information

Article

Research on Anti-Occlusion Correlation Filtering Tracking Algorithm Based on Adaptive Scale

Xifeng Guo, Turdi Tohti *, Mayire Ibrayim and Askar Hamdulla

College of Information Science and Engineering, Xinjiang University, Urumqi 830017, China; xguo@stu.xju.edu.cn (X.G.); mayire401@xju.edu.cn (M.I.); askar@xju.edu.cn (A.H.)
* Correspondence: turdy@xju.edu.cn; Tel.: +86-139-9999-4696

Abstract: Target tracking has always been an important research direction in the field of computer vision. The target tracking method based on correlation filtering has become a research hotspot in the field of target tracking due to its efficiency and robustness. In recent years, a series of new developments have been made in this research. However, traditional correlation filtering algorithms cannot achieve real-time tracking in complex scenes such as illumination changes, target occlusion, motion deformation, and motion blur due to their single characteristics and insufficient background information. Therefore, a scale-adaptive anti-occlusion correlation filtering tracking algorithm is proposed. First, solve the single feature problem of traditional correlation filters through feature fusion. Secondly, the scale pyramid is introduced to solve the problem of tracking failure caused by scale changes. In this paper, two independent filters are trained, namely the position filter and the scale filter, to locate and scale the target, respectively. Finally, an occlusion judgment strategy is proposed to improve the robustness of the algorithm in view of the tracking drift problem caused by the occlusion of the target. In addition, the problem of insufficient background information in traditional correlation filtering algorithms is improved by adding context-aware background information. The experimental results show that the improved algorithm has a significant improvement in success rate and accuracy compared when with the traditional kernel correlation filter tracking algorithm. When the target has large scale changes or there is occlusion, the improved algorithm can still keep stable tracking.

Keywords: target tracking; correlation filtering; feature fusion; scale adaptation; occlusion judgment; context-aware

Citation: Guo, X.; Tohti, T.; Ibrayim, M.; Hamdulla, A. Research on Anti-Occlusion Correlation Filtering Tracking Algorithm Based on Adaptive Scale. *Information* **2022**, *13*, 131. https://doi.org/10.3390/info13030131

Academic Editor: Gholamreza Anbarjafari (Shahab)

Received: 11 January 2022
Accepted: 2 March 2022
Published: 4 March 2022

1. Introduction

With the continuous development and progress of modern technology, the discipline of artificial intelligence has entered a new era, and the use of artificial intelligence technology to replace traditional manual labor has become a development trend. Computer vision, as an important branch of artificial intelligence, contains many kinds of technology research for different application scenarios, and target tracking algorithm is one of the research directions with practical significance. Target tracking usually means that the initial position of the target is given in the first frame, and the position and shape of the tracked target are estimated in the subsequent video sequences to obtain information about the moving direction and trajectory of the target [1]. Its essence is to use feature extraction and feature association techniques to match the features in different frames of images that are most likely to belong to the same target according to the given target image, and then connect the matched targets in the video frames to get the target's motion trajectory and finally achieve target tracking. The research and application of target tracking methods, as an important branch of computer vision, is widely used in various fields such as autonomous driving, intelligent video surveillance, medical care, and military [2].

Although the target tracking algorithm based on correlation filtering has made great progress, there are still some problems to be solved:

(1) During the movement of the target (a variety of factors can affect the performance and efficiency of the tracking algorithm) the algorithm must not only solve the target itself by the light, deformation, and rapid movement brought about by the interference, but also to solve the target by obstacles or even by other targets partially obscured or completely obscured. Furthermore, in some image sequences the target moved out of the field of view, in the actual application scenario there may also be complex climate conditions change, excessive external noise and other factors of interference. These are the main technical challenges of the current tracking algorithms to ensure tracking accuracy while taking into account the real-time tracking performance;
(2) Model update strategy and update interval selection problem: if the update is too frequent, it will lead to a large amount of model computation and thus affect the real-time problem, and may lead to the loss of some feature information; if the update is too slow, there may be a drift of the tracking frame caused by feature changes occurring too quickly;
(3) Search box size selection problem: If the search box is too small, it is not easy to detect fast-moving targets; if the search box is too large, it will introduce a lot of useless background information, and even some backgrounds similar to the target will interfere with the target tracking process, leading to the degradation of the model and the phenomenon of tracking drift.

In order to better deal with these problems, this paper proposes a new algorithm. Our main contributions can be summarized as follows:

(1) For the defects of single features, a serial feature fusion strategy is used to fuse multiple features for feature extraction;
(2) A scale adaptation strategy is proposed for the target frame size fixation problem to cope with the large deformation of the target during the tracking process;
(3) To address the problem of contamination of the extracted target appearance features due to occlusion, an anti-occlusion detection model update strategy is proposed;
(4) For some complex application scenarios, the discriminative power is enhanced by adding contextual information blocks.

2. Related Work

In 2010, Bolme et al. [3] proposed the Minimized Output Sum of Squared Error (MOSSE) model in their CVPR work, which first used the correlation filter class algorithm for target tracking and achieved a breakthrough in tracking speed, and subsequently, more and more tracking algorithms have been improved on this basis. CSK [4] introduced circular matrix and kernel function to improve the operation rate, and CN [5] color features extended CSK as a multi-channel color tracker and combined with adaptive dimensionality reduction strategy to improve the performance of the algorithm while reducing the computational cost. Henriques et al. proposed a new kernelized correlation filter KCF (Kernelized Correlation Filter) [6] based on CSK, which adds a circular matrix to alleviate the effect of the few learning samples, extending single-channel features to multi-channel directional gradient histogram features, using the gradient information of the image to improve the performance of the algorithm. Besides, learning the target detector by ridge regression results in a faster tracking speed. However, the algorithm fixes the target size, tracking is not robust when scale changes occur, and the circular matrix is prone to boundary effects. In 2015, Denelljan's team designed a spatially regularized discriminative correlation filter SRDCF [7] for tracking, which not only suppresses the background response and expands the search, but also solves the boundary effect problem of the KCF algorithm to obtain better performance in complex background scenes. However, the algorithm only has a speed of only four FPS and cannot perform real-time tracking. Since then, after years of research by domestic and foreign researchers, a large number of classical and excellent algorithms have emerged, and there are many practical applications in engineering projects.

However, due to many interference factors such as various motions of objects in reality and changes in the background, the current tracking algorithms are far from being able

to meet the needs of practical applications in terms of accuracy, robustness and real-time performance, and still face a series of challenges. Wu et al. summarized these difficulties caused in the OTB100 dataset as deformation, illumination changes, fast movement, background interference, rotation, scale changes, and occlusion [8]. For existing target tracking algorithms, few can deal with both background changes and target changes at the same time, so how to establish a stable target appearance model to cope with these difficulties is still the focus and difficulty of future research. Naiyan Wang et al. [9] divided a tracking system into five parts: motion model, feature extraction, observation model, model update, and integration method.

Among them, the motion model mainly describes the motion trend and motion state information of the target to be tracked in the continuous image sequence, and the model generates some candidate regions or bounding boxes that may contain the target object in the current frame. The model is commonly used in algorithms such as particle filtering [10] and Kalman filtering [11]; the apparent model is an algorithm used to extract the features of the target image to be tracked by calculating the similarity between image features to mark out the same target in different images, and then connect these matching targets to obtain the motion trajectory of the target, and finally achieve the purpose of tracking; the role of the observation model is to predict the state of the target based on the candidate states provided by the apparent and motion models, and most tracking methods focus on the design of this part; the model update is mainly for the update of appearance model and motion model to adapt to the change of the target appearance and prevent drift in the tracking process; the integrated approach is beneficial to improve the prediction accuracy of the model and is often seen as an effective means to improve tracking accuracy. Inspired by this, we improve the original KCF algorithm in terms of motion model, apparent model, and model update mainly for single feature, scale change, sample problem, and target occlusion problem to improve the accuracy and robustness of the associated filtered target tracking algorithm.

3. Analysis of KCF Algorithm

3.1. Ridge Regression

The KCF algorithm extends single-channel features to multi-channel HOG features, using the gradient information of the image to improve the performance of the algorithm. It employs ridge regression and circular matrix in the matrix operation process, which effectively reduces the computational effort.

In order to solve the problem of insufficient training samples, the KCF algorithm constructs the training sample set (x_i, y_i) of the filter from around the target by a cyclic shift method. The sample training process is actually a ridge regression problem, and the minimum generation mean square error of the trainer can be constructed by the least-squares method as follows:

$$\min_{w} \sum_{i=1}^{n} (f(x_i) - y_i)^2 + \lambda \|w\|^2 \tag{1}$$

in which $f(x_i) = w^T x_i$, where w denotes the weights of the coefficients and λ is the regularization parameter to prevent overfitting. Simplification in the frequency domain using the Fourier transform leads to $w = \sum_i \alpha_i \varphi(x_i)$, where $\varphi(x_i)$ denotes the mapping function.

The closed-form solution of the ridge regression can be obtained as:

$$\alpha = (k + \lambda I)^{-1} y$$

where α denotes the solution in the dyadic space, k is the kernel matrix, and I is the unit matrix.

3.2. Model Updates

To meet the changes in the target tracking process, the appearance model x and the classifier coefficients α need to be updated in real-time, and the model update formula is as follows:

$$\begin{aligned} x_i &= (1-\eta)x_{i-1} + \eta x_i \\ \alpha_i &= (1-\eta)\alpha_{i-1} + \eta \alpha_i \end{aligned} \tag{2}$$

where x_i and x_{i-1} are the target feature models of the current frame and the previous frame, respectively, α and α_{i-1} denote the coefficient matrices of the current frame and the previous frame and η is the learning coefficient.

4. Improved KCF Algorithm

4.1. Overall Flow of the Algorithm

The overall flow of the algorithm in this paper is as follows: (1) Input image containing the target image and its surrounding contextual background information; (2) Extracting manual features HOG, CN, Gray, fusing them and obtaining the fused features; (3) Bringing the fused features into the KCF algorithm to calculate the location and multi-scale response, where the optimal scale of the target is obtained using the scale pyramid approach; (4) To avoid the tracking drift problem caused by the update of wrong frames during the tracking process, we adopt an anti-occlusion detection strategy to update the corresponding model only for the images that meet the requirements. (5) An efficient correlation operation is performed by using the Discrete Fourier Transform (DFT). The response graph is obtained through the inverse Fourier transform (IFFT). The position of the maximum response score is the new position and the best scale of the target in the current frame, then the relevant filter is updated accordingly. The specific flow chart is shown in Figure 1.

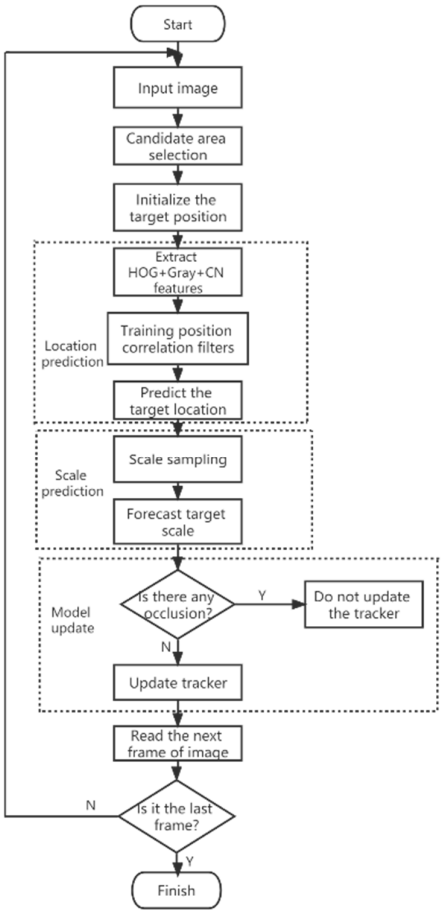


Figure 1. The flow chart of the proposed algorithm.

4.2. Feature Extraction Based on Multi-Feature Fusion

Currently, the main manual features used in tracking algorithms are color features, grayscale features, directional gradient histogram features, texture features and shape features, etc. These features quantify and describe some information of the image from different perspectives, and the expressiveness of the resulting feature operators varies.

KCF extracts the HOG features by calculating the target direction gradient histogram, which mainly describes the edge information of the target [12]. More than this, it can better cope with scenarios such as background color interference, background chaos and complexity. However, the expressiveness of this feature decreases when the target undergoes complex conditions such as deformation, image blurring, and rotation. The CN feature has a better tracking effect for motion blur, light intensity change, background confusion, etc., however, shows poor performance for similar color interference. Gray feature has low computational complexity and fast computing speed. However, it is difficult to achieve accurate tracking in most cases. Feature extraction is one of the key links in the process of target tracking. If there is only a single feature, when the scene changes in the process of tracking, the single feature may not be able to cope with the complex scene changes and become invalid or unreliable. And multi-feature fusion can effectively deal with this situation. The fused features improve the performance of the tracking algorithm by using the complementarity between multiple features. Even when a feature is unreliable, the complementary features are still effective. Through this way, the target location and tracking can be completed in most cases. The feature fusion methods mainly include serial fusion and parallel fusion. Among them, the serial fusion strategy can effectively identify the profile information of the target and eliminate most of the redundant information, so that the fusion observation function has a higher confidence, which would improve tracking accuracy and real-time performance. Therefore, we use the serial feature fusion strategy to fuse HOG + CN + Gray features for feature extraction. When serial feature fusion is performed, let the two features, S(A dimension) and T(B dimension), be defined on the same space where A and B are the dimensions of the two features, respectively, then the dimensionality of the features after using serial feature fusion is A + B dimension [13]. To address the shortcomings of single features, the specific steps of our proposed serial feature fusion are as follows:

Extract manual features, where the channel numbers of Gray feature, CN feature, and HOG feature are 1, 10, and 31, respectively:

(1) Serial fusion of these features is performed to obtain a total number of 42 feature channels;
(2) The feature maps of these 42 channels are pixel summed to obtain a single channel manual feature map;
(3) The obtained manual features are adjusted to the same size to get the final fused features.

The experimental results show that the use of feature serial fusion strategy can realize the complementary advantages between different features, effectively solve the shortcomings of single feature discrimination performance, and achieve better performance under the condition that feature extraction is not time-consuming.

4.3. Scale Adaptive Evaluation

Robust scale estimation has been a challenging problem in the field of visual target tracking. In the traditional correlation filter tracking algorithm, a fixed size sampled selection frame is used, the algorithm always uses the same size selection frame for image feature extraction during the target tracking of the video sequence. Fixed checkboxes are difficult to adapt to changes in the scale of the target, and if the target area becomes larger, the tracking box cannot fully contain the target, resulting in the absence of some information about the target. When the target area becomes larger, the tracking frame contains not only the information of the target but also the background information of the target, which can lead to the inclusion of too much interference information in the

tracking frame [14] and affect the robustness of the tracking results. This paper proposes an improved scheme for this problem of scale variation, that is, the scale estimation is carried out by introducing the scale pyramid [15], and the scale and position of the target are calculated independently. The center position of the target is first determined, and then 33 kinds of scales are performed on its basis. The specific steps of scale adaptation are as follows:

In order to adapt to the scale change of the target, it is necessary to select a series of samples to construct the scale pyramid, and the scale of the samples is selected as:

$$a^n P \times a^n R, n \in \left\{ -\frac{S-1}{2}, \ldots, \frac{S-1}{2} \right\} \tag{3}$$

where $a = 1.02$ is the scale factor and $S = 33$ is the number of images of different scales. P and R are the width and height of the target window in the previous frame, respectively. To avoid the image sampling being too large to affect the calculation speed, these 33 images of different scale sizes are adjusted to a uniform size, and then their features are calculated separately. The minimum loss function is constructed and the scaled filter model is solved as follows:

$$\varepsilon = \| \sum_{l=1}^{d} h^l \times f^l - g \|^2 + \lambda \sum_{l=1}^{d} \| h^l \|^2 \tag{4}$$

where g is the desired correlation output and the parameter $\lambda \geq 0$ controls the effect of the regularization term.

Solving the above equation yields the filter as:

$$H^l = \frac{\overline{G} F^l}{\sum_{k=1}^{d} \overline{F^k} F^k + \lambda} \tag{5}$$

According to the filter obtained from training, the FHOG features of different scale images are extracted in the next frame with the target point as the center, and the extracted FHOG features are concatenated into a two-dimensional matrix to obtain the scale feature vector Z. The filter is then used to discriminate and obtain the scale response y_s:

$$y_s = F^{-1} \left\{ \frac{\sum_{l=1}^{d} \overline{A^l} Z^l}{B + \lambda} \right\} \tag{6}$$

The scale with the largest response value is selected as the best target scale. After getting the best scale, the model needs to be updated to accommodate the change in scale. Updating the numerator A_t and denominator B_t of the associated filter H separately:

$$\begin{aligned} A_t &= (1-\eta) A_{t-1} + \eta \overline{G_t} F_t \\ B_t &= (1-\eta) B_{t-1} + \eta \sum_{k=1}^{d} \overline{F_t^k} F_t^k \end{aligned} \tag{7}$$

Among them, η is a learning rate parameter.

4.4. Model Update Strategy

In the target tracking algorithm, in addition to feature extraction and scale estimation [16–18], the model update mechanism is also a very important part. In the process of the model update, judging the occlusion of the target has always been a big problem in the field of visual tracking. During the tracking process, when occlusion occurs, the extracted target appearance features will be contaminated and the target tracking model will have different degrees of tracking drift, such as drifting to other similar objects in the background, which will lead to the failure of tracking results. KCF is updated frame-by-frame without the determination of the reliability of tracking results. Each tracking target corresponds to a Gaussian label response, and the response map oscillates gently when the target is

not occluded, and, conversely, oscillates violently with multiple peaks when the target is occluded.

We introduce the *APCE* occlusion discrimination strategy proposed by Wang et al. in the literature [19] to address the problem of easy tracking failure when the target is occluded.

$$APCE = \frac{|F_{\max} - F_{\min}|^2}{mean\left(\sum_{w,h}(F_{w,h} - F_{\min})^2\right)} \tag{8}$$

Where $F_{\max}$ represents the peak value of the response map, $F_{\min}$ is the minimum value of the response map, and $F_{w,h}$ denotes the response value at the (w,h) position. The *APCE* criterion can reflect the degree of oscillation of the response map. When the target is not obscured, the *APCE* value will be relatively large, and its peak will be a single peak with a more prominent peak. When the target is occluded or lost, the response peak will be reduced, or even multiple peak points will appear, which will make the *APCE* value decrease. In the proposed filter update, it is considered that the model is updated only when both the *APCE* and the response peak $F_{\max}$ are greater than the historical mean by a certain ratio, and the filter is not updated whenever either of the two conditions is not satisfied. This reduces the number of model updates necessary to achieve acceleration and also reduces the problem of filter corruption due to the introduction of too much background information.

4.5. Context-Aware Correlation Filters

The traditional correlation filtering target tracking algorithm is limited by the cosine window and search area so that the correlation filtering template does not learn much background information, and the tracking effect will be limited when the target has deformation, occlusion, fast motion, etc., resulting in tracking loss. For example, the KCF algorithm considers that the samples obtained after circular shifting are approximately equal to the real samples, which not only results in little background information being learned, but also causes the problem of sample redundancy. To address this problem, this article refers to the framework proposed by the CACF algorithm [20] to add more contextual background information and include this contextual information into the learned filters. Figure 2 compares the traditional correlation filtering method with our method of adding contextual background information.

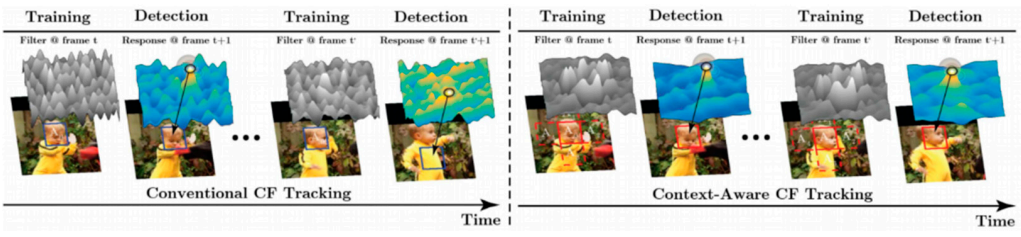


Figure 2. Comparing conventional CF tracking to proposed context-aware tracking.

As can be seen from Figure 2, in the training phase, after adding a context to the top, bottom, left, and right of the target, the response map has less interference and better discriminative ability. The tracking is still very effective after the target has been significantly rotated.

$$\min_w \|A_0 w - y\|_2^2 + \lambda_1 \|w\|_2^2 \tag{9}$$

where A_0 denotes the matrix obtained by cyclic shifting for a given image block. Adding the above equation to the background penalty term yields:

$$\min_w \|A_0 w - y\|_2^2 + \lambda_1 \|w\|_2^2 + \lambda_2 \sum_{i=1}^{k} \|A_i w\|_2^2 \tag{10}$$

where A_i denotes the circular matrix composed of context blocks A_i, with the middle A_0 marked as positive samples and the nearby context regions A_i marked as negative samples. A penalty term is added to make the response as small as possible when the template w to be trained does correlation with A_i.

5. Experimental Results and Analysis

5.1. Experimental Environment and Configuration

The software platforms we used were MATLAB_2019b and vs2019; the hardware configuration was: Intel(R) Core(TM) i7-10750H CPU @ 2.60 GHz 2.59 GHz processor with 8 G RAM. The graphics card is NVIDIA GeForce GTX 1660 Ti; the operating system environment is Ubuntu 16.04 and windows 10 dual system. The experimental parameters are set as follows: The regularization factor lambda1 = 0.0001 lambda2 = 0.5, the learning rate of the filter is 0.015, the learning rate of the scale filter is 0.025, the padding size is 1.7,$\gamma_1 = 0.4$, $\gamma_2 = 0.6$, the scale pool size is 33, and the scale benchmark factor is 1.02.

5.2. Evaluation Indicators

To make a uniform comparison of the accuracy of all tracking algorithms, we use the publicly available dataset OTB-2015, which contains 100 video sequences with 11 possible complications during target tracking: scale variation (SV), shift out of field of view (OV), low resolution (LR), background clutter (BC), occlusion (OCC), in-plane rotation (IPR), out-of-plane rotation (OPR), deformation (DEF), fast movement (FM), motion blur (MB), and illumination variation (IV). One-Pass Evaluation (OPE) was performed using Overlap Precision (OP) and Distance Precision (DP) metrics. Overlap accuracy (OP) is the ratio of the number of video frames whose overlap between the indicator note frame and the tracking prediction frame is greater than a certain threshold to the total number of video frames currently being tracked. The center position error *CLE* is the Euclidean distance between the center of the true labeled frame (x_b, y_b) and the center of the predicted frame (x_c, y_c):

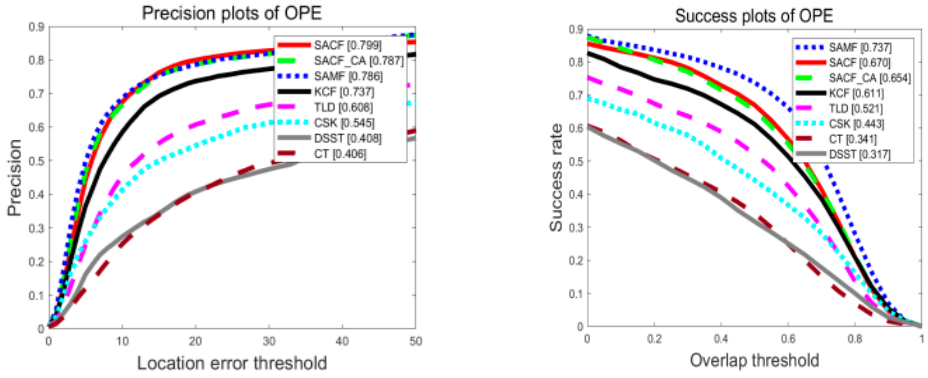
$$CLE = \sqrt{(x_b, y_b)^2 + (x_c, y_c)^2} \tag{11}$$

Distance accuracy DP is the ratio of the number of video frames whose center position error *CLE* is less than a set threshold to the total number of video frames. We have an overlap success rate threshold of 0.5 and a distance accuracy threshold of 20 pixels.

5.3. Quantitative Analysis

In this section, our improved scale-adaptive anti-occlusion correlation filter algorithm SACF (containing three improvements of feature fusion, scale adaption, and model update) and scale-adaptive anti-occlusion correlation filter for context-aware SACF_CA (containing four improvements of feature fusion, scale adaption, model update, and context-aware) are compared with several current classical correlation filter target tracking algorithms, including several algorithms of KCF, DSST, CSK, CT, SAMF, TLD, and the comparison chart is as Figure 3:

Under the aforementioned hardware and software experimental conditions, the average FPS of our proposed improved algorithm SACF is 70.138. As can be seen from Figure 3, the algorithm in this paper ranks first among all algorithms in tracking accuracy and second among all algorithms in tracking success rate, second only to SAMF algorithm. The tracking accuracy and success rate are improved by 6.2% and 5.9%, respectively, compared with the original KCF algorithm, the tracking accuracy reaches 79.9% and the success rate reaches 67%. The proposed improved algorithm SACF_CA has a tracking accuracy of 78.7%, which ranks second among all algorithms, a success rate of 65.4%, which ranks third among all algorithms, which improves the tracking accuracy and success rate by 5% and 4.3%, respectively, compared to the original algorithm.



(**a**) (**b**)

Figure 3. Distance accuracy and success rate curves of eight algorithms in OTB-2015. (**a**) Location error threshold. (**b**) Overlap threshold.

We compare the tracking performance on problems such as in-plane rotation, out-of-plane rotation, occlusion, fast motion, background interference, deformation, illumination change, and scale change frequently encountered in the target tracking process, and the corresponding accuracy and success rate comparison curves are shown below.

It can be seen from Figure 4, both of our proposed algorithms, SACF and SACF_CA, have high tracking accuracy and tracking robustness in some specific real-world complex scenarios. Among them, SACF_CA is very robust in complex video sequences such as deformation, lighting changes, and motion blur.

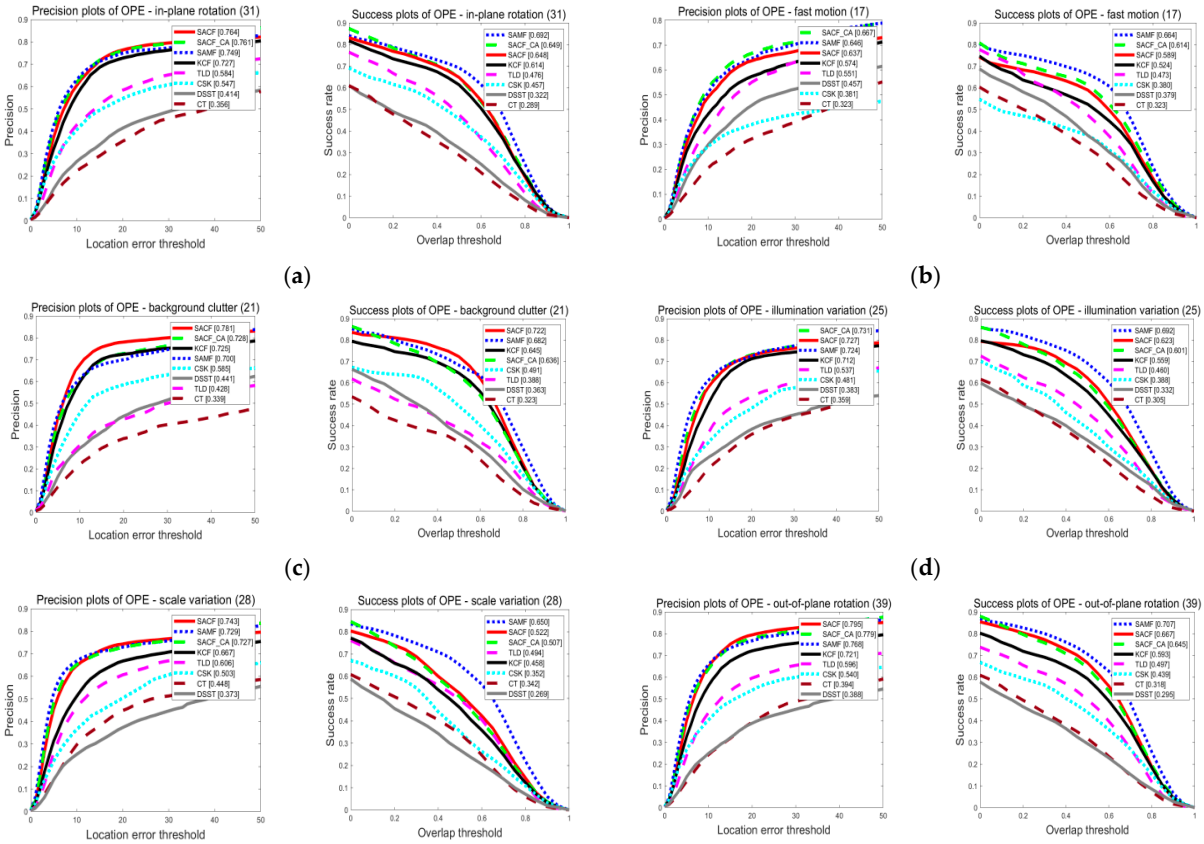


Figure 4. Comparison of tracking accuracy and success rate of various algorithms on (**a**–**f**) video sequences. (**a**) In-plane rotation sequence. (**b**) Fast motion sequence. (**c**) Background clutter sequence. (**d**) Lighting change sequence. (**e**) Scale change sequence. (**f**) Out-of-plane rotation sequence.

The performance comparison of six algorithms and the DP values under different scene attributes are given in Tables 1 and 2, respectively, and the values in bold are the optimal values.

Table 1. Performance comparison of six algorithms.

Algorithm	CSK	KCF	DSST	SACF(ours)	SACF_CA(ours)	SAMF
Mean DP	0.545	0.737	0.408	**0.799**	0.787	0.786
Mean OP	0.433	0.611	0.317	0.670	0.654	**0.737**
Mean FPS	**297.765**	169.72	63.854	70.138	43.062	15.346

Table 2. DP values of the six algorithms for different scene attributes.

	CSK	KCF	DSST	SACF(ours)	SACF_CA(ours)	SAMF
IV	0.481	0.712	0.383	0.727	**0.731**	0.724
MB	0.342	0.611	0.435	0.560	**0.648**	0.621
SV	0.503	0.667	0.373	**0.743**	0.727	0.729
OV	0.379	0.555	0.293	0.726	**0.751**	0.729
LR	0.411	0.379	0.360	0.475	**0.616**	0.520
BC	0.585	0.725	0.441	**0.781**	0.728	0.700
OCC	0.500	0.740	0.377	0.789	**0.826**	0.808
IPR	0.547	0.727	0.414	**0.764**	0.761	0.749
OPR	0.540	0.721	0.388	**0.795**	0.779	0.768
DEF	0.476	0.745	0.378	0.809	**0.839**	0.823
FM	0.381	0.574	0.457	0.637	**0.667**	0.646

Table 1 compares the proposed SACF method and SACF_CA with the most advanced HOG-based tracker on the OTB2015 dataset. It can be seen from the experimental results that the proposed SACF method obtains the highest average distance accuracy value on the OTB2015 dataset, reaching 0.799, and the DP value of the SACF_CA method is second only to SACF, which is 0.787. Among the six trackers, the SAMF method has the highest average overlap rate, followed by our method. In terms of average tracking speed, the CSK method has the fastest tracking speed, reaching 297.765, and KCF's FPS ranks second. The real-time tracking speeds of the two trackers proposed in this paper, SACF and SACF_CA, are 70.138 and 43.062, respectively. The average tracking speed of SAMF is only 15.346.

As can be seen from Table 2, the six top-ranked algorithms are compared in terms of DP values for 11 complex scenarios, among them, this article proposed improved algorithms rank in the top two and outperform the other algorithms overall. It shows that proposed algorithms not only improves the tracker's resistance to occlusion and deformation but also has strong robustness to various other types of interference and can adapt to various complex application scenarios.

5.4. Qualitative Analysis

To demonstrate more intuitively the effectiveness of our algorithm in real scenarios, in our experiments we extracted several sets of representative video sequences for comparison, and compared our proposed algorithm SACF with SACF_CA and four classical algorithms DSST, KCF, SAMF, and TLD, as shown in Figure 5.

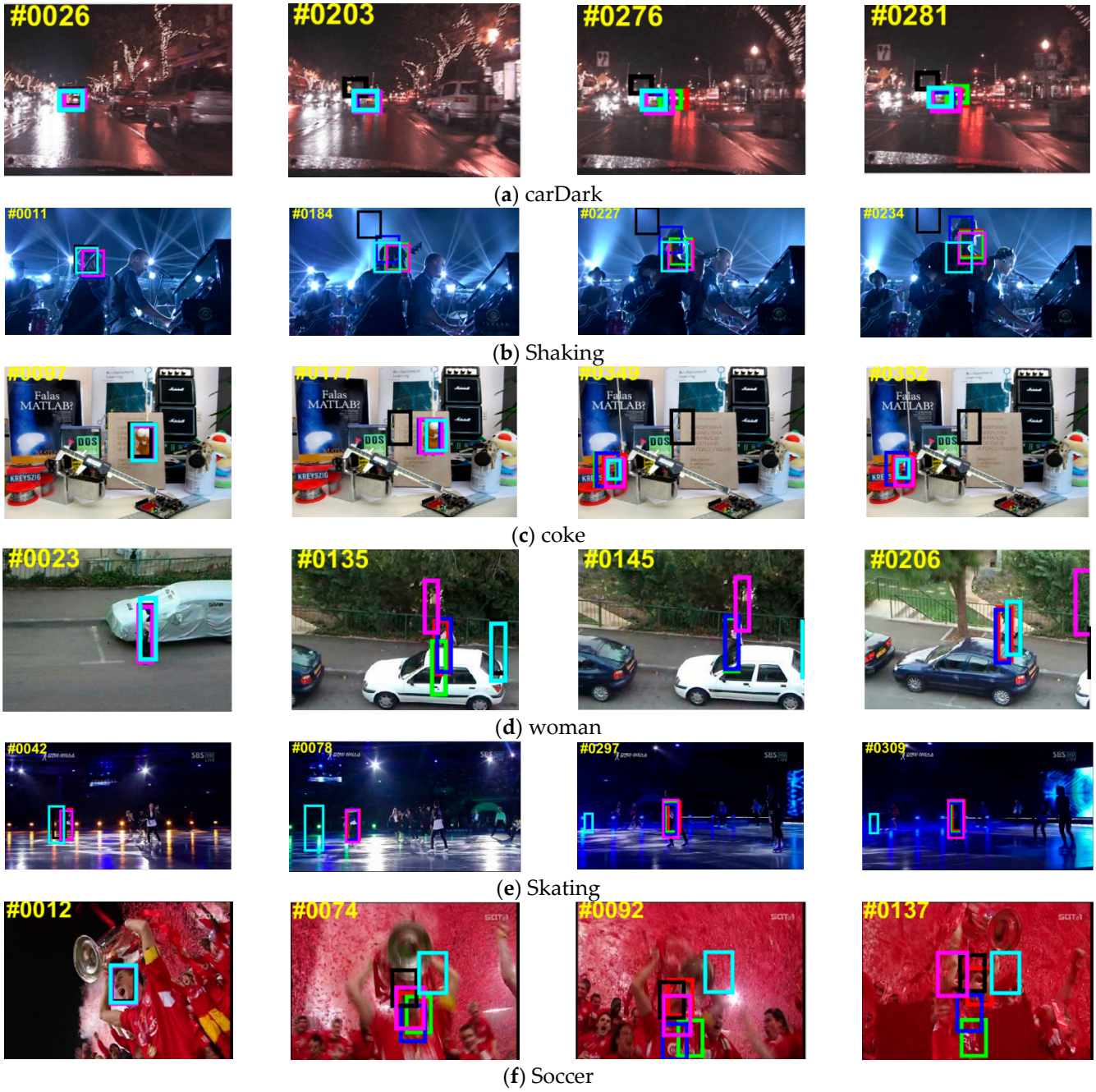


Figure 5. Partial video sequence tracking results. SACF, SACF_CA, SAMF, KCF, CSK, TLD.

Among them, the first two groups of video sequences have the interference of illumination changes and cluttered backgrounds, the third and fourth groups have occlusions, and the fifth group has large scale changes. In the sixth group of video sequences, there are various interferences such as occlusion, background confusion, motion blur, etc. It can be seen from the first two video sequences that the two algorithms SACF and SACF_CA proposed in this paper can resist the interference of the cluttered background and can accurately locate the target when the illumination changes or the target is disturbed by the background chaos, followed by TLD, and KCF loses the target. When the target is occluded, it can be divided into partial occlusion and full occlusion according to the degree of occlusion. In Figure 5, the coke test sequence has long-term total occlusion between frames 177 and 349, while the woman test sequence has short-term occlusion from 135 frame to

145 frame. Both algorithms proposed in this paper can track the target smoothly, even in the woman sequence, SACF_CA algorithm has a short drift, it can still relocate and track the target at the 145th frame. Although there are many complex situations such as background clutter and motion blur in the soccer video sequence, the SACF algorithm can still resist all kinds of interferences, which shows good robustness, and correctly represents the target. The experimental results show that the improved algorithms proposed in this paper can cope well with various complex practical application scenarios.

6. Conclusions

The improved algorithm adaptively updates the tracking frame size by fusing the directional gradient histogram, grayscale, and color features, and adding adaptive scale analysis to the targets that are prone to deformation during tracking, while performing some processing on the input samples. It is worth mentioning that for the occlusion interference factor, the model update mechanism proposed in this paper can cope well with the occlusion of the target. The filter is not updated when it detects that the target is occluded, which reduces the tracking drift caused by the contamination of the filter template by the background information to a certain extent. The validation results on the OTB2015 dataset show that compared with the original algorithm, our algorithm improves 6.2% and 5.9% in tracking accuracy and success rate, respectively, with higher tracking accuracy and robustness, and the improved algorithm also has stronger robustness compared to some classical mainstream algorithms in correlated filtered target tracking.

Author Contributions: Conceptualization, X.G.; methodology, X.G.; software, A.H.; validation, X.G., T.T. and M.I.; formal analysis, X.G.; investigation, T.T.; resources, A.H.; data curation, X.G.; writing—original draft preparation, X.G.; writing—review and editing, T.T. and M.I.; visualization, X.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been supported by the National Natural Science Foundation of China (62166042, 62166043 and U2003207), and Natural Science Foundation of Xinjiang, China (2021D01C076 and 2020D01C045), and Youth Fund for scientific research program of Xinjiang (XJEDU2019Y007).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available dataset was analyzed in this study. Our dataset can be obtained from (http://cvlab.hanyang.ac.kr/tracker_benchmark/datasets.html, accessed on 25 August 2021).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chen, Z.; Hong, Z.; Tao, D. An experimental survey on correlation filter-based tracking. *arXiv* **2015**, arXiv:150905520.
2. Li, J.; Zhou, X.; Chan, S.; Chen, S. A novel video target tracking method based on adaptive convolutional neural network feature. *J. Comput. Aided Comput. Graph.* **2018**, *30*, 273–281. [CrossRef]
3. Bolme, D.S.; Beveridge, J.R.; Draper, B.A.; Lui, Y.M. Visual object tracking using adaptive correlation filters. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 2544–2550. [CrossRef]
4. Henriques, J.F.; Caseiro, R.; Martins, P.; Batista, J. Exploiting the circulant structure of tracking-by-detection with kernels. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 702–715.
5. Danelljan, M.; Khan, F.S.; Felsberg, M.; van de Weijer, J. Adaptive color attributes for real-time visual tracking. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1090–1097.
6. Henriques, J.F.; Caseiro, R.; Martins, P.; Batista, J. High-speed tracking with kernelized correlation filters. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *37*, 583–596. [CrossRef] [PubMed]
7. Danelljan, M.; Hager, G.; Khan, F.S.; Felsberg, M. Learning spatially regularized correlation filters for visual tracking. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 4310–4318.
8. Wu, Y.; Lim, J.; Yang, M.-H. Online Object Tracking: A Benchmark. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 2411–2418.

9. Wang, N.; Shi, J.; Yeung, D.Y.; Jia, J. Understanding and diagnosing visual tracking systems. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 3101–3109.
10. Salmond, D.; Birch, H. A particle filter for track-before-detect. In Proceedings of the 2001 American Control Conference (Cat. No. 01CH37148), Arlington, VA, USA, 25–27 June 2001; Volume 5, pp. 3755–3760. [CrossRef]
11. Aidala, V.J. Kalman filter behavior in bearings-only tracking applications. *IEEE Trans. Aerosp. Electron. Syst.* **1979**, *AES-15*, 29–39. [CrossRef]
12. Jiang, N.; Wu, Y. Unifying spatial and attribute selection for distracter-resilient tracking. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 3502–3509.
13. Jie, C.; Wei, L. Object tracking algorithm based on multi-feature fusion. *J. Lanzhou Univ. Technol. Pap.* **2011**, *37*, 80–84.
14. Yang, Y. Research on Robust Visual Object Tracking Algorithm Based on Corelation Filter. Master's Thesis, School of Computer Science & Technology, Huazhong University of Science and Technology, Wuhan, China, 2016; pp. 1–35.
15. Danelljan, M.; Häger, G.; Khan, F.; Felsberg, M. Accurate scale estimation for robust visual tracking. In Proceedings of the British Machine Vision Conference 2014, Nottingham, UK, 1–5 September 2014. [CrossRef]
16. Xiong, C.; Che, M.; Wang, R.; Yan, L. Adaptive model update via fusing peak-to-sidelobe ratio and mean frame difference for visual tracking. *Acta Photonica Sin.* **2018**, *47*, 910001. [CrossRef]
17. Cheng, Y.H.; Wang, J. A motion image detection method based on the inter-frame difference method. *Appl. Mech. Mater.* **2014**, *490*, 1283–1286. [CrossRef]
18. Yang, W.; Shen, Z.-K.; Li, Z.-Y. The application of difference method to dim point target detection in infrared images. In Proceedings of the National Aerospace and Electronics Conference (NAECON'94), Dayton, OH, USA, 23–27 May 1994; pp. 133–136. [CrossRef]
19. Wang, M.; Liu, Y.; Huang, Z. Large margin object tracking with circulant feature maps. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4021–4029.
20. Matthias, M.; Neil, S.; Bernard, G. Context-aware correlation filter tracking. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1387–1395.