IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 16, 2021, accepted July 25, 2021, date of publication July 27, 2021, date of current version August 3, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3100557

Research on Visual Tracking Algorithm Based on Peak Sidelobe Ratio

YIN CUI [1] **AND HONGGE REN** [2]

[1]College of Electrical Engineering, North China University of Science and Technology, Qinhuangdao 063210, China
[2]School of Control and Mechanical Engineering, Tianjin Chengjian University, Tianjin 300384, China

Corresponding author: Hongge Ren (renhg@tcu.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 illumination changes, target deformation and background clutter in the target tracking field, a visual tracking algorithm based on peak sidelobe ratio is proposed. The object-interference model is used to represent the target appearance model, and context information is added to the relevant filtering framework. Training the filter internally to enhance the ability of filter discrimination. In the model update process, it is easier to introduce samples that cannot characterize the target, and use peak sidelobe comparison to update the tracking parameters, which can enhance the generalization ability of the model. Tested with some classic and recently algorithms in the OTB50, OTB100, UAV123, TC128 experimental video data set, the experiment' results show that the visual tracking algorithm that is proposed in the article can track the target more accurately. It has important research on the development of intelligent video surveillance value.

INDEX TERMS Target tracking, machine vision, correlation filter, peak sidelobe ratio.

I. INTRODUCTION

In recent years, people innovated a lot of things, and computer vision technology has continued to develop, and the problem of target tracking has become more significant and it is the basis in the field of vision. A large number of scholars in this direction have conducted research and made substantial progress, but there are still many problems that need to be further improved. For example, problems such as the target being occluded, the target scale change, the target being partially or completely occluded, the illumination change in the object, and the background clutter in the tracking process could cause the target tracking failure.

Video target tracking usually obtains a target motion model based on the background and target information in the frame sequence. After a series of processing, it predicts the target's position, size, shape, trajectory and other behavior states in the video, which is the target behavior. The foundation for advanced tasks such as understanding and analysis and behavior prediction. Target tracking also has comprehensive application value in traffic flow monitoring [1], robots [2], human-computer interaction [3], autonomous driving [4], medical diagnosis [5], etc. It is a variety of intelligent systems that can play different roles. It is an indispensable part of the process, but there is still much room for improvement in tracking the target object in a complex environment to achieve the accuracy and success rate requirements. Therefore, it is vital to study a visual target tracking algorithm that takes into account tracking speed and accuracy.

Traditional target tracking algorithms include particle filter method [6], EKF interactive multi-model algorithm [7], mean shift method [8], Struck method [9] and so on. And now, with the development of two main ways that include deep learning frameworks and related filtering frameworks, current tracking methods mainly adopt regression and classification models for research, mainly discriminant methods [10]. Using strong generalization and characteriza-tion capabilities, deep learning models have received more and more attention, and many algorithms based on neural networks with excellent performance have been proposed one after another. For example, STCT [11] based on early neural network, SINT [12] based on main network. Deep features are used in many trackers. These trackers can deal with complex problems and track object easily [13]–[15]. Ma et al proposed HCFT [16] method and the method improved system performance based on the correlation filter and used the convolutional features. Correlation filtering methods are also popular and it has a fast speed.

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

The correlation filtering method regards the process of performing correlation filtering operations on the image region of interest to obtain the result as the process of target tracking, which has certain advantages in terms of speed and accuracy. In 2010, Henriques *et al.* [17] proposed the MOSSE method, the processing speed reached 669fps, but the accuracy was not very good. The CSK algorithm [18] added a regular term on the basis of the MOSSE algorithm, proposed the concept of a classifier, and improved the accuracy. The subsequent KCF [19] algorithm reached 172 fps in terms of speed, which fully met the requirements of real-time performance, and achieved good results in terms of accuracy. It has become a popular algorithm in the field of target tracking and has been used by a large number of scholars for algorithm comparison and improvement.

As stated in the convolution theorem, the multiplication of the elements in the Fourier domain corresponds to the correlation in time domain. Recently, CF is researched by a large number of scholars [20]. Sui *et al.* [21] proposed a new tracking algorithm that could enhance the performance of identifying the peak of the correlation response. Yan *et al.* [22] used multi-channel HOG features to track small object. And they combined two responses to achieve accuracy. Yuan *et al.* [23] proposed a target-focusing convolutional regress-ion model to complicated situations and achieved better performance. Moreover, Yuan *et al.* [24] used multiple features to build the target model for object tracking to deal with the tracking accuracy problem caused by a single feature. Zhang *et al.* [25] combined the texture features and color features to perform optimal similarity matching. And they proposed scale normalization to track in the candidate frames. Xiong *et al.* [26] proposed a rotation parameter estimation and target scale method and it could deal with problems caused by long-term target tracking. When the object was lost, the method started to determine the possible position of the target in the current frame. Joakim Johnander *et al.* proposed DCCO [27], which regularized the sub-filter locations with an affine deformation model to avoid over-fitting. And the method improved the performance in challenging situations.

Nan *et al.* [28] combined spatial reliabilities of samples and the time regularization with the channel for the first time. And they used ADMM to solve calculation problem. Qi e al. [29] proposed a multi-cue CPF for short visual tracker and combined the correlation filter and particle filter to prove the tracker's effectiveness. Yuan *et al.* [30] matched current frames and previous frames and added a motion detection algorithm to correlation filter tracking. It has greatly advantage for fast motion scenes.

The hybrid between CF and CNN trackers had great performance in target tracking. CFNet [31] used conv1, conv2 and conv5 in the proposed algorithm framework to compare the performance differences, and concluded that CFNet-Conv2 was more suitable for completing the tracking task. Danelljan M et al proposed the ECO tracker [32], which used factorized convolution operator to reduce the dimensions of HOG, CN and CNN. They also reduced the training parameters, thereby effectively reducing the computational complexity. ECO+ [14] changed the cha-racteristic network to ResNet-50 and divided into shallow features and deep features to get the great result.

The new CNN-based trackers also had high accuracy to track target. Faster R-CNN [33] introduced a RPN that simultaneously predicted object bounds and objectness scores at each position. Wang *et al.* proposed the SiamMask [34] and they replaced the shallow network AlexNet that extracted features with the deeper network ResNet, and added a mask branch to RPN in parallel to further improve the accuracy of tracking. Siam R-CNN [35] was introduced as a Siamese two-stage full-image redetection architecture with a Tracklet Dynamic Programming Algorithm. And it showed strong results for long-term tracking.

Our main contributions are as follows:

1) The color feature and HOG feature are combined to extract features of the target.
2) Adds contextual information and assign it to the tracker in the learning phase and enhances the discriminative ability of the filter.
3) Use the peak sidelobe ratio in the object template update process to improve the ability of related filtering algorithms to deal with tracking problems and achieve the goal of track accurately.

II. RELATE WORKS

Correlation filter has good tracking performance and it is an efficient tracking method. We used CF as our basic framework to track object. Our method can well deal with tracking scenes. In the following, we will briefly review some workers related to our algorithm.

A. CORRELATION FILTER TRACKER

Correlation filter trackers have got a lot of attention due to their rapid speed and performance. MOSSE tracker used gray feature to build the target model, and the speed was very fast. It was the pioneer of correlation filter tracker. Li and Zhu proposed the SAMF tracker [36] and added the scaling pool which calculated the target scale by matching the features of seven scales. Then, KCF tracker used HOG feature and correlation filters in a kernel space. And it adopted a circulant matrix to train the target sample and solved the problem of insufficient samples. But boundary effects were produced due to the cyclic shift operation of the target sample, Danneljan *et al.* [37] proposed spatial regularization to solve the problem. Bertinetto *et al.* [38] combined HOG feature and global color histogram to build Staple tracker, it developed the inherent structure of each patch by maintaining two independent regression questions and was inherently robust to both deformations and color changes. The tracker's calculation efficiency and accuracy had been significantly improved, and the frame rate had reached 80FPS.

With the rapid development of deep learning, many correlated filter trackers that combined deep convolution features had emerged. Based on the KCF, [39] introduced rich hierarchical convolution features in the correlation filtering framework for visual tracking. It used specific three-layer features trained on ImageNet to replace the original HOG features for modeling.

B. OBJECT MODEL UPDATE

Object model update plays an important role in target tracking and it largely affects the quality of the tracking algorithm. Generally, it is impossible for the target to maintain the same posture in the video frame sequence. There will be various situations in the target update. For example, Illumination Variation, Occlusion, Background Clutters and so on. Therefore, adopting simple update methods are often inaccurate. Many researchers have proposed a large number of algorithms to optimize the above interference factors, and achieving good tracking results in specific image sequences. Wang *et al.* proposed LMCF [36] tracker and introduced the speed advantage of correlation filter. They used structure SVM as a classifier, and used online update method. While taking into account the accuracy, the calculation efficiency was greatly accelerated.

Recently, DPT [40] used iterative direct method for efficient optimization and constrained the geometric features and visual to a convex cost function. The tracker used bottom-up update to achieve goal tracking. Fan *et al.* [41] proposed adaptive and complementary correlation filter to achieved relative high tracking speeds. Ping *et al.* [42] used APCE and the response peak to update object model. Once the tracking failed, the algorithm would quickly make a judgment. The tracker had achieved good tracking results and solved the long-term object tracking problem in complex scenes.

III. ALGORITHM DESIGN

In this section, we will introduce our tracker in detail. The overall framework is shown in the Fig.2.

A. OBJECT-INTERFERENCE MODEL

The color feature has achieved good tracking results in the scenes of target rotation, translation changes and background clutter, which is reflected in the DAT [43] tracker. This article partly shows that the target model of interest is in the form of a color histogram, and a Bayesian classifier is used to judge the target position on the input frame.

The target area of interest is defined as $R \in \{\alpha, \beta\}$ in the search stage, where α is the target area, and β is the interference area. For the object pixel $x \in O$, applying Bayesian rule to obtain the object likelihood at position x is

$$P(x \in O|\alpha, \beta, c_x) = \frac{P(c_x|x \in \alpha)P(x \in \alpha)}{\sum_{\gamma \in \{\alpha,\beta\}} P(c_x|x \in \gamma)P(x \in \gamma)} \quad (1)$$

Among them, the input picture is I. Where c_x is the histogram corresponding to the interval of x in the color histogram H on the image area R. The foreground histogram model is represented by H_α^I, and the background histogram model is represented by H_β^I. We use the probability of the color histogram to estimate the likelihood term, namely:

$$P(c_x|x \in \alpha) \approx H_\alpha^I(c_x)/\,|\alpha| \quad (2)$$

$$P(c_x|x \in \beta) \approx H_\beta^I(c_x)/\,|\beta| \quad (3)$$

$$P(x \in \alpha) \approx |\alpha|\,/(|\alpha| + |\beta|) \quad (4)$$

Therefore, the probability simplification of $x \in O$ (belonging to the object area) in the input picture I and (1) establishes the object-interference model as

$$P(x \in O|\alpha, \beta, c_x) = \begin{cases} \dfrac{H_\alpha^I(c_x)}{H_\alpha^I(c_x) + H_\beta^I(c_x)} & if\ I(x) \in (\alpha \cup \beta) \\ 0.5 & otherwise. \end{cases} \quad (5)$$

The object-interference model is shown in Figure 1. In the formula, the value of the invisible pixels is set to 0.5 in the image. Equation (5) can be used to calculate the likelihood probability by looking up the table to obtain a color probability map that separate goals and distracting factors.

The object-interference model is established by extracting color features, and the color features are prone to tracking drift phenomenon under illumination changes. The interference model is not contributing to the illumination changes. The illumination changes occur in the video frame scene.

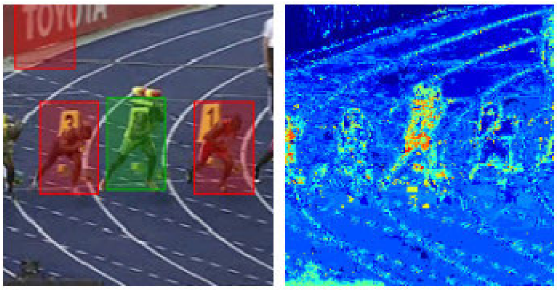


FIGURE 1. Object-interference model. The green area is object model and the red area is interference model.

B. KERNELIZED CORRELATION FILTERS

CACF [44] provides a framework for the correlation filter tracker, which adds contextual information and assign it to the tracker in the learning phase, and enhances the discriminative ability of the filter.

Inspired by CACF, Q context image blocks $b_i \in R^n$ and target image blocks $b_0 \in R^n$ are sampled by uniform sampling around the target on each input image sequence. The circulant matrices formed by them are B_i and B_0. The contextual image blocks are negative samples, and they allow the filter to better distinguish the context that is about to become an occluder or background. We want to learn a filter $h \in R^n$ that is close to zero in the context information. To the target image block, it has a high response. The filter obtained is in accordance with the decomposition strategy of the CACF article to reduce the complexity of the system. Therefore, we propose the filter model as

$$\min_h \|B_0 h - y\|_2^2 + \delta_1 \|h\|_2^2 + \delta_2 \sum_{i=1}^{Q} \|B_i h\|_2^2 \quad (6)$$

In the formula, h is expressed as a correlation filter. The regression target y is a 2D Gaussian vectorized image, and δ_1 and δ_2 are parameters.

In this paper, the features are HOG features and color features, so (6) is used in the original domain for multi-channel features and a joint filter h is learned, so that the dimension of all features is m, and the multi-channel original objective function is constructed as

$$f(h) = \|Ch - \bar{y}\|_2^2 + \delta_1 \|h\|_2^2 \tag{7}$$

$$C = \begin{bmatrix} B_0 \\ \sqrt{\delta_2}\mathrm{B}_1 \\ \vdots \\ \sqrt{\delta_2}\mathrm{B}_k \end{bmatrix} \tag{8}$$

$$\bar{y} = \begin{bmatrix} y \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{9}$$

And $C \in R^{(k+1)n\times nm}$ is a new data matrix, which is obtained by stacking the target image on the context image block. The new regression target $\bar{y} \in R^{(k+1)n}$ is connected with zero.

Since the f(h) function is convex, use the method of setting the gradient to zero to minimize it, and get

$$h = (C^T C + \delta_1 I)^{-1} C^T \bar{y} \tag{10}$$

Diagonalize the new data matrix to get

$$\widehat{h} = \begin{bmatrix} D_{11} & \cdots & D_{1m} \\ \vdots & \ddots & \vdots \\ D_{m1} & \cdots & D_{mm} \end{bmatrix}^{-1} \begin{bmatrix} diag(\widehat{b}_{01}^* \odot \widehat{y}) \\ \vdots \\ diag(\widehat{b}_{0m}^* \odot \widehat{y}) \end{bmatrix} \tag{11}$$

In the formula, $\odot$ is the dot product of the elements. Each feature dimension $l, v \in \{1, \cdots, m\}$ in the target image block and the context image block is represented by a_{0l} and a_{lv}, respectively. The role of diag is to extract the diagonal elements of the matrix. $\wedge$ is the Fourier transform, and $*$ represents the conjugate operation.

$$\begin{cases} D_{ll} = diag(\widehat{b}_{0l}^* \odot \widehat{b}_{0j} + \delta_2 \sum_{i=1}^{Q} \widehat{b}_{il}^* \odot \widehat{b}_{ij}) + \delta_1 I \\ D_{lv} = diag(\widehat{b}_{0l}^* \odot \widehat{b}_{0v} + \delta_2 \sum_{i=1}^{Q} \widehat{b}_{il}^* \odot \widehat{b}_{iv}), j \neq l \end{cases} \tag{12}$$

C. DETECTION PHASE

The learned filter $\widehat{h}$ is convolved with the image block z in the next frame in the video frame, where Z is the cyclic matrix.

$$S_p = Zh \tag{13}$$

The above convolution operation can be simplified to

$$S_{p1} = \widehat{z} \odot \widehat{h} \tag{14}$$

In the above formula, the image block z and the learned filter are both m-dimensional.

Calculate the sum of the pixel probabilities in the color probability map that are equal to the object size rectangle in the video frame, and set the length of the rectangle as a and the width as e. The response of color histogram is:

$$\begin{aligned} S_{p2} = (&J(r, u) + J(r+a, u+e) - J(r+a, u) \\ &- J(r, u+e)) \end{aligned} \tag{15}$$

$$J(p_a, p_e) = \sum_{p_{a1} \le p_a, p_{e1} \le p_e} I(p_{a1}, p_{e1}) \tag{16}$$

In (15), S_{p2} represents the response of color histogram. And (r,u) is the upper left corner of the rectangular box. The response of color histogram can be easily obtained by looking up the four-fold integral graph on the probability graph. Integral map $J(p_a, p_e)$ is the sum of all pixel values $I(p_{a1}, p_{e1})$ in the upper left corner of rectangular position (p_a, p_e).

Integrate the color feature response and the HOG feature response, and set the fusion factor to η get the final response score as

$$S_{p3} = (1 - \eta) S_{p1} + \eta S_{p2} \tag{17}$$

D. UPDATE STRATEGY

In target tracking algorithms, most discriminant tracking algorithms use linear interpolation for the template update problem. This method generally uses a fixed learning rate, and the learning rate is difficult to select. A small error will cause tracking drift for long-term tracking. Inspired by the article [45], an adaptive update strategy based on the peak sidelobe ratio is used to update the object-interference model as

$$H_\alpha^I(c_x)_{(t)} = (1 - \rho\eta) H_\alpha^I(c_x)_{(t-1)} + \rho\eta H_\alpha^I(c_x)_{(t)} \tag{18}$$

$$H_\beta^I(c_x)_{(t)} = (1 - \rho\eta) H_\beta^I(c_x)_{(t-1)} + \rho\eta H_\beta^I(c_x)_{(t)} \tag{19}$$

$$\eta = \frac{\mu_t}{\max}(\mu_i | i \in [1, 2, \cdots, n]) \tag{20}$$

$$\mu_t = \frac{\varphi - \omega}{K} \tag{21}$$

In the above formula, ρ is the learning factor, and t is the current number of frames. η is the relative weight of the target change. μ_t is the peak sidelobe ratio. φ, ω, K is the peak value, mean value and variance of the region with the size of 11×11 and the peak value of the confidence map as the center. The model parameter adopts the relative deviation of the target confidence of an adjacent video frames, and uses the relative deviation as a metric for adaptive updating.

In the article, first use the PSR value of the final position response map of the current frame to determine whether the target has tracking drift and occlusion. And use PSR to re-determine the update rate of the model.

$$\rho^* = \begin{cases} 0.025 & PSR_t > T \\ 0 & PSR \le T \end{cases} \tag{22}$$

In (22), ρ^* is the new learning factor. Through a lot of experimental operations, we found that the PSR value of the target during normal movement exceeds 7. In this situation, replacing ρ in the previous frame in (18) and (19) with the new learning factor ρ^*, and use it in the next array. And when the PSR value is less than 7, the target may track drift and occlusion and we keep the target template information of the previous frame to adaptively update the model. So, the value of T is 7 to effectively update the object model.

E. ALGORITHM DESCRIPTION

According to the above theoretical derivation and formula analysis, the tracking algorithm flow is shown in Figure 2. First initialize the target, start to input the video frame, and in the current frame extract the color feature and HOG feature, construct the feature model and calculate the response of the two models, use (17) to obtain the final response score, and the predict target position use maximum value. And the established model is updated by an adaptive update strategy that is based on the peak sidelobe ratio. After the update is completed, the next frame operation is performed until all the video frame sequences are tracked to the target.

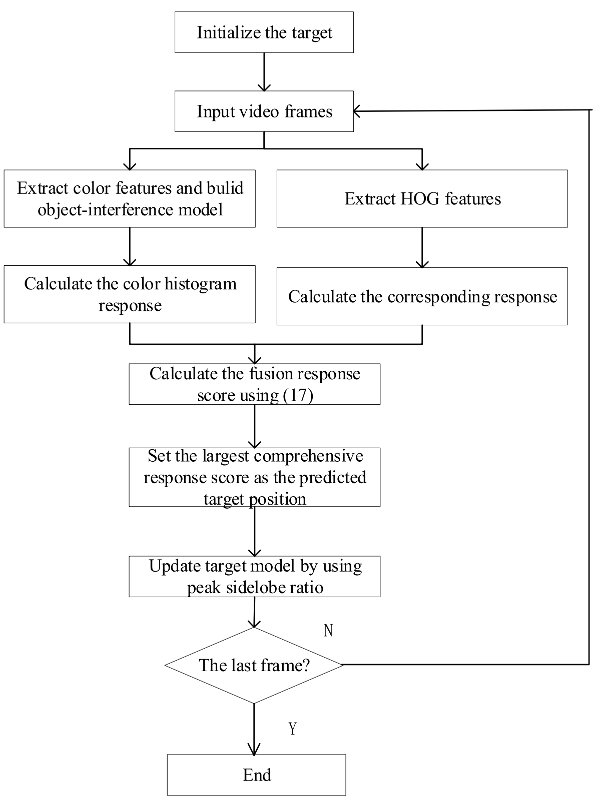


FIGURE 2. **The tracking algorithm flow.**

IV. EXPERIMENTS

A. IMPLEMENTATION DETAILS

The visual tracking algorithm based on peak sidelobe ratio proposed in this paper uses an object-interference model and a kernel-related filter to deal with target deformation and occlusion problem. In order to verify the proposed algorithm's effectiveness, it is compared with tracking algorithms. The experimental platform is simulated on Matlab R2014a, and the experimental environment is Intel(R) Core (TM) i5-8265U CPU @ 1.6GHz 1.8GHz. To ensure the accuracy of the experiment, we use OTB50 [46], OTB100 [47], UAV123 [48] and TC128 [49] for testing.

We use the success rate curve and the one-time evaluation precision curve for analysis. The algorithm proposed in this paper is represented by Pro1. The tracker follows the setting in CACF. It is better than the KCF algorithm using HOG feature alone.

B. RESULTS ON OTB50

OTB50 contains 49 challenging video sequences, it is manually tagged with 11 attributes that infect different problems, including occlusion (OCC), scale variation (SV), motion blur (MB), illumination variation (IV), deformation (DEF), background clutter (BC), fast motion (FM), in-plane rotation (IPR), out-of-plane rotation (OPR), low resolution (LR), out-of-view (OV), which are shown in Table 1. We compared pro1 with other five algorithms on OTB50. The algorithms are KCF, SAMF_AT [50], MOSSE_CA [44], DCF_CA [44], and KCF_SC [51].

The results on OTB50 are shown in Figure 3. It can be learned from the success rate graph that the algorithm has a success rate of 67.5% when the overlap rate threshold is [0, 1], which is much better than other algorithms. From the one-time evaluation accuracy curve, it can be concluded that the accuracy of the algorithm proposed in this paper is 74.9%. In order to clearly see the comparison between this algorithm and other algorithms, Table 2 lists the success rate and accuracy of various algorithms in the sequence. And we add ECO to compare with various algorithms.

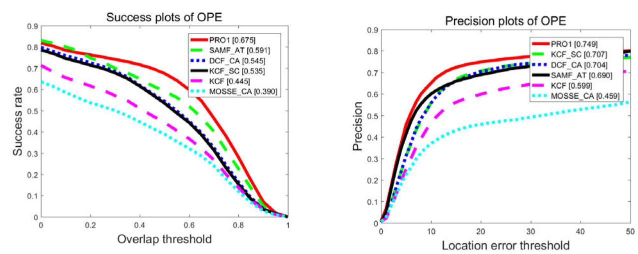


FIGURE 3. **Success plots of OPE and precision plots of OPE of six trackers on OTB50.**

Compared with the SAMF_AT algorithm, our method improves the success rate by 8.4% and precision by 5.9%, which illustrates the superiority of the algorithm in this paper.

C. RESULTS ON OTB100

OTB100 contains 98 challenging video sequences and it has a total of 100 test scenarios. The sequences are manually tagged with 11 attributes that infect different problems. Each attribute is a problem that needs to be solved in target tracking. It contains occlusion (OCC), scale variation (SV), motion blur (MB), illumination variation (IV), deformation (DEF), background clutter (BC), fast motion (FM), in-plane rotation

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

Sequence	OCC	SV	MB	IV	DEF	BC	FM	IPR	OPR	LR	OV
Biker	√	√	√				√		√	√	√
Bolt	√				√			√	√		
Car1		√	√	√		√	√			√	
CarDark				√		√					
Liquor	√	√	√	√		√	√		√		√
Girl	√	√						√	√		
Deer			√			√	√	√		√	
Shaking		√		√		√		√	√		
MotorRolling		√	√	√		√	√	√		√	
Walking20	√	√								√	
Sylvester				√				√	√		
Soccer	√	√	√	√		√	√	√	√		

TABLE 2. Performance comparison of various algorithms on OTB50. ↑ means bigger is better.

Tracker	OTB50	
	Success↑	Precision↑
PRO1	0.675	0.749
KCF	0.445	0.599
KCF_SC	0.535	0.707
SAMF_AT	0.591	0.690
DCF_CA	0.545	0.704
MOSSE_CA	0.390	0.459
ECO	0.668	0.716

(IPR), out-of-plane rotation (OPR), low resolution (LR), out-of-view (OV).

The results on OTB100 are shown in Figure 4. We evaluated PRO1 and five correlation filtering algorithms. The success plots of OPE and precision plots of OPE are shown that the proposed algorithm's performance has been significantly improved. We compared the trackers including KCF, SAMF_AT [50], MOSSE_CA [44], DCF_CA [44], and KCF_SC [51] on the challenging video sequences. In order to clearly see the comparison between this algorithm and other algorithms, Table 3 lists the success rate and accuracy of various algorithms in the sequence. And we add ECO and DCCO to compare with various algorithms.

Compared with the second-placed tracker ECO, our method improves the success rate by 5% and precision by 0.7%, which illustrates the superiority of the algorithm in this paper.

Figure 5 chooses 3 out of 100 sequences with more complex sequences, and compares the algorithm proposed in this paper with the five algorithms proposed by previous researchers. The KCF algorithm uses the circulant matrix and the fast calculation results of the Fourier transform, and the speed is very fast, up to 172FPS, and it has a good tracking effect on occlusion and deformation.

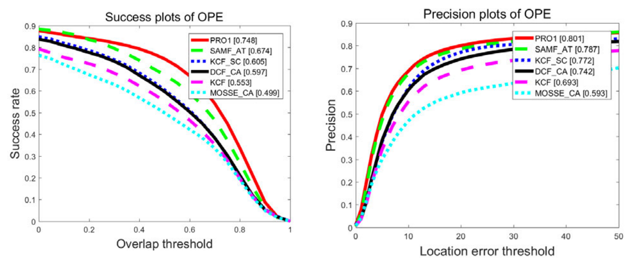


FIGURE 4. Success plots of OPE and precision plots of OPE of six trackers on OTB100.

TABLE 3. Performance comparison of various algorithms on OTB100. ↑ means bigger is better.

Tracker	OTB100	
	Success↑	Precision↑
PRO1	0.748	0.801
KCF	0.553	0.693
KCF_SC	0.605	0.772
SAMF_AT	0.674	0.787
DCF_CA	0.597	0.742
MOSSE_CA	0.499	0.593
ECO	0.698	0.794
DCCO	0.690	0.595

In Figure 5(a), the target is deformed. The algorithm in this paper quickly tracks the target at the beginning, and KCF, MOSSE_CA tracking algorithms have tracking drift. After the 244 frames of the sequence where the athletes cross the line, the proposed tracking algorithm and KCF_SC, DCF_CA, SAMF_AT algorithms can track target well. And other algorithms have deviated from the object, causing tracking failure. In bolt sequence, the proposed algorithm tracked the target well in fast motion and target deformation.

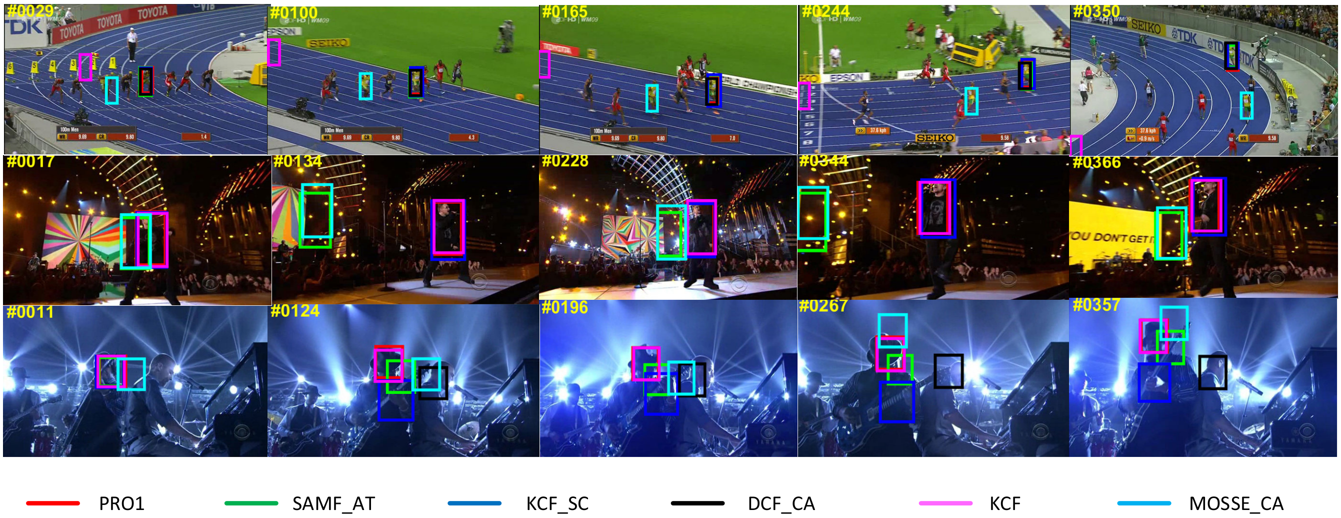


FIGURE 5. **(a) Bolt sequence (b) singer2 sequence (c) shaking sequence. Comparison PRO1 and SAMF_AT, DCF_CA, KCF_SC, KCF, MOSSE_CA on OTB100.**

In Figure 5(b), a singer occurs in the scene and the illumination changes. The singer is dancing on the stage, his body is constantly deforming, and the big screen behind him is constantly playing different videos. The background is very messy. The tracking algorithm proposed in the article can track the target well. As time goes by, at 344 frames, the SAMF_AT algorithm and MOSSE_CA algorithm have failed to track the object, showing the superiority of the algorithm in this paper.

In Figure 5(c), the target of interest in the video is a person holding a guitar, and the lighting in the scene has changed significantly. At the beginning of the video frame, the MOSSE_CA algorithm has deviated from the target, other algorithms can track correctly. The band in the scene has many people and equipment, which affects the algorithm to accurately track the target. At the 196th frame, the lighting is relatively strong. The proposed tracker in this paper has good robustness to illumination. At the end frame, only the PRO1 and KCF can track the target.

Figure 6 is a comparison of the deformation scene and illumination scene in the tracking challenge between the PRO1 and the state-of-the-art algorithm. The peak sidelobe ratio is used in the article to update the model. The algorithm has obvious advantages and can easily cope with challenges. The accuracy of the two scenarios reaches 0.764 and 0.759 respectively.

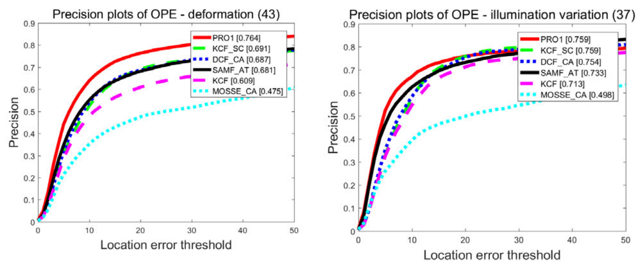


FIGURE 6. **Precision plots of OPE of six trackers in deformation scene and illumination scene.**

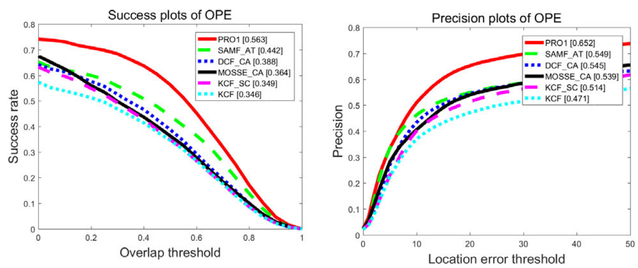


FIGURE 7. **Success plots of OPE and precision plots of OPE of six trackers on UAV123.**

D. RESULTS ON UAV123

UAV123 contains 91 videos, which has 123 short sequences and the characteristic is that the background is clean and the viewing angle changes a lot. The sequences are a dataset of special scenes, and are all shot with drones.

The results on UAV123 are shown in Figure 7. We evaluated PRO1 and five correlation filtering algorithms. We compared the trackers including KCF, SAMF_AT [50], MOSSE_CA [44], DCF_CA [44], and KCF_SC [51] on the challenging video sequences. The plots of OPE are shown that the proposed algorithm's performance has been greatly improved. In order to clearly see the comparison between this algorithm and other algorithms, Table 4 lists the accuracy and success rate of various algorithms in the sequence. And we add ECO to compare with various algorithms.

Our method improves SAMF_AT with an absolute gain of 12.1% and 10.3%, in terms of success and precision. The SAMF_AT tracker and PRO1 show great performance. Compared with DCF_CA, PRO1 has improved success and precision by 17.5% and 10.7%. Moreover, our method can accurately track the target and has great performance.

TABLE 4. Performance comparison of various algorithms on UAV123. ↑ means bigger is better.

Tracker	UAV123	
	Success↑	Precision↑
PRO1	0.563	0.652
KCF	0.346	0.471
KCF_SC	0.349	0.514
SAMF_AT	0.442	0.549
DCF_CA	0.388	0.545
MOSSE_CA	0.364	0.539
ECO	0.530	0.617

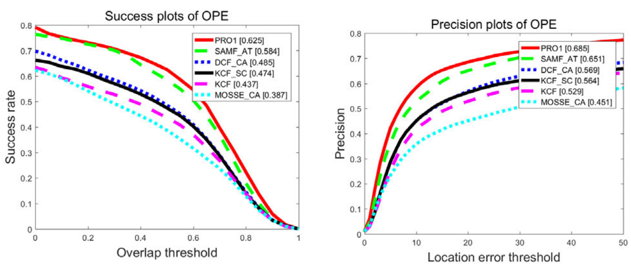


FIGURE 8. Success plots of OPE and precision plots of OPE of six trackers on TC128.

TABLE 5. Performance comparison of various algorithms on TC128. ↑ means bigger is better.

Tracker	TC128	
	Success↑	Precision↑
PRO1	0.625	0.685
KCF	0.437	0.529
KCF_SC	0.474	0.564
SAMF_AT	0.584	0.651
DCF_CA	0.485	0.569
MOSSE_CA	0.387	0.451
ECO	0.601	0.669
DCCO	0.594	0.660

E. RESULTS ON TEMPLE COLOR 128

Temple Color 128 contains 128 challenging video sequences. The data set includes the original 50 videos, and then 78 sequences that have not been seen before are added on this basis. The color sequence contains 11 different problem attributes, and the color model is very helpful for tracking the target under target deformation and illumination variation.

The results on TC128 are shown in Figure 8. We evaluated PRO1 and five correlation filtering algorithms. We compared the trackers including KCF, SAMF_AT [50], MOSSE_CA [44], DCF_CA [44], and KCF_SC [51] on the challenging video sequences. In order to clearly see the comparison between this algorithm and other algorithms, Table 5 lists the results of various algorithms in the sequence. And we add ECO and DCCO to compare with various algorithms.

The success rate of our tracker is 62.5% and the precision is 68.5%. Compared with ECO, PRO1 achieves some improvement by a gain of 2.4% and 1.6%, in terms of success and precision. Moreover, our method performs better and it ranks first and our tracker can track the target more accurately in the color sequence.

V. CONCLUSION

In order to accurately establish the target appearance model, the object-interference model is used to characterize the target, and context information is added to the filter in the learning phase within the framework of related filtering, which better improves the discriminative ability of the filter. The peak sidelobe ratio update strategy is adopted to avoid the problem that the linear interpolation method is difficult to select due to the fixed learning rate, and to achieve accurate tracking of the interested target. The results on the experimental sequences OTB50, OTB100, UAV123 and TC128 show that compared with other algorithms, proposed tracker is the best. The tracker has high accuracy and speed in video challenges such as lighting changes, deformations, rotations and complex backgrounds, which is of great significance to the development of intelligent video surveillance.

REFERENCES

[1] B. Tian, Q. Yao, Y. Gu, K. Wang, and Y. Li, "Video processing techniques for traffic flow monitoring: A survey," in *Proc. 14th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2011, pp. 1103–1108.

[2] J. M. B. Onate, D. J. M. Chipantasi, and N. D. R. V. Erazo, "Tracking objects using artificial neural networks and wireless connection for robotics," in *Proc. JTEC*, 2017, pp. 161–164.

[3] Y. Jing, H. Kolivand, V. E. Balas, A. Paul, and V. Ramachandran, "Research on fuzzy English automatic recognition and human-computer interaction based on machine learning," *J. Intell. Fuzzy Syst.*, vol. 39, no. 4, pp. 5809–5819, Oct. 2020.

[4] M. Brown, J. Funke, S. Erlien, and J. C. Gerdes, "Safe driving envelopes for path tracking in autonomous vehicles," *Control Eng. Pract.*, vol. 61, pp. 307–316, Apr. 2017.

[5] S. Xiao and L. Chen, "The emerging landscape of nanotheranostic-based diagnosis and therapy for osteoarthritis," *J. Controlled Release*, vol. 328, pp. 817–833, Dec. 2020.

[6] K. Nummiaro, E. Koller-Meier, and L. Van Gool, "An adaptive color-based particle filter," *Image Vis. Comput.*, vol. 21, no. 1, pp. 99–110, Jan. 2003.

[7] C. Y. Gao, P. Dong, H. Gao, and H. Q. Wang, "Application of EKF interacting multiple model algorithm in target tracking," *Machinery Des. Manuf.*, vol. 2, no. 2, pp. 284–287, 2020.

[8] D. Kai, Y. F. Ju, Y. L. Jin, L. Gang, Y. Li, and S. Qian, "Object tracking based on improved mean shift and SIFT," in *Proc. 2nd Int. Conf. Consum. Electron., Commun. Netw.*, 2012, pp. 2716–2719.

[9] S. Hare, P. H. Torr, A. Saffari, V. Vineet, M. Cheng, S. L. Hicks, and P. H. Torr, "Struck: Structured output tracking with kernels," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 10, pp. 2096–2109, Oct. 2016.

[10] Y. M. Tang, Y. F. Liu, and H. Huang, "Survey of single-target visual tracking algorithms," *Meas. Control Technol.*, vol. 39, no. 8, pp. 21–34, 2020.

[11] L. Wang, W. Ouyang, X. Wang, and H. Lu, "STCT: Sequentially training convolutional networks for visual tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1373–1381.

[12] J. Choi, H. J. Chang, T. Fischer, S. Yun, K. Lee, J. Jeong, Y. Demiris, and J. Y. Choi, "Context-aware deep feature compression for high-speed visual tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 479–488.

[13] Z. He, Y. Fan, J. Zhuang, Y. Dong, and H. Bai, "Correlation filters with weighted convolution responses," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops. (ICCVW), Oct. 2017*, pp. 1992–2000.

[14] G. Bhat, J. Johnander, M. Danelljan, F. S. Khan, and M. Felsberg, "Unveiling the power of deep tracking," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 483–498.

[15] B. Huang, T. Xu, S. Jiang, Y. Bai, and Y. Chen, "SVTN: Siamese visual tracking networks with spatially constrained correlation filter and saliency prior context model," *IEEE Access*, vol. 7, pp. 144339–144353, 2019, doi: 10.1109/ACCESS.2019.2945846.

[16] C. Ma, J.-B. Huang, X. Yang, and M.-H. Yang, "Robust visual tracking via hierarchical convolutional features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 11, pp. 2709–2723, Nov. 2019.

[17] 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.*, Jun. 2010, pp. 2544–2550.

[18] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2012, pp. 702–715.

[19] 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.

[20] T. Zhang, S. Liu, C. Xu, B. Liu, and M.-H. Yang, "Correlation particle filter for visual tracking," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2676–2687, Jun. 2018.

[21] Y. Sui, G. Wang, and L. Zhang, "Correlation filter learning toward peak strength for visual tracking," *IEEE Trans. Cybern.*, vol. 48, no. 4, pp. 1290–1303, Apr. 2018.

[22] J. Yan, J. Du, R. Young, C. Chatwin, and P. Birch, "Real time UAV tracking of fast moving small target on ground," *J. Electron. Imag.*, vol. 27, no. 5, pp. 1–12, 2018.

[23] 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, Jun. 2019.

[24] D. Yuan, X. Li, Z. He, Q. Liu, and S. Lu, "Visual object tracking with adaptive structural convolutional network," *Knowl.-Based Syst.*, vol. 194, Apr. 2020, Art. no. 105554.

[25] J. Zhang, H. Huang, J. Wang, and J. Bao, "An improved TLD real-time target tracking algorithm based on CN algorithm," *Comput. Eng. Sci.*, vol. 42, no. 7, pp. 1215–1225, Jul. 2020.

[26] D. Xiong, H. Lu, J. Xiao, and Z. Zheng, "Robust long-term object tracking with adaptive scale and rotation estimation," *Acta Autom. Sinica*, vol. 45, no. 2, pp. 289–304, Apr. 2019.

[27] M. Felsberg, A. Heyden, and N. Krüger, "DCCO: Towards deformable continuous convolution operators for visual tracking," in *Proc. Int. Conf. Comput. Anal. Images Patterns*. Cham, Switzerland: Springer, 2017, pp. 55–67.

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

[29] Y. Xiao and D. Pan, "Research on robust visual tracker based on multi-cue correlation particle filters," *IEEE Access*, vol. 8, pp. 19555–19563, Jan. 2020.

[30] C. Liu, J. Gong, J. Zhu, J. Zhang, and Y. Yan, "Correlation filter with motion detection for robust tracking of shape-deformed targets," *IEEE Access*, vol. 8, pp. 89161–89170, May 2020.

[31] J. Valmadre, L. Bertinetto, J. Henriques, A. Vedaldi, and P. H. S. Torr, "End-to-end representation learning for correlation filter based tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2805–2813.

[32] 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.

[33] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[34] Q. Wang, L. Zhang, L. Bertinetto, W. Hu, and P. H. S. Torr, "Fast online object tracking and segmentation: A unifying approach," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5050–5068.

[35] P. Voigtlaender, J. Luiten, P. H. S. Torr, and B. Leibe, "Siam R-CNN: Visual tracking by re-detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 6578–6588.

[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)*, Jul. 2017, pp. 4021–4029.

[37] 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)*, Dec. 2015, pp. 4310–4318.

[38] 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)*, Jun. 2016, pp. 1401–1409.

[39] 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.

[40] A. Lukežič, L. Č. Zajc, and M. Kristan, "Deformable parts correlation filters for robust visual tracking," *IEEE Trans. Cybern.*, vol. 48, no. 6, pp. 1849–1861, Jun. 2018.

[41] F. Wu, T. Xu, L. Wu, Y. Zhang, and X. Li, "Adaptive and complementary correlation filter with dynamic contextual constraints," *IEEE Access*, vol. 8, pp. 141895–141909, Aug. 2020.

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

[43] H. Possegger, T. Mauthner, and H. Bischof, "In defense of color-based model-free tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2113–2120.

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

[45] Q. Zhao and P. Xie, "Object tracking algorithm in infrared seeker based on multi-view feature collaborative fusion," *J. Chongqing Univ. Posts Tele-Commun., Natural Sci. Ed.*, vol. 32, no. 4, pp. 639–647, 2020.

[46] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A bench-mark," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2411–2418.

[47] 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, doi: 10.1109/TPAMI.2014.2388226.

[48] 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.

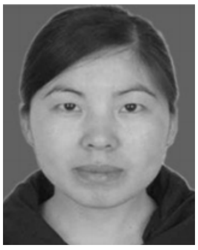
[49] P. Liang, E. Blasch, and H. Ling, "Encoding color information for visual tracking: Algorithms and benchmark," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5630–5644, Dec. 2015.

[50] A. Bibi, M. Mueller, and B. Ghanem, "Target response adaptation for correlation filter tracking," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Cham, Switzerland: Springer, Oct. 2016, pp. 419–433.

[51] Q. Hu, Y. Guo, Y. Chen, J. Xiao, and W. An, "Correlation filter tracking: Beyond an open-loop system," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, London, U.K., Jul. 2017, pp. 152.1–152.12.

YIN CUI received the B.E. degree in electrical engineering and automation from QingGong College, North China University of Science and Technology, Qinhuangdao, Hebei, China, in 2018, where he is currently pursuing the master's degree.

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



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

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