IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 17 August 2022, accepted 7 September 2022, date of publication 8 September 2022,
date of current version 28 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3205406

RESEARCH ARTICLE

Multifeature Fusion Tracking Algorithm Based on Self-Associative Memory Learning Mechanism

HONGGE REN[1], JINGJING QIAO[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, Hebei 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 A multi-feature fusion tracking algorithm updated with a self-associative memory learning mechanism is proposed to address the problems of short-time disappearance, re-emergence of the target and instability of single features in the kernelized correlation filtering algorithm. When extracting features, directional gradient histogram features, color features, and scale invariant features are fused instead of single features to collect more features of the target and increase the feature robustness. In the detection stage, the bimodal detection is proposed to judge whether the target model needs updating. Bimodal detection is used to judge the maximum target response in the search domain and predict the location of the target in the next frame. The self-associative memory learning mechanism was added into the updating template, and the original algorithm framework was improved to cope with the change of target model. The new algorithm update is biogenic, can recover fragment information, deal with complex and changeable tracking situation. Simulation experiments were conducted on the OTB50, OTB100, and UAV123 video datasets for the classical and new algorithms. The simulation verified that the proposed tracking algorithm has a high success and accuracy rate, which has research value. The tracking success rate improved by 23.6% and the accuracy rate improved by 18.8%.

INDEX TERMS Self-associative memory mechanism, target tracking, correlation filter, template update, multi-feature fusion.

I. INTRODUCTION

The field of computer vision [1] includes many types of techniques researched for different application scenarios, and target tracking algorithms are one of the research directions used in this field for video content analysis. Target tracking technology is now used in many vision fields, such as virtual reality, intelligent surveillance, human-computer interaction and so on. In most scenarios, the conditions can be complex, such as target occlusion, light change, target scale change, and target pose change, which can render the tracking algorithm challenging. Therefore, it is important to develop a tracking algorithm that is adopted to real-life scenarios with a high success rate and accuracy.

The essence of target tracking is to analyze and predict the possible location of the target in next the adjacent frame based on known information about the current frame. Tracking algorithms can be divided into two categories, depending on the appearance model: generative and discriminative. Generative tracking [2] learns online to model directly and then searches for the closest candidate domain to the target with the help of the model to determine the next position of the target. However, the generative class of algorithms does not consider background information regarding the target, which is prone to tracking failure. To improve the algorithm tracking success rate, extensive research has been conducted on the use of background information and discriminative tracking has been proposed. Discriminative transforms tracking into a binary classification problem [3], collecting background information and target information samples to train

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

the classifier and obtain the target's location. The discriminative model considers both target and background information, which has obvious advantages over existing tracking methods. As a result of the efforts of researchers, a correlation filter (CF) [4] was successfully introduced into the field of visual tracking.

The classical algorithms of CF-based trackers are as follws: Henriques *et al.* [5] proposed a circulant structure with kernels (CSK) to improve the computational speed, but the grayscale features used can only adapt to simple environments and are susceptible to complex image backgrounds and similar colors of the target backgrounds. Henriques *et al.* optimized CSK by replacing the Histogram of Oriented Gradient (HOG) features with grayscale and proposed a kernelized correlation filter (KCF) algorithm [6], which has the problems of not adapting to large target movements and single feature instability that have not been solved. The color names (CN) tracker [7] uses color attributes in the filter tracking algorithm and adopts an adaptive dimensionality reduction strategy to reduce eleven dimensional color features to two dimensional, which improves the algorithm performance while ensuring efficient tracking. The scale adaptive multiple feature (SAMF) [8] tracker based on multi-feature fusion simultaneously fuses the original image grayscale information, color attributes, and HOG multiple features to obtain more robust results. Danelljan *et al.* proposed a discriminative scale space tracker (DSST) [9] to study the scale problem and propose a solution algorithm. Balancing the weight ratio of the scale and location filters still needs to be studied. For complex illumination, background information, and the use of contextual information [10], [11], [12], [13] have been studied to propose the multiple kernelized correlation filters (MKCF) tracking algorithm, to make full use of the discriminative invariance of the power spectrums (power spectrums) of various features and further improve the performance of the algorithm. The multi-feature fusion algorithm STAPLE-CA uses two complementary features [14], HOG and global color histogram, to model the appearance of the target, and utilizes the inherent characteristics of each feature to be transformed into a ridge regression problem solver. The attentional correlation filter network for adaptive visual tracking [15] selects appropriate correlation filters that enable the algorithm to adaptively adjust the selected features according to the tracking situation.

With the development of computer vision technology, deep learning has been used for target tracking to enhance the tracking effect. Its representative tracking algorithms are twinnetwork based tracker, recurrent tracker, attention-based classifier, convolutional neural network-based tracker, and unit learning-based tracker [16], where some deep learning tracks are improved on the basis of correlation filtering. Zhang *et al.* [17] proposed Siamese FC for target tracking, using a convolutional layer fusion strategy for target tracking and obtained good feature representation. Based on this, Mueller *et al.* [18] can be considered as a special correlation filter added to the twin network, and back propagation is derived by defining the network output to output a probabilistic heat map of target locations. Liu *et al.* [19] proposed a dynamic twin network that enables efficient online learning from historical frames by rapidly transforming the learning model to learn target appearance changes and background suppression, and then integrate the network output using multi-level deep features through elemental multilayer fusion. The empirical verification of deep learning joins tracking to achieve good performance. However, owing to the complexity of deep learning networks, the algorithm framework is complex to build and the computational speed decreases. It is worth studying how to keep the algorithm accurate and improve the computational speed at the same time, so that the algorithm can accurately target real-life situations.

To understand the human brain's processing of complex information, theoretical studies on memory learning mechanisms have emerged one after another. The human brain has been found to rely heavily on associative memory mechanisms, which is a hot research topic at the intersection of cognitive neuroscience and computer science. Associative memory can be divided into self-associative and hetero-associative memory, depending on the stimulus source. Among them, the self-associative memory mechanism has been an active research topic and has received extensive attention from scholars in the fields of cognitive neuroscience and neural networks [20]. Post first proposed the biological associative memory model [21]. Sui *et al.* proposed the concept of an associative machine to store entities represented by bit patterns in a distributed manner and retrieve all entities from a part of them [22]. Based on the signal processing characteristics of the human brain, a self-organizing feature mapping neural network was proposed [23], which mimics the self-organization process of the brain for network tuning of network neurons. Zhang *et al.* proposed a neurobiologically based concept of pattern association and self-association [24]. Pattern association plays an important role in the frontal lobe of the eye fossa cortex. This mechanism also acts throughout the cerebral cortex and plays an important role in visual memory recall processes. Subsequently, Kelley and Cassenti [25] gave a typical structure of the pattern association memory network and explained the operation of the pattern associators.In 2006, Cai *et al.* proposed a computational theory of hippocampal function, suggesting that the hippocampus uses self-association to form situational memories [26]. The self-associative memory model is also called an attractor neural network. This model enables memory and efficient memory retrieval, and memory is stored in recurrent synaptic connections between the network neurons. During memory storage, each memory is represented by a neural activity pattern. When a memory fragment is provided to the model, the self-associative network [27] can recall associated memory from the network. A self-associative network with fast synaptic plasticity can learn from each memory in a single trial. Owing to its fast learning ability, this type of network is well suited for the storage of situational memories.

Self-associative memory learning mechanisms mimic the information storage, recall, and recognition functions of the human brain, and are widely used in image processing.

Our main contributions are as follows:

1) HOG features, color features, and scale invariant features are fused instead of single feature extraction features.
2) Double-peak detection is added to the detection module to increase the detection accuracy.
3) The self-associative memory learning mechanism is added to the update template to improve the original algorithm framework.

II. RELATED WORK

A. CORRELATION FILTERS

The advantage of the KCF algorithm is that a circular matrix is used to densely sample the samples, and the introduction of the kernel method makes the correlation filtering algorithm more robust and capable of handling nonlinear classification problems. Many algorithms have selected KCF as the basic algorithm framework to improve it and propose algorithms that can adapt to complex environment tracking. Mueller *et al.* [14] proposed the CACF tracking algorithm, which uses regularization to strengthen the target and weaken the background information. The disadvantage is that the tracker can make judgment errors, leading to tracking failures. Henriques *et al.* [6] used multi channel features and kernel tricks in a DCF algorithm to improve the target model. Lukezic *et al.* [28] used a tracker to generate spatial reliability maps and proposed CSRDCF tracking, which led to improved target tracking accuracy. Dai *et al.* [29] proposed an adaptive spatial regularization correlation filtering algorithm (ASRCF), which makes the filter better adapted to the target morphology. Considering that the kernelized filter can be computed quickly, KCF filtering is selected as the basis of this study, and the maximum response value is obtained by double-peak detection for tracking using multi-feature modeling.

B. FILTERING UPDATE MECHANISM

The KCF algorithm uses linear weighting for template update, and updating after each detection leads to computational degradation of the algorithm and redundant information. Some scholars use the interval method of updating, because we cannot tell whether the target is deformed or not; when the target is obscured or deformed in large cases, updating again will lead to tracking errors. Bolme *et al.* [30] proposed PSR detection, which measures whether to perform an update by the detection ratio. Wang *et al.* [31] proposed the APCE calculation, which selects the response value of the highest confidence in the next. Huang *et al.* [32] proposed ARCF to process a response map to improve the detection accuracy. It is clear from previous research results that the update mechanism is very important for the algorithm. Choosing an appropriate update mechanism leads to improved algorithm performance. By continuously studying human brain memory, we found that the self-associative memory mechanism has fast learning ability and is suitable for use in situational learning. The literature [33] investigates the self-associative memory learning mechanism from three aspects: the learning algorithm, architecture and application area. In [34], the self-associative memory algorithm was used to solve problems in image scaling. The self-associative memory learning mechanism simulates the information storage, recall, and recognition functions of the human brain, and is widely used in image processing [35]. In this study, a bionic algorithm was selected to incorporate the self-associative memory mechanism into the tracking algorithm.

After learning the KCF algorithm, it was found that its single feature with HOG alone is not sufficient to cope with complex environment tracking. To address this problem, multiple features were proposed instead of single features to improve the tracking accuracy of the algorithm. When the target is occluded or in a complex environment, the tracking sample may contain useless information. If information is accumulated for non-targets, it will lead to an algorithm tracking failure. Inspired by human memory learning, a self-associative memory learning mechanism is introduced to improve the update module in the original tracking framework, which allows the algorithm to have human-like memory to deal with the problem that the algorithm disappears and recreates the target for a short time.

III. ALGORITHM DESIGN

In response to the problems of single feature tracking, over reliance on the appearance model, and inappropriate updates leading to model drift and inability to cope with complex environments in KCF algorithm, a multi-feature fusion tracking algorithm updated with a self-associative memory learning mechanism is proposed and named the SMFCF algorithm. The new algorithm ensures the robustness of sample detection and improves tracking accuracy. The self-associative memory learning mechanism has the features of fast learning, large storage space, and the recovery of fragmented images, which can solve the problem of transient disappearance and reappearance. Therefore, the new algorithm uses a multi-feature fusion method to extract the target features, and obtaining more information can increase the robustness of the target appearance model. Combining multifeature fusion and self-associative memory learning mechanisms, the feature and update modules of the traditional tracking framework are improved to enhance the tracking effect.

A. MULTI-FEATURE MODELING

Although the HOG feature can adapt to different lighting situations, it is a local search and lacks rotational invariance, which is still difficult in the face of complex tracking environment challenges. Therefore, the color, HOG, and SIFT features were extracted from the image, and the three features were fused to obtain the maximum response value for tracking.

1) HISTOGRAM OF ORIENTED GRADIENTS FEATURE

The target tracking map is segmented into multiple small blocks to extract features, resulting in an HOG that is resistant to illumination changes and geometric deformations but is not applicable to large scale changes. The HOG features [36] can cope with small physical changes in passersby when the general pose of the human body is constant. First-order discrete differential equations for pixel points in the horizontal and vertical directions.

$$\begin{cases} G_x(x,y) = I(x+1,y) - I(x-1,y) = A \\ G_y(x,y) = I(x,y+1) - I(x,y-1) = B \end{cases} \tag{1}$$

In the above equation, I(x,y) is the input pixel value, and G_x (x,y) and G_y (x,y) are the horizontal and vertical gradients of (x,y), respectively. A amplitude and direction of the gradient equations.

$$G(x,y) = \sqrt{A^2+B^2} \tag{2}$$

$$\alpha(x,y) = \tan^{-1} A\big/B \tag{3}$$

2) COLOR (CN) FEATURES

CN features are target-oriented to perform global acquisition, and the common extraction method is the color histogram, where the color distribution is calculated for the target appearance map to obtain the frequency information of individual colors. Image appearance and shape changes did not affect the extraction of CN features, and showed scale invariance. Hue, saturation and brightness are common characteristics of the human eye to observe things, so (hue, saturation, value) HSV space is used to extract color to the target.

3) SCALE INVARIANT TRANSFORMATION FEATURES

SIFT features are local features that describe the target. The extreme value point was found in the domain and scale, and the scale, position and rotation of this point were also captured. First, the convolution kernel is smoothed on the source image to construct the target scale space L.

$$\mathrm{L}(\mathrm{x},\mathrm{y},\sigma) = \mathrm{G}(\mathrm{x},\mathrm{y},\sigma) \otimes \mathrm{I}(\mathrm{x},\mathrm{y}) \tag{4}$$

$$\mathrm{G}(\mathrm{x},\mathrm{y},\sigma) = \exp\frac{\mathrm{x}^2+\mathrm{y}^2}{2\sigma^2}\Big/2\pi\sigma^2 \tag{5}$$

In Eq.(5), the scale coordinate σ expresses the smoothness of the image.

Next, two adjacent Gaussian scale maps were subtracted to obtain the difference of gaussian (DOG). The local maxima are searched in the DOG, and the adjacent two-layer maps are compared with the key points to determine the function maxima. Finally, the descriptor is calculated using a Gaussian function.

4) INTEGRATION OF MULTIPLE FEATURES

Select the video image at moment t, the target for different feature acquisitions, to obtain m_t^{HOG}, m_t^{HSV} and m_t^{SIFT}. The three extracted features are fast fourier calculation, to obtain the corresponding response values r_t^{HOG}, r_t^{HSV} and r_t^{SIFT}. After calculating the response value linear weighting to obtain the final response value r_t, the formula is as follows The formula is as follows.

$$\mathrm{r_t} = \alpha \mathrm{r_t^{HOG}} + \beta \mathrm{r_t^{HSV}} + \vartheta \mathrm{r_t^{SIFT}} \tag{6}$$

In the formula,$\mathrm{r_t}$ is the integrated response value, α, β and ϑ are the weight parameters of HOG, HSV, and SIFT response values respectively. $r_{t.max}$ is considered as the maximum value of the integrated response value, which is the most likely location of the next position of the target for subsequent tracking. $\alpha + \beta + \vartheta = 1$,,$\alpha = 0.5$, $\beta = 0.3$ and $\vartheta = 0.2$.

B. CONSTRUCTING FILTERS

The KCF algorithm constructs the sample set via cyclic shifting after selecting the target image, associates the shifted samples with the filter, learns the RLS classifier, solves it to obtain the filter template, and filters it in the candidate area until the desired target is obtained. The objective function was obtained from the circular sample set training.

$$f(x) = W^T x \tag{7}$$

Find the minimum error sum of squares for Equation 1.

$$\underset{W}{min} \sum\nolimits_i (f(x_i){-}y_i)^2 + \lambda \, \|W\|^2 \tag{8}$$

W is the weight matrix, $\mathrm{x_i}$ is the sample set, λ is the regularization parameter, $\mathrm{f(x_i)}$ is the classifier, and $\mathrm{y_i}$ is the actual output sample set. If the samples are linearly divisible, the derivative of Eq. (8) is set to zero. The ridge regression is then solved.

$$W = (X^T X + \lambda I)^{-1} X^T Y \tag{9}$$

Diagonalization properties of circular matrices:

$$\mathrm{X} = \mathrm{Fdiag}\left(\hat{\mathrm{X}}\right)\mathrm{F^H} \tag{10}$$

Combining equations (9) and (10):

$$\hat{\mathrm{W}} = \hat{\mathrm{x}}^* \cdot \hat{\mathrm{y}} \big/ (\hat{\mathrm{x}}^* \cdot \hat{\mathrm{y}} + \lambda) \tag{11}$$

In the above equation, $\hat{\mathrm{x}}^*$ is the complex conjugate of $\hat{\mathrm{x}}$ and · is the corresponding element multiplication.

C. TWIN-PEAK DETECTION

When the target disappears or only partially appears in the frame, KCF cannot obtain an accurate tracking response value in the detection phase. When the target disappears for a short period of time or when a similar tracked object appears, the response image shows a multi peaked response [37], as shown in Figure 1. If the wrong response value is selected, it leads to tracking failure. Therefore in the detection module, multi-peak detection of the response image is added.

In Figure 1, (b) and (d) are the response plots corresponding to the different cases when (a) and (c) are targets tracked, respectively. As can be seen from the plots, the corresponding maximum response values are not singular after the short disappearance of the tracked target. In this case, alternative

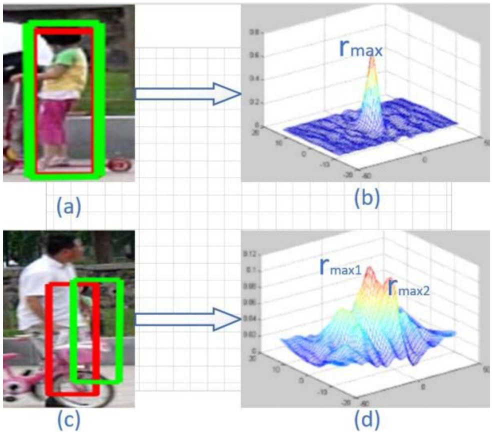


FIGURE 1. **Response image comparison chart.**

sample ranges are redetected to obtain the possible locations of the target. Because the peak response is the main influencing factor in determining the location, the peaks are sorted from the largest to smallest, and the first two peak values in the response map are saved. The judgment is made using the following formula.

$$\Gamma = \begin{cases} P(r_{max1}), & \dfrac{r_{max1}}{r_{max2}} > \zeta \\ P(r_{max1}, Pr_{max2}), & \dfrac{r_{max1}}{r_{max2}} \leq \zeta \end{cases} \tag{12}$$

In formula (12), Γ is the final maximum response value used for tracking, $P(r_{max1})$ is the highest peak response position, and $P(r_{max1}, Pr_{max2})$ is the pooled region of the two peak positions. If the ratio of the highest peak to the second highest peak is greater than ζ, the output is based on the highest peak position, and vice versa, the output is based on the pooled area of the two peak positions.

The set of alternative samples consists of an image and its cyclically shifted samples at time t. The formula is:

$$\mathrm{Z_i = P^iZ} \tag{13}$$

P is the permutation matrix and Z is the initial input sample. In the detection module, the responses of the selected region and the filter are calculated.

$$\mathrm{f(Z) = (K^Z)^T}\alpha \tag{14}$$

In the frequency domain the equation is as follows.

$$\mathrm{K^Z} = \Phi(\mathrm{Z}) \cdot \Phi(\mathrm{Z}) = \mathrm{C(k^{xz})} \tag{15}$$

$\mathrm{K^Z}$ is the kernel matrix of the training sample and the sample to be tested.

D. THE UPDATE MODULE

KCF using only a linear weighted update is not sufficient to cope with complicated tracking situation, and the self-associative memory learning mechanism is used to update the strategy to ensure tracking accuracy and improve the robustness of the KCF algorithm. The self-associative memory learning mechanism is a method to simulate human brain memory using neural networks [38]. When there is a similar fragment or a missing fragment, self-association can retrieve the relevant memory from the repository and recover the original complete features, to complete the delayed matching and recognition tracking tasks. In summary, the self-associative memory learning mechanism was selected for updating, which made the algorithm more robust.

The learning memory of things is abstracted from concrete objects, and the sample vector pairs $(\mathrm{x^p}, \mathrm{y^q})$ are stored in the self-associative memory matrix A. An association matrix associative memory model stores two corresponding vectors, in the form of an association matrix. As shown in Equation (16).

$$\mathrm{A} = \sum\nolimits_{\mathrm{P}} \mathrm{C_p} \left(x^p\right)^{\mathrm{T}} y^p = \sum\nolimits_{\mathrm{P}} \left\|x^p\right\|^{-2} \left(x^p\right)^{\mathrm{T}} y^p \tag{16}$$

When there is a sample x^γ input, the output sample is shown in Equation (17).

$$\mathrm{y}^\gamma = \mathrm{Ax}^\gamma = \mathrm{C}_\gamma \left(\mathrm{y}^\gamma\right)^{\mathrm{T}} \mathrm{y}^\gamma \mathrm{x}^\gamma + \sum\nolimits_{\mathrm{p}\neq\gamma} \mathrm{C_p} \left(\mathrm{x^p}\right)^{\mathrm{T}} \mathrm{y^p} \mathrm{x}^\gamma \tag{17}$$

If the input samples are orthogonal to each other, the formula changes.

$$\mathrm{(x^i)^T x^j} = \begin{cases} \left\|\mathrm{x^i}\right\|^2, & \mathrm{i = j} \\ 0, & \mathrm{i \neq j} \end{cases} \tag{18}$$

$$\mathrm{y}^\gamma = \left\|\mathrm{x}^\gamma\right\|^2 \mathrm{y}^\gamma \mathrm{C}_\gamma \tag{19}$$

For Y=AX, $\mathrm{A = YX^{-1}}$ is only available when X is a full rank matrix. The mathematical formula is more idealized, and there are always unknown factors that interfer with the actual problem. Assume that W is an s×q matrix of rank r, then a generalized inverse of W is a q×s array G, such that X=GY is a solution of the equation WX=Y, so that $\mathrm{G = W^+}$, which is the generalized inverse of W.

Let Y_k be an n×k array, which represents the ith output sample; X_k, which is an s×k array, represents Y_k with the corresponding input sample. The generalized inverse formula for the input sample $X_k = (X_{k-1} \left| \mathrm{x^k} \right.)$ is Equation (20).

$$\mathrm{X_k^+ = (Z_k/b_k)_s^{(k-1)s}} \tag{20}$$

where X_k is the input kth sample and $\mathrm{X_{k-1}^+}$ is the generalized inverse of X_{k-1}.

$$\mathrm{X_k^+ = (X_{k-1}^+ - X_{k-1}^+ x^k b_k/b_k)_s^{(k-1)s}} \tag{21}$$

The self-associative memory matrix $\mathrm{A_k}$ expression is equation (22).

$$\mathrm{A_k = A_{k-1} - A_{k-1} x^k b_k + y^k b_k} \tag{22}$$

$$\begin{cases} \mathrm{b_k = c_k^+ + (1 - c_k^+ c_k)(1 + d_k^T d_k)^{-1} d_k^T d_k^+ X_{k-1}} \\ \mathrm{c_k = x^k - X_{k-1} d_k} \\ \mathrm{d_k = X_{k-1}^+ x^k} \end{cases} \tag{23}$$

Learning the first set of sample vectors (x^1, y^1), the formula is as follows.

$$X_1^+ = ((x^1)^T(x^1)^{-1})(x^1)^T \tag{24}$$

$$A_1 = y^1((x^1)^T(x^1)^{-1})(x^1)^T \tag{25}$$

Starting from equations (24) and (25), the storage is continuously iterated until it stops including all features of the sample, thus forming self-association matrix A. The module is updated using equations (17) to (19), and when a defect occurs in the tracker, the stored features are called or the search is delayed by virtue of the existing information to ensure the accuracy of the tracking.

E. ALGORITHMIC FRAMEWORK

From the above theoretical derivation and formula learning, Flowchart 2 of the SMFCF algorithm is shown as follows: initialize the target and input the first frame image. HOG, CN and SIFT features were extracted from the image. After constructing the robust appearance model, the response values of the three features were calculated separately and accumulated using certain weights to obtain the integrated response values. Bimodal detection was performed on the response map to determine the maximum integrated response value for the target detection. The self-associative memory mechanism was added to the update module to complete the main framework of the new algorithm. Determine whether the video is in the last frame, and end if it is. Otherwise, multi-feature extraction is performed again, and the above process is repeated until the end of the video playback to complete the tracking task.

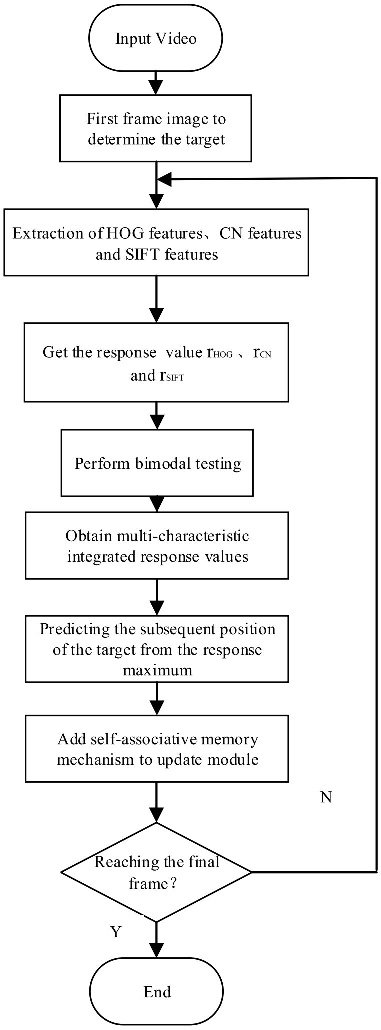

Input Video
First frame image to determine the target
Extraction of HOG features、CN features and SIFT features
Get the response value r_{HOG}、r_{CN} and r_{SIFT}
Perform bimodal testing
Obtain multi-characteristic integrated response values
Predicting the subsequent position of the target from the response maximum
Add self-associative memory mechanism to update module
N
Reaching the final frame?
Y
End

FIGURE 2. SMFCF algorithm flow chart.

IV. ANALYSIS OF SIMULATION RESULTS

CSK [5], KCF [6], DSST [9], STAPLE_CA [14], ACFN [15] and the new algorithm are compared on the OTB50 [39], OTB100 [40] and UAV123 [41] datasets to verify whether the proposed algorithm can achieve the desired criteria. The experimental device is an Intel(R) Core(TM) i7-10870H CPU @ 2.20GHz 2.21GHz, simulated with MATLAB R2019b software. The parameters were set as follows: linear adaptive rate of 0.06, Gaussian kernel standard deviation of 0.7, regularization of 10^{-3}. feature fusion parameters $\alpha = 0.5$, $\beta = 0.3$, $\vartheta = 0.2$.

A. ANALYSIS OF THE RESULTS OF THE OTB50

The dataset provides both grayscale and color image sequences, and the sample categories in the set contain 11 common problem cases in tracking tasks. It contains illumination variation(IV), scale change(SV), occlusion(OCC), deformation(DEF), motion blur(MB), fast motion(FM), in plane rotation(IPR), low resolution(LR),out of field(OV), background clutter(BC), and out of plane rotation(OPR). Each of these image sequences corresponds to two or more difficult problems and is often used to check whether the algorithm tracking is accurate.

Graphs of the success rate and accuracy of SMFCF and the other five algorithms were obtained by simulation, as shown in Figure 3.

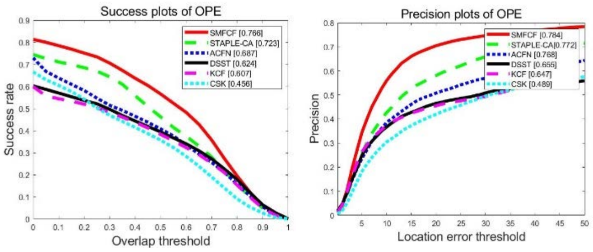


FIGURE 3. Success plots and precision plots of algorithms on OTB50.

From Figure 3 and Table 1, it can be seen that the SMFCF algorithm is in the leading position, and its success rate of accuracy is the best among the selected algorithms. the success rate of SMFCF is 0.766 and the success rate of KCF is 0.607, with a difference of 0.159; the accuracy rate of SMFCF is 0.784 and the accuracy rate of KCF is 0.647, with a

TABLE 1. Data summary of various algorithms on OTB50.

Percentages	SMFCF	STAPLE-CA	ACFN	DSST	KCF	CSK
Success	**0.766**	0.723	0.687	0.624	0.607	0.456
Precision	**0.784**	0.772	0.768	0.655	0.647	0.489

difference of 0.137. The data confirm that the improved algorithm is more accurate for tracking targets, and the success possibility is improved.

B. ANALYSIS OF THE RESULTS OF THE OTB100

The dataset contains 100 videos, providing both grayscale and color image sequences, and the sample categories in the set contain common 11 problem cases in tracking tasks. It is often used as a dataset for tracking algorithms for validation. The experimental results of the six algorithms on this dataset are shown in Figure 4.

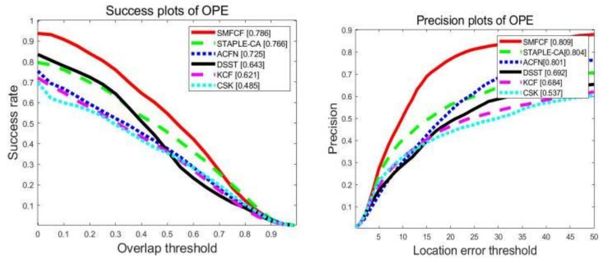


FIGURE 4. Success plots and precision plots of algorithms on OTB100.

The experimental data after running on 100 videos, are shown in Fig. 4 and Table 2. It can be seen that the success rate and precision rate of SMFCF are 0.786 and 0.809, respectively, while the success rate of KCF has a precision rate of 0.621 and 0.684, respectively. The new algorithm improved the success rate of KCF by 26.6% and a precision rate of 13.3% in terms of performance.

TABLE 2. Data summary of various algorithms on OTB100.

Percentage	SMFCF	STAPLE-CA	ACFN	DSST	KCF	CSK
Success	**0.786**	0.766	0.725	0.643	0.621	0.485
Precision	**0.809**	0.804	0.801	0.692	0.684	0.537

C. ANALYSIS OF THE RESULTS OF THE UAV123

The image sequences provided by the UAV123 dataset are all colored. This dataset is mainly used in the field of UAV target tracking research, in which image data are captured by UAV photography. The filming angles of the dataset were highly variable, and the backgrounds are relatively clean. In addition, the tracking targets in the video sequences are small and the videos are too long, which tends to degrade the target model and requires a high performance of the tracking algorithm.

The experimental results of multiple algorithms on the UAV123 dataset are shown in Figure 5. Table 4 shows the specific values of success rate and precision rate of each algorithm. The numerical comparison shows that the SMFCF tracking success rate is 0.607, and the accuracy rate is 0.675, which ranks first compared with other algorithms. The SMFCF algorithm incorporates SIFT features and can successfully lock the target even after the target has changed. There is experimental data to know that proposed SMFCF algorithm builded an effective appearance model for tracking and detection in long tracking videos of small targets as a way to ensure algorithm accuracy.

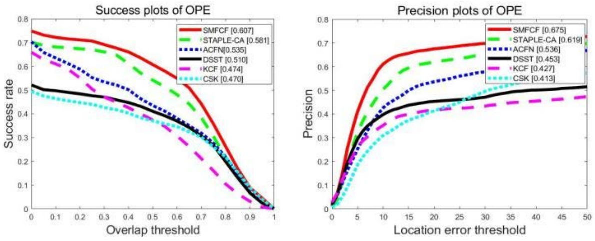


FIGURE 5. Data summary of various algorithms on UAV123.

D. COMPARISON OF THE NEW ALGORITHM WITH KCF

Figure 6 shows the graphs obtained after running SMFCF and KCF on the 20 videos. In the figure, the success rate of the SMFCF algorithm is 0.691, which is a 0.132 improvement on the original algorithm, and the accuracy rate is 0.851, which is an improvement of 0.135 for the original algorithm.

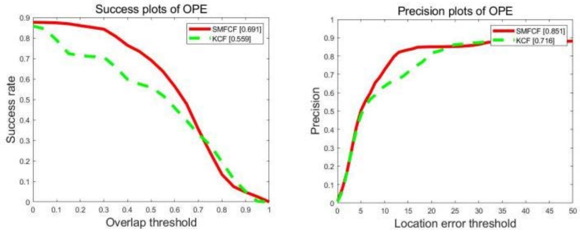


FIGURE 6. Success plots and precision plots of KCF improvement.

In summary, the comparison shows that SMFCF improves the KCF algorithm, and the robustness and success possibility are greatly improved. Compared with the STAPLE-CA algorithm, which also has feature fusion, the new algorithm exhibits a slightly higher performance in each case. Compared to CSK and DSST, the addition of multi-feature fusion and self-associative memory learning mechanisms was found to improve the speed and accuracy of the algorithm. Compared with the ACFN by the attention mechanism, it is concluded that the self-associative memory learning mechanism in the new algorithm deepens the memory of the target more

TABLE 3. Video of some experiments in 11 tracking situations.

	IV	OPR	SV	OCC	DEF	MB	FM	IPR	OV	BC	LR
Box	√	√	√	√		√		√	√	√	√
Car1	√		√			√	√			√	√
Couple		√	√		√		√			√	
Dancer		√	√		√			√			
Doll	√	√	√	√				√			
DragonBaby		√	√	√		√	√	√	√		
Girl		√	√	√				√			
Humman6		√	√	√	√		√		√		
Liquor	√	√	√	√		√	√		√	√	
Matrix	√	√	√	√			√	√		√	
Shaking	√	√	√					√		√	
Tiger1	√	√		√	√	√	√	√			

TABLE 4. Data summary of various algorithms on UAV123.

Percentage	SMFCF	STAPLE-CA	ACFN	DSST	KCF	CSK
Success	**0.607**	0.581	0.535	0.510	0.474	0.470
Precision	**0.675**	0.619	0.536	0.453	0.427	0.413

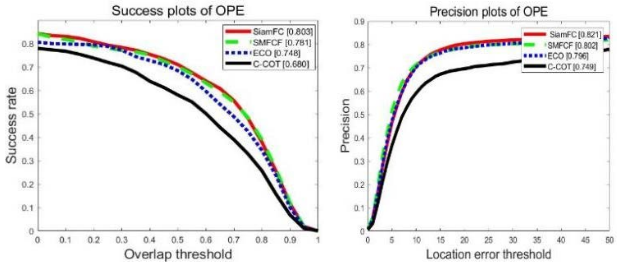
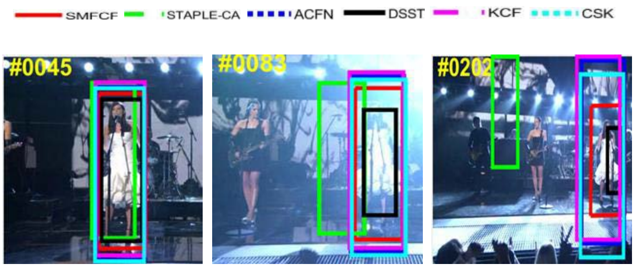


FIGURE 7. Comparison of deep learning algorithm and SMFCF algorithm.

TABLE 5. Data summary between deep learning algorithm and SMFCF algorithm.

Percentage	SiamFC	SMFCF	ECO	C-COT
Success	**0.803**	0.781	0.748	0.680
Precision	**0.821**	0.802	0.796	0.749



(a) (b) (c)

FIGURE 8. Rotating video multi-algorithm comparison.

than increasing the attention of the target enables the algorithm to capture the target quickly.

E. COMPARISON OF SOME DEEP LEARNING TRACKING ALGORITHMS WITH SMFCF ALGORITHMS

With the continuous development of deep neural networks, deep learning With the continuous development of deep neural networks, deep learning tracking algorithms appear. Danelljan *et al.* [42] proposed Continuous Convolution Operator Tracker (C-COT). In order to improve efficiency, he proposed an Efficient model update strategy [43], Efficient Convolution Operators (ECO). This tracking algorithm simplifies the feature information of C-COT to achieve the purpose of speed improvement. Bertinetto *et al.* proposed SiamFC [27], which opened the way for deep learning methods to gradually overtake correlation filtering methods.In summary, these three deep tracking algorithms are selected for experimental simulation on the OTB dataset with the algorithm proposed in this paper.

The simulation results are shown in Figure 5. Although SMFCF algorithm is not the fastest, it ranks the second among the four algorithms. The difference between the success rate of SMFCF algorithm and the success rate of the first place is 0.022, which is not big and needs to be improved.

F. ANALYSIS OF SOME SPECIFIC VIDEO RESULTS

In the video where the experiment was conducted, three groups were selected for comparison to demonstrate the performance of the algorithm under different tracking situations. The first two groups are multi-algorithm comparisons and the last group is a comparison chart before and after the KCF algorithm improvement.

The video is of a singer performing on stage, and at the beginning as shown in frame 45, all algorithms can determine the tracker. Then as the stage lights change from dark to

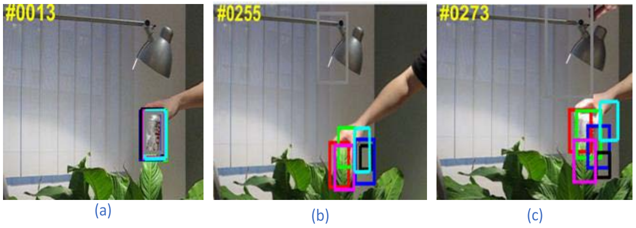


FIGURE 9. Beverage video multi-algorithm comparison.

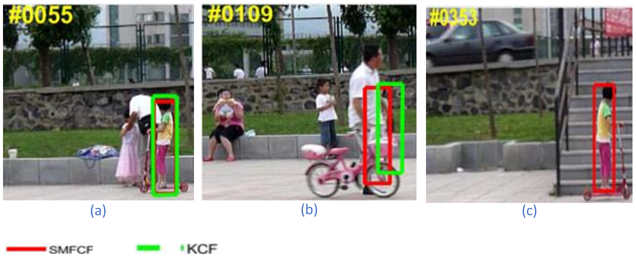


FIGURE 10. Comparison of algorithms for mobile girl video.

bright, as shown in frame 83, the singer is exposed, and the intense lighting causes some of the algorithms to fail in tracking. Finally, after the stage lights return to normal again, as shown in frame 202, the SMFCF algorithms always track the target accurately, but some algorithms show a tracking drift.

The video shows a person moving continuously with a cola drink in hand to verify the tracking success of the algorithm. In frame 13, the algorithm was successful in locating the tracking target. At frame 255, the drink is artificially placed behind the plant, presenting a situation where the tracking object is obscured, at which point the SMFCF accurately locates the drink. The video person then takes the drink out from behind the plant and finds that the multi-algorithm tracking occurs in a chaotic scene, whereas the SMFCF algorithm is able to successfully track the cola drink.

This video shows a girl playing freely using a pulley car. The KCF algorithm before and after the improvement is used to track the girl in this video to verify its superiority of the improved algorithm. Both old and new algorithm successfully tracked a target when there was no occlusion. Around frame 109, a pedestrian cart walks and blocks the target girl with her body. The old and new algorithm tracking appear different; SMFCF can determine the girl's position, but KCF's tracking frame contains most of the background information, which affects tracking. Waiting for the pedestrian to walk by, the girl driving the scooter undergoes a position change, as shown in frame 353, presenting the new algorithm tracking success and KCF tracking failure. Observing three sets of photos with different situations and tracking targets, the SMFCF algorithm demonstrates its robustness and success in dealing with lighting changes, object reappearance after brief occlusion, and scale changes.

V. CONCLUSION

The SMFCF algorithm is a multi-feature fusion and self-associative memory learning mechanism update. For the shor-t time disappearing and reappearing problem, the self-asso- ciative memory learning mechanism is introduced as an update; facing the complex tracking background situation, the algorithm uses multi-feature fusion for appearance modeling to ensure tracking stability. The experimental results after the simulation indicate that the tracking is more accurate for the short-time disappearing and reappearing problem of the target, which is convenient for the tracking algorithm to be used in practice.

REFERENCES

[1] Q. Shu, H. Lai, L. Wang, and Z. Jia, "Multi-feature fusion target relocation tracking based on correlation filters," *IEEE Access*, vol. 9, pp. 28954–28964, 2021.

[2] M. Liang, X. Wu, Y. Wang, Z. Zhu, B. Cao, and J. Xu, "Multi-model and multi-expert correlation filter for high-speed tracking," *IEEE Access*, vol. 9, pp. 52326–52335, 2021.

[3] M. Islam, G. Hu, W. Dan, C. Lyu, and Q. Liu, "Correlation filter based moving object tracking with scale adaptation and online re-detection," *IEEE Access*, vol. 6, pp. 75244–75258, 2018.

[4] M. Yu, Y. Zhang, Y. Li, Z.-L. Lin, J. Li, and C. Wang, "Saliency guided visual tracking via correlation filter with log-Gabor filter," *IEEE Access*, vol. 8, pp. 158184–158196, 2020.

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

[6] 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.

[7] T. Ran, L. Yuan, and J. B. Zhang, "Scene perception based visual navigation of mobile robot in indoor environment," *ISA Trans.*, vol. 109, pp. 389–400, Mar. 2021.

[8] Y. Li and J. Zhu, "A scale adaptive kernel correlation filter tracker with feature integration," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 254–265.

[9] 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.

[10] Y. Z. Zhang, Y. S. Sun, and C. K. Xia, "Multi-scene fast motion planning of robotic arm incorporating long and short term memory mechanism," *Control Decis.*, vol. 35, no. 12, pp. 2968–2975, 2020.

[11] A. Scoboria, K. A. Wade, D. S. Lindsay, T. Azad, D. Strange, J. Ost, and I. E. Hyman, "A mega-analysis of memory reports from eight peer-reviewed false memory implantation studies," *Memory*, vol. 25, no. 2, pp. 146–163, 2017.

[12] T. Zhang and S. Yang, "Attribute topology-based analysis of human brain forgetting characteristics," *Digital Design*, vol. 6, no. 2, pp. 1–8, 2017.

[13] D. Wang and Y. Y. Yang, "A study of self-associative memory neural networks," *Comput. Technol. Develop.*, vol. 21, no. 3, pp. 109–114, 2011.

[14] 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.

[15] D. Yuan, X. Chang, P. Huang, and Q. Liu, "Self-supervised deep correlation tracking," *IEEE Trans. Image Process.*, vol. 30, pp. 976–985, 2021.

[16] W. Huang, J. Gu, X. Ma, and Y. Li, "Correlation-filter based scale-adaptive visual tracking with hybrid-scheme sample learning," *IEEE Access*, vol. 6, pp. 125–137, 2018.

[17] S. Zhang, Y. Sui, S. Zhao, and L. Zhang, "Graph-regularized structured support vector machine for object tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 6, pp. 1249–1262, Jun. 2017.

[18] M. Mueller, N. Smith, and B. Ghanem, "Context-aware correlation filter tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1387–1395.

[19] 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, 2020.

[20] Y. Liu, F. B. Zheng, and X. Y. Zuo, ‘‘Framework of cross-modal semantic mapping based on cognitive computing of visual and auditory sensations,’’ *High Technol. Lett.*, vol. 22, pp. 90–98, Mar. 2016.

[21] T. Liu, G. Wang, and Q. Yang, ‘‘Real-time part-based visual tracking via adaptive correlation filters,’’ in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 4902–4912.

[22] Y. Sui, S. Zhang, and L. Zhang, ‘‘Robust visual tracking via sparsity-induced subspace learning,’’ *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4686–4700, Dec. 2015.

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

[24] H. Zhang, S. Hu, X. Zhang, and L. Luo, ‘‘Visual tracking via constrained incremental non-negative matrix factorization,’’ *IEEE Signal Process. Lett.*, vol. 22, no. 9, pp. 1350–1353, Sep. 2015.

[25] T. D. Kelley and D. N. Cassenti, ‘‘Theoretical explorations of cognitive robotics using developmental psychology,’’ *New Ideas Psychol.*, vol. 29, no. 3, pp. 228–234, 2017.

[26] Z. Y. Cai, Y. Gao, and Z. L. Yu, ‘‘Image retrieval based on ternary convolutional neural network,’’ *J. Xian Univ. Posts Telecommun.*, vol. 6, pp. 60–64, 2016.

[27] L. Bertinetto, J. Valmadre, and J. F. Henriques, ‘‘Fully convolutional Siamese fusion networks for object tracking,’’ in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, 2016, pp. 850–865.

[28] A. Lukezic, T. Vojir, L. C. Zajc, J. Matas, and M. Kristan, ‘‘Discriminative correlation filter with channel and spatial reliability,’’ in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6309–6318.

[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)*, Jun. 2019, pp. 4670–4679.

[30] 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.

[31] 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.

[32] 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)*, Oct. 2019, pp. 2891–2900.

[33] H. H. Li, F. Q. Si, and Z. A. Xu, ‘‘A sensor fault diagnosis method based on robust self-associative neural network,’’ *Chin. J. Electr. Eng.*, vol. 32, no. 14, pp. 116–121, 2012.

[34] H. Z. Xu and H. L. Li, ‘‘Research on image scaling based on self-associative memory algorithm,’’ *Sci. Eng. Technol.*, vol. 13, no. 18, pp. 5381–5384, 2013.

[35] W. Zhang, L. Jiao, Y. Li, and J. Liu, ‘‘Sparse learning-based correlation filter for robust tracking,’’ *IEEE Trans. Image Process.*, vol. 30, pp. 878–891, 2021.

[36] D. Bouget, M. Allan, D. Stoyanov, and P. Jannin, ‘‘Vision-based and marker-less surgical tool detection and tracking: A review of the literature,’’ *Med. Image Anal.*, vol. 35, pp. 633–654, Jan. 2017.

[37] Y. F. Liu, Y. He, Q. Tian, and J. Yang, ‘‘Improved KCF tracking algorithm using outlier detection and relocation,’’ *Comput. Eng. Appl.*, vol. 20, no. 54, pp. 166–171, 2018.

[38] D. Du, Y. Qi, H. Yu, Y. Yang, K. Duan, G. Li, W. Zhang, Q. Huang, and Q. Tian, ‘‘The unmanned aerial vehicle benchmark: Object detection and tracking,’’ in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 370–386.

[39] 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.

[40] 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.

[41] 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.

[42] M. Danelljan, A. Robinson, F. S. Khan, and M. Felsberg, ‘‘Beyond correlation filters: Learning continuous convolution operators for visual tracking,’’ in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016.

[43] 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, Qinhuangdao, 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 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.

JINGJING QIAO received the bachelor's degree in electrical engineering and automation from Hubei Nationalities University, Enshi, Hubei, China. She is currently pursuing the master's degree in engineering with the North China University of Science and Technology. Her research interests include computer vision, correlated filters, 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.