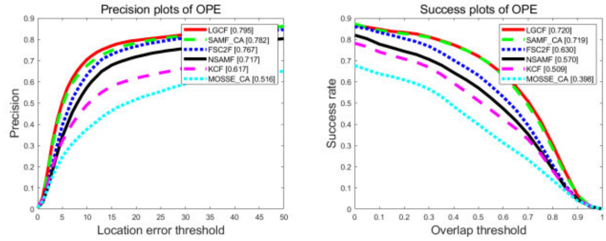


FIGURE 6. LGCF and five trackers on UAV123 success rate graph and accuracy graph.

In Figure 5(c), in frame 137, with the rapid movement of the doll, the SAMF_CA algorithm tracking fails and the other algorithms' tracking is ready to stabilize. By 226 frames, the SAMF_CA and MOSSE_CA algorithms were already tracking far off by the time the doll's occlusion challenge, rotation challenge, and fast-moving challenge were underway. LGCF algorithm has always tracked accurately, LGCF algorithm extracts richer target appearance features and can track the object rotation, and displacement after multiple template update strategies, constantly adding accurate tracking templates makes the algorithm can track accurately, and tracking performance is far more than the other five algorithms.

D. RESULTS ON UAV123

The UAV123 dataset, containing 91 video sequences, has a total number of tracks of 123. The UAV123 dataset was filmed using professional drones, and each video sequence is long and contains a rich set of challenging attributes.

The performance of the algorithm was tested using the LGCF algorithm with NSAMF algorithm, FSC2F algorithm, SAMF_CA algorithm, KCF algorithm, and MOSSE_CA

TABLE 4. Performance comparison between LGCF and five tracking algorithms on UAV123. means bigger is better.

Tracker	UAU123	
	Precision ↑	Success ↑
LGCF	0.795	0.720
SAMF_CA	0.782	0.719
FSC2F	0.767	0.630
NSAMF	0.717	0.570
KCF	0.617	0.509
MOSSE_CA	0.516	0.398
ECO	0.775	0.698

algorithm on the UAV123 dataset by selecting 30 of the video test sequences. The six algorithms of this chapter work on this UAV123 dataset for target tracking, and the test results are obtained as shown in Figure 6:

Due to the complex challenge attributes of the UAV123 dataset and the variety of shooting angles, the tracking accuracy and success rate of the tested algorithms were low. The test results are shown by the accuracy rate graph and the success rate graph, which show that the LGCF algorithm outperforms all the remaining five tested algorithms. The success rate is 0.720 and the accuracy is 0.795. The performance far exceeds that of the FSC2F algorithm, with a 2.8% increase in accuracy and a 9.0% increase in success rate, reflecting the high accuracy and stability of the LGCF algorithm in tracking targets.

To better compare the tracking performance of the algorithms, the statistics are listed in Table 4. Add the ECO algorithm and LGCF algorithm, NSAMF algorithm, FSC2F algorithm, SAMF_CA, KCF algorithm, and MOSSE_CA for comparison.

The LGCF algorithm proposed in this paper improves the accuracy rate by 2.0% and the success rate by 2.2% over the joined ECO algorithm. The ECO algorithm is a relatively new algorithm with more stable algorithm tracking performance in recent years, and the improvement of the accuracy rate and success rate of the LGCF algorithm reflects the good performance of the LGCF algorithm. The UAV123 dataset has variable and tricky shooting angles and long video sequences, making the accuracy rate of the MOSSE_CA algorithm only 0.516 and the success rate only 0.398. The superior performance of the LGCF algorithm in the UAV dataset highlights the effectiveness of this algorithm research work.

Twenty-five more video sequences containing light change and scale change challenge attributes were selected on UAV123 for the light and scale change experiments of the LGCF algorithm with the test algorithm. The experimental results are shown in Figure 7:

The test results show that the LGCF algorithm has an accuracy rate of 0.792 and a success rate of 0.729 in the test sequence results with different scale changes of illumination. the accuracy rate of the KCF algorithm is 0.616 and the success rate is 0.520. the accuracy rate and success rate of the LGCF algorithm proposed in this paper far exceed those

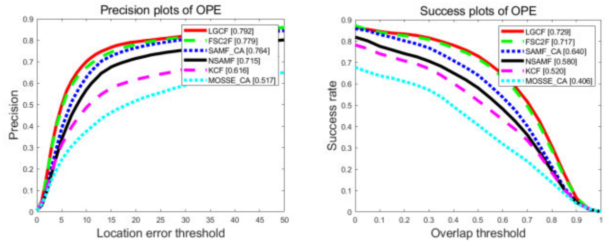


FIGURE 7. OPE accuracy plots for six trackers in light change scenes and scale change scenes.

TABLE 5. Performance comparison between LGCF and five tracking algorithms on TC128. means bigger is better.

Tracker	TC128	
	Precision ↑	Success ↑
LGCF	0.807	0.708
FSC2F	0.783	0.673
SAMF_CA	0.773	0.672
NSAMF	0.687	0.524
KCF	0.617	0.500
MOSSE_CA	0.519	0.400
ECO	0.795	0.693
ARCF	0.789	0.675

of the classical algorithm KCF, which reflects that the LGCF algorithm can adapt to the background changes in the scenes of illumination and scale changes and perform stable tracking of the target. It demonstrates that the LGCF can adapt to the background changes and track the target stably in the scenes with changing illumination and scale.

E. RESULTS ON TC128

The TC128 dataset contains 128 color video sequences and focuses on evaluating the tracking performance of the algorithm on color challenges. In this paper, 35 challenge video sequences are selected on the TC128 dataset, which contains various challenge attributes and rich color models, and the LGCF algorithm proposed in this paper is tested against five comparison algorithms. The test results are shown in Figure 8:

The LGCF algorithm has a success rate of 0.708 and an accuracy rate of 0.807 for the 35 color challenge sequence test results. The algorithm performance ranked first in the tested algorithms, reflecting the superior tracking effect of the LGCF algorithm on the color model to other algorithms. To better reflect the superiority of the algorithms, Table 5 is drawn and the ECO algorithm and ARCF algorithm are added for evaluation.

The ECO algorithm has an accuracy rate of 0.795 and a success rate of 0.693 out of 35 test sequences. The LGCF algorithm proposed in this paper has improved the accuracy rate by 1.2% and the success rate by 1.5% over the ECO algorithm. The second-ranked FSC2F algorithm has an accuracy rate of 0.783 and a success rate of 0.673. The LGCF algorithm

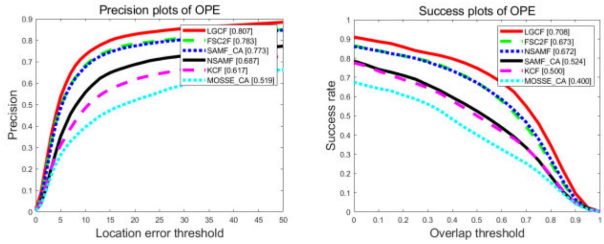


FIGURE 8. LGCF and five trackers on TC128 success rate graph and accuracy graph.

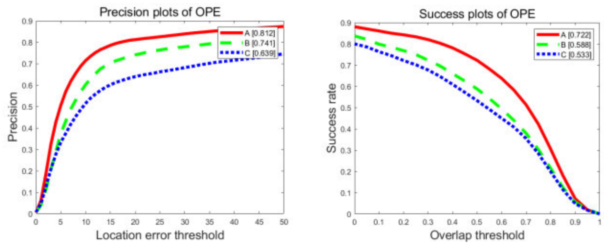


FIGURE 9. Plot of test results of algorithms A, B and C on the dataset.

improves the success rate by 2.4% and the accuracy rate by 3.5% over the FSC2F algorithm. The test results show that the LGCF algorithm ranks first in success rate and accuracy rate under color video evaluation, which reflects the powerful appearance model of the LGCF algorithm and stable tracking effect.

F. ABLATION EXPERIMENTS

To test the effectiveness of manual feature and depth feature fusion for constructing target appearance models. In this paper, we select 40 test challenge sequences on the OTB100 dataset, which contains 11 of the challenge attributes. The LGCF algorithm that fuses the three manual features and two layers of depth features proposed in this paper is denoted by Algorithm A. The algorithm that adds the extraction of Conv4-4 and Conv5-4 layer depth features to the correlation filtering framework using the VGG-19 network alone is denoted by Algorithm B. The algorithm that adds the fusion of HOG features, CN features, and ULBP features into the correlation filter tracking framework is denoted by the C algorithm. Algorithm A, Algorithm B, and Algorithm C were tested on the selected sequence of 40 complex challenge attributes, and the test results are shown in Figure 9:

Algorithm A has a success rate of 0.722 and an accuracy rate of 0.812. Algorithm B has a success rate of 0.588 and an accuracy rate of 0.741. The success rate of the C algorithm is 0.533 and the accuracy rate is 0.639. The results show that the algorithm proposed in this paper for the fusion of depth features and three manual features is 13.4% more successful and 7.1% more accurate than the algorithm using depth features alone. Compared to the C algorithm using three manual feature fusions alone, the success rate is 18.9% higher and the accuracy rate is 17.3% higher. The superiority of the LGCF algorithm in extracting target features using two layers of depth features and the fusion of three manual features is demonstrated. The representation of texture information by the manual features and the semantic information by the depth features complement each other, enabling the filter to extract more effective target features and generate an appearance feature model with a good tracking effect.

TABLE 6. FPS values before and after using the ADMM method on the challenge sequence.

Tracker	Challenge sequence
	FPS
ADMM-	24.805
ADMM+	31.263

Secondly, the speed of the filter will affect the tracking speed of the algorithm when performing complex calculations. This paper uses the ADMM method to optimize the calculation speed of the filter. The algorithm speed was tested on 40 test challenge sequences on the selected OTB100 dataset. The FPS value before acceleration is 24.805, and the FPS value after acceleration is 31.263, which is an increase of 6.458. By drawing Table 6, the experimental results are more clearly displayed.

V. CONCLUSION

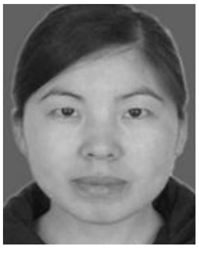
We use convolutional neural networks to extract deep features and fuse them with hand-crafted features to increase the texture information and semantic information of the appearance model and avoid the feature singularity of the appearance model. The background learning-related filter is added to enhance the performance of the tracker by continuously learning the tracking background. Then, the target template is updated by using the multi-template update strategy with high confidence, which eliminates the feature information interference caused by the pollution template and updates the correct template in time. Extensive experiments on classical datasets demonstrate the effectiveness of our work. The algorithm in this paper does not have strong richness in the feature selection stage. In the future, a more lightweight convolutional neural network model and better manual features can be selected for fusion when the algorithm is stable. In terms of tracking speed, better optimization methods can be selected to improve the calculation speed of the algorithm and the tracking ability of the algorithm.

REFERENCES

[1] T. Zhou and Y. Liu, ''Long-term person tracking for unmanned aerial vehicle based on human–machine collaboration,'' *IEEE Access*, vol. 9, pp. 161181–161193, 2021.
[2] Z. Tang, ''Intelligent target detection and tracking algorithm for martial arts applications,'' *Wireless Commun. Mobile Comput.*, vol. 2022, pp. 1–10, Mar. 2022.
[3] T. Yang, S. Jia, B. Yang, and C. Kan, ''Research on tracking and registration algorithm based on natural feature point,'' *Intell. Autom. Soft Comput.*, vol. 28, no. 3, pp. 683–692, 2021.
[4] P. Feng, C. Xu, Z. Zhao, F. Liu, J. Guo, C. Yuan, T. Wang, and K. Duan, ''A deep features based generative model for visual tracking,'' *Neurocomputing*, vol. 308, pp. 245–254, Sep. 2018.

[5] R. B. Devi, Y. J. Chanu, and K. M. Singh, "Discriminative object tracking with subspace representation," *Vis. Comput.*, vol. 37, no. 5, pp. 1207–1219, May 2021.
[6] B. Jiang, Y. Zhang, B. Luo, X. Cao, and J. Tang, "STGL: Spatial–temporal graph representation and learning for visual tracking," *IEEE Trans. Multimedia*, vol. 23, pp. 2162–2171, 2021.
[7] X. Hu, J. Li, Y. Yang, and F. Wang, "Reliability verification-based convolutional neural networks for object tracking," *IET Image Process.*, vol. 13, no. 1, pp. 175–185, Jan. 2019.
[8] T. Jiang, Q. Zhang, J. Yuan, C. Wang, and C. Li, "Multi-type object tracking based on residual neural network model," *Symmetry*, vol. 14, no. 8, p. 1689, Aug. 2022.
[9] Z. Li and J. Xu, "Target adaptive tracking based on GOTURN algorithm with convolutional neural network and data fusion," *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–12, Aug. 2021.
[10] Z. Hu, D. Yang, K. Zhang, and Z. Chen, "Object tracking in satellite videos based on convolutional regression network with appearance and motion features," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 783–793, 2020.
[11] 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.*, vol. 9914, 2016, pp. 850–865.
[12] R. Tao, E. Gavves, and A. W. M. Smeulders, "Siamese instance search for tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 1420–1429.
[13] K. Chen, X. Guo, L. Xu, T. Zhou, and R. Li, "A robust target tracking algorithm based on spatial regularization and adaptive updating model," *Complex Intell. Syst.*, vol. 9, no. 1, pp. 285–299, Feb. 2023.
[14] Z. Yuan and D. Shan, "Algorithms based on correlation filter in target tracking: A survey," in *Proc. IEEE Int. Conf. Electr. Eng. Mechatronics Technol. (ICEEMT)*, Qingdao, China, Jul. 2021, pp. 786–790.
[15] 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.*, San Francisco, CA, USA, Jun. 2010, pp. 2544–2550.
[16] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *Computer Vision—ECCV 2012*, vol. 7575. Berlin, Germany: Springer-Verlag, 2012, pp. 702–715.
[17] 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.
[18] M. Danelljan, G. Häger, F. Shahbaz Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," in *Proc. Brit. Mach. Vis. Conf.*, 2014, p. 65.
[19] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg, "Discriminative scale space tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 8, pp. 1561–1575, Aug. 2017.
[20] 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)*, Santiago, Chile, Dec. 2015, pp. 3074–3082.
[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[22] Z. Li, K. Nai, G. Li, and S. Jiang, "Learning a dynamic feature fusion tracker for object tracking," *IEEE Trans. Intell. Transport. Syst.*, vol. 23, no. 2, pp. 1479–1491, Feb. 2020.
[23] S. Zhao, T. Xu, X.-J. Wu, and X.-F. Zhu, "Adaptive feature fusion for visual object tracking," *Pattern Recognit.*, vol. 111, Mar. 2021, Art. no. 107679.
[24] D. Elayaperumal and Y. H. Joo, "Robust visual object tracking using context-based spatial variation via multi-feature fusion," *Inf. Sci.*, vol. 577, pp. 467–482, Oct. 2021.
[25] K. Nai, Z. Li, and H. Wang, "Dynamic feature fusion with spatial–temporal context for robust object tracking," *Pattern Recognit.*, vol. 130, Oct. 2022, Art. no. 108775.
[26] J. Li, J. Wang, and W. Liu, "Moving target detection and tracking algorithm based on context information," *IEEE Access*, vol. 7, pp. 70966–70974, 2019.
[27] 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. 1144–1152.
[28] D. Yuan, X. Shu, and Z. He, "TRBACF: Learning temporal regularized correlation filters for high performance online visual object tracking," *J. Vis. Commun. Image Represent.*, vol. 72, Oct. 2020, Art. no. 102882.
[29] K. Dai, D. Wang, H. Lu, C. Sun, and J. Li, "Visual tracking via adaptive spatially-regularized correlation filters," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 4665–4674.
[30] D. Li, X. Xiang, Y. Kang, and W. Tao, "Adaptive visual tracking via temporal-spatial regularized multi-patches," in *Proc. IEEE 6th Inf. Technol. Mechatronics Eng. Conf. (ITOEC)*, Chongqing, China, Mar. 2022, pp. 1043–1047.
[31] C. Deng, S. He, Y. Han, and B. Zhao, "Learning dynamic spatial–temporal regularization for UAV object tracking," *IEEE Signal Process. Lett.*, vol. 28, pp. 1230–1234, 2021.
[32] C. Wang, Z. Shi, K. Si, Y. Su, Z. Li, and E. Wang, "Anti-occlusion and scale adaptive target tracking algorithm based on kernel correlation filter," in *Proc. IEEE 20th Int. Conf. Trust, Secur. Privacy Comput. Commun. (TrustCom)*, Shenyang, China, Oct. 2021, pp. 130–138.
[33] Y. Xiao and D. Pan, "Robust visual tracking via multilayer CaffeNet features and improved correlation filtering," *IEEE Access*, vol. 7, pp. 174495–174506, 2019.
[34] L. Zhou, Y. Jin, H. Wang, Z. Hu, and S. Zhao, "Robust DCF object tracking with adaptive spatial and temporal regularization based on target appearance variation," *Signal Process.*, vol. 195, Jun. 2022, Art. no. 108463.
[35] L. Wang and C. Pan, "Visual object tracking via a manifold regularized discriminative dual dictionary model," *Pattern Recognit.*, vol. 91, pp. 272–280, Jul. 2019.
[36] M. Wang, Y. Liu, and Z. Huang, "Large margin object tracking with circulant feature maps," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 4800–4808.
[37] D. Wu, "Research on target tracking method of sports video based on multi-template matching," in *Proc. Int. Conf. Virtual Reality Intell. Syst. (ICVRIS)*, Zhangjiajie, China, Jul. 2020, pp. 82–85.
[38] T. Xu, Z.-H. Feng, X.-J. Wu, and J. Kittler, "Learning adaptive discriminative correlation filters via temporal consistency preserving spatial feature selection for robust visual object tracking," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5596–5609, Nov. 2019.
[39] S. Wang, X. Liu, S. Liu, K. Muhammad, A. A. Heidari, J. D. Ser, and V. H. C. de Albuquerque, "Human short long-term cognitive memory mechanism for visual monitoring in IoT-assisted smart cities," *IEEE Internet Things J.*, vol. 9, no. 10, pp. 7128–7139, May 2022.
[40] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010.
[41] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, Jan. 1996.
[42] S. Yan and M. Yang, "Alternating direction method of multipliers with variable stepsize for partially parallel MR image reconstruction," in *Proc. 36th Chin. Control Conf. (CCC)*, Dalian, China, Jul. 2017, pp. 10886–10889.
[43] H. Ma, S. T. Acton, and Z. Lin, "SITUP: Scale invariant tracking using average peak-to-correlation energy," *IEEE Trans. Image Process.*, vol. 29, pp. 3546–3557, 2020.
[44] 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. 1387–1395.
[45] Z. Qu, W. Yi, R. Zhou, H. Wang, and R. Chi, "Scale self-adaption tracking method of Defog-PSA-Kcf defogging and dimensionality reduction of foreign matter intrusion along railway lines," *IEEE Access*, vol. 7, pp. 126720–126733, 2019.
[46] J. Fang, Z. Li, and J. Xue, "Spatial-sequential-spectral context awareness tracking," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Beijing, China, Sep. 2017, pp. 2582–2586.
[47] M. Kristan, J. Matas, A. Leonardis, M. Felsberg, and R. Pflugfelder, "Visual tracking via fast saliency-guided continuous correlation filters," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2019, pp. 2206–2241.
[48] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2411–2418.
[49] 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.

[50] 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.
[51] M. Mueller, N. Smith, and B. Ghanem, "A benchmark and simulator for UAV tracking," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Cham, Switzerland: Springer, Oct. 2016, pp. 445–461.
[52] Z. Huang, C. Fu, Y. Li, F. Lin, and P. Lu, "Learning aberrance repressed correlation filters for real-time UAV tracking," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 2891–2900.
[53] 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)*, Jul. 2017, pp. 6638–6646.



HONGGE REN received the B.E. degree in measurement and control technology and instrument from the North China University of Science and Technology, Tangshan, Hebei, China, in 2003, and the M.E. and Ph.D. degrees from the Beijing University of Technology, Beijing, China, in 2007 and 2011, respectively.

She is currently a Professor with the School of Control and Mechanical Engineering, Tianjin Chengjian University. Her research interests include cognitive robots, image processing, and object tracking.

LEIGANG XING received the B.E. degree in automation from the Hebei Normal University of Science and Technology, Qinhuangdao, Hebei, China, in 2021. He is currently pursuing the master's degree with the North China University of Science and Technology.

His research interests include computer vision, correlation filter, and object tracking.

TAO SHI received the Ph.D. degree in control science and engineering from the Beijing University of Science and Technology, Beijing, China, in 2015.

He is currently an Associate Professor with the School of Electrical Engineering and Automation, Tianjin University of Technology. His research interests include brain-like intelligent robots, robot vision, and biologically inspired intelligent computing.