Research Article

Adaptive Spatial Regularization Target Tracking Algorithm Based on Multifeature Fusion

Turdi Tohti,1,2 Xifeng Guo,1,2 and Askar Hamdulla1,2

1School of Information Science and Engineering, Xinjiang University, Urumqi 830017, China
2Xinjiang Key Laboratory of Signal Detection and Processing, Urumqi 830017, China

Correspondence should be addressed to Askar Hamdulla; askar@xju.edu.cn

Received 15 August 2022; Revised 3 December 2022; Accepted 8 December 2022; Published 29 December 2022

Copyright © 2022 Turdi Tohti et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The accuracy and robustness of object-tracking algorithms are challenging tasks in the field of artificial intelligence. The discriminative correlation filter has fast tracking speed and target discrimination ability in visual target tracking, but it will be affected by unnecessary boundary effects. Although spatially regularized discriminative correlation filters (SRDCF) effectively solve this problem, it faces a slow tracking speed and cannot meet the real-time requirement. Moreover, the constraints added by this algorithm are fixed during the tracking process, which needs to reflect better the characteristics and appearance changes of a particular object. Therefore, we propose an adaptive spatial regularization target tracking model based on multifeature fusion and learn the spatial regularization weights by adaptive spatial regularization terms, so that the filter can adapt to the changes of the target, which can highlight the target area and suppress the background area more accurately. At the same time, in order to cope with the problem of target loss when the target is obscured by a large area or rotated to a large extent, texture features are introduced to improve the deficiencies of HOG features and CN features, so as to achieve the complementary advantages among different features. In addition, the benchmark algorithm SRDCF uses Gaussian–Seidel iterative method to solve the filter, which makes the tracking speed very slow and cannot be tracked in real time. Therefore, this paper proposes an ADMM (alternating direction method of multiplier) algorithm to optimize the solution of the filter and achieve the tracking real-time requirement. The experimental results on the OTB-2015 dataset show that the accuracy and success rate of the improved algorithm reach 86% and 81.9%, respectively. At the same time, in the OTB-2013 dataset, it has increased by 3.4% and 6.3%, respectively, on the basis of the benchmark algorithm. It is verified that the improved model not only improves the robustness of the algorithm but also can better adapt to the changes of the target, while effectively reducing the computational overhead and meeting the tracking real-time requirements.

1. Introduction

Now, artificial intelligence has entered an era of explosion, which is driving a new round of global intelligence innovation boom. Computer vision technology [1] is an important research direction of artificial intelligence, which can be used to simulate the biological vision system. It has related technical tasks such as image classification, target detection, target tracking, semantic segmentation, and instance segmentation. Among them, target tracking technology [2] is research that needs to be deeply explored and improved in the field of computer vision. Because of this, it has attracted a large number of experts and scholars to research and discusses, making many classical target-tracking algorithms come into being [3]. Target tracking first obtains the position of the target in the first frame according to manual annotation or image detection, then estimates the location and scale of the object in the next frame, and finally obtains the target’s motion trajectory and contour information according to the relevant information obtained. In recent years, due to the widespread use of cameras, there has been a boom in research on target tracking algorithms in order to analyze and process the captured images, with extensive attention from scholars both at home and abroad. Technology is inseparable from human life in the development of artificial intelligence, and with the continuous
maturation of the technology, it is now often used in various fields such as autonomous driving, defense and security, biomedicine, human-machine interaction, and video surveillance [4–6]. In addition, this technology is also widely used in the field of UAV tracking [7].

Usually, the trackers based on correlation filtering use a large number of cyclic shift samples for learning, which achieves good tracking performance. However, it also makes the correlation filter model train some unreal samples, leading to the boundary effects problem [9]. At the same time, due to many interfering factors in the actual scene, such as various motions of objects, changes in the background, etc. [9], the current tracking algorithms cannot meet the needs of practical applications in terms of accuracy, robustness, and real-time performance and still face a series of challenges.

This paper studies how to effectively solve the boundary effect problem in a complex tracking environment, adding an adaptive spatial regularization term based on the SRDCF algorithm, and proposes an adaptive spatial regularization correlation filter target tracking algorithm to learn the spatial regularization weight, so that the filter can be well adapted to the target. Furthermore, the multifeature fusion strategy is used to improve the tracking performance of the algorithm in different scenarios. The main motivations and contributions of this paper are as follows:

(1) In case of small partial occlusion of the target or under relatively stable lighting conditions, the fusion of HOG features and CN features can achieve better tracking. However, when the target is obscured by a large area or under a significant degree of rotation, the target tracking will be lost. As a typical texture feature, the local binary patterns (LBP) can effectively cope with some complex conditions such as lighting changes and rotation, with less computation and higher time efficiency, but it is easily disturbed by noise and more sensitive to information in the direction. Therefore, this paper combines the three features of HOG + CN + LBP as the input feature of the algorithm to achieve complementary advantages among different features [8]. And this complementary way can also complete the localization tracking of the target in most cases.

(2) Since the SRDCF algorithm uses the Gaussian–Seidel iterative method in solving the filter, this method requires high conditions for convergence, which makes the tracking speed slow and prevents real-time tracking. The ADMM method uses the divide-and-conquer idea to effectively reduce the cost while reducing the computational effort. Therefore, this paper adopts ADMM to replace the Gauss–Seidel iterative algorithm to optimize the model, which could cut down the computing costs and promote the real-time performance of tracking.

(3) Although the standard spatial term solves the boundary effect to some extent, this constraint is fixed during the tracking process and does not reflect well the variation in the characteristics and appearance of a particular object. The adaptive spatial regularization term is introduced to mitigate the boundary effect problem while allowing the spatial regularization weights to be adaptively adjusted with the change of the target profile. The stability of the tracking is enhanced.

2. Related Work

2.1. Discriminative Correlation Filters. The target tracking algorithm can be divided into two types according to the model category: the generative target tracking algorithm and the discriminant. Among them, the generative target tracking methods [10, 11] mainly describe the appearance features with strong feature description capability. In the search area of the target, the position of the target is located according to the appearance information of the target, and the running path of the target is determined. However, due to the negligence of image background information, when encountering some complex environmental interference such as partial or complete occlusion, internal or external rotation, etc., timely and proper processing is not possible, which may cause the tracking frame to drift to similar objects in the background, thus leading to tracking failure [12]. Discriminative target tracking methods are highly preferred, and most classical target tracking methods also belong to discriminative target tracking methods. Usually, the method regards the target and background as a binary problem; the tracked target is considered a positive sample, and the background information around the tracked target is considered a negative sample. This binary problem is solved by training a classifier to discriminate the target from the background and locate the target’s position [13]. Compared with the generative methods, the discriminative tracking methods based on correlation filtering have been favored by researchers for their speed and good tracking effect. They thus have become the mainstream methods in the field of target tracking. Currently, most of the existing trackers estimate the target state based on classifiers and multiscale estimation. Although the trackers have become more stable, the tracking accuracy is stagnant. Literature [14] solved this problem by proposing a new tracking method based on distance IoU (DIoU) loss, making the proposed tracker include target estimation and target classification. The proposed method achieves competitive tracking accuracy and real-time tracking speed. In the literature [15], an image super-resolution reconstruction method using a feature map attention mechanism is proposed to facilitate the reconstruction of original low-resolution images to multiscale super-resolution images, which effectively improves the visual effect of images. The following is a brief description of the development history and research results of target tracking.

The correlation filtering algorithm was first used in the field of target tracking in the minimum output sum of squared error (MOSSE) model proposed by Bolme [16] in 2010. Subsequently, the correlation filtering method has
been further developed. The circulant structure kernel (CSK) [17] improves the execution efficiency of the algorithm to a great extent by introducing the circulant matrix and kernel function. Moreover, the color feature CN [18] extends the CSK of the single-channel grayscale feature to a multi-channel color tracker, enhancing the algorithm’s robustness. Then, Henriques et al. proposed a new kernel correlation filter (KCF) [19], which utilizes the idea of circular shift and ridge regression. However, it is also prone to boundary effects. To solve this thorny problem, Denellán’s team designed a spatially regularized discriminative correlation filter SRDCF [20] for tracking at ICCV in 2015, a tracker based on the spatial penalty. Its main contributions are as follows: first, to solve the boundary effect problem and provide effective suppression of the background region of the target, the authors include a spatial regularity term in the process of solving the filter. Second, the algorithm roughly specifies several different scales, which performs well in most cases with small-scale changes but needs to be more flexible to cope with sudden changes in the target scale. The Gaussian–Seidel iterative method is used to solve the filter, which significantly reduces the tracking speed. The background-aware correlation filter tracking algorithm (BACF) [21] was proposed by Galoogahi et al. in 2017, which effectively increases the number of samples, improves the quality of the samples using the cropping operation, and uses the alternating direction method of multiplier ADMM [22] to optimize the filter, which speeds up the tracking speed. However, the algorithm does not introduce spatial regularization; it will quickly lead to tracking failure when background interference is encountered. After that, Li et al. [23] proposed a method of learning spatial-temporal regularized correlation filters for visual tracking (STRCF), which added time regularization based on SRDCF to prevent model corruption. However, because its spatially regularized weights have no learning ability, it cannot effectively suppress the background when significant changes in target appearance are encountered. Visual tracking via adaptive spatially regularized correlation filters (ASRCF) was proposed in the literature [21]. The tracker enhances the anti-interference capability of the filter by incorporating a spatial regularization term, which essentially solves the boundary effect and noise problem. But the algorithm is prone to overfitting when the target encounters large deformations due to the absence of a temporal relationship between the filters, which makes the tracking in subsequent frames poor.

With the rapid development of deep learning, much research on related filtering methods incorporating in-depth features has emerged due to the difficulty of adapting manual features to various variations of targets and in-depth features due to their powerful feature processing and representation learning performance. Zhang et al. [24] proposed a target tracking framework combining correlation filtering tracking and Siamese-based tracking, which combines depth features with manual features to evaluate the robustness of tracking results. Many manually labeled training samples are usually required in the training of feature extraction networks, which undoubtedly increases the training cost. For this reason, Yuan et al. [25] proposed a self-supervised learning method based on multicycle consistency loss to pretrain deep feature extraction networks, which can improve the robustness of improving feature extraction networks by using a large number of unlabeled video samples instead of a limited number of manually labeled samples. The Siam-Corners [26] network introduces a layered feature fusion strategy that can enable the corner pooling module to predict multiple corner points of a tracked target in a deep network. It ensures that it achieves experimental results comparable to state-of-the-art trackers while maintaining a high operating speed. Learning dual-level deep representation for thermal infrared tracking [27] model proposes a two-layer feature model containing TIR-specific discriminative features and fine-grained correlation features for robust TIR target tracking to distinguish interferers more effectively. In order to effectively address the elderly activity recognition problem, the literature [28] focuses on fusing multimodal features to effectively aggregate action and interactive discrimination information from RGB video and skeleton sequences. A new extended squeeze-excitation fusion network (ESE-FN) is proposed, and experimental results show that the model achieves the best accuracy.

Therefore, to improve the algorithm’s performance, more and more scholars are currently fusing manual features with depth features to enhance the discriminative properties of the algorithm in subsequent studies. Furthermore, a nonregular arbitrary quadrilateral tracking frame is used to replace the traditional rectangular tracking frame.

2.2. Standard DCF Training and Detection. The correlation filter tracking algorithm is mainly seen in two phases: training and detection. In the training phase, the multi-channel convolutional filter template $f(M \times N \times L)$ is mainly learned from a series of samples, where $M$ denotes the height of the filter, $N$ denotes the filter width, $M \times N$ denotes the single-channel filter template size, and $L$ denotes the total number of feature channels. In the detection phase, a response map is obtained by using the previous frame’s filter and the current frame’s features. Then, the target’s position can be determined from the position of the maximum value in the response map. The response graph of the correlation filter can be expressed as follows:

$$S_f(x) = \sum_{i=1}^{L} x_i \ast f_i,$$  \hspace{1cm} (1)

where $S$ is the response map, $x_i$ is the target feature of the $i$th channel, $f_i$ and $\ast$ denote the filter and circular convolution, respectively. The location of the peak of the response map $S$ is the center of the detection target. How to solve the filter $f$ is a critical step in the correlation filter. The traditional correlation filter solves the filter $f$ in the following way:

$$E(f) = \frac{1}{2} \sum_{i=1}^{L} x_i \ast f_i - y^2 + \frac{\lambda}{2} \sum_{i=1}^{L} \| f_i \|^2,$$  \hspace{1cm} (2)

where $y$ denotes the desired output, a two-dimensional Gaussian distribution function centered on the target. $\lambda \geq 0$
is the weight of the regularization term, and the term after the plus sign is the regularization term used to avoid overfitting. Usually, the filter $f$ of the $t$th frame is trained by formula (2). The target position of the trained filter at frame $t + 1$ is detected by (1), and finally, the result obtained from the $t + 1$ frame is used for training to track the target in a cycle continuously. When solving the filter using (2), it is usually converted to the frequency domain by Fourier transform to increase the computational speed. However, in this process, since the training samples are generated by cyclic shift, it will lead to edge effects and insufficient discriminative ability of the classifier. In addition, during the update phase, the model updates the filter every frame, which degrades the filter performance in complex tracking environments, eventually leading to tracking failure and degradation of tracking performance.

2.3. The Main Principles and Ideas of the SRDCF Algorithm.

The SRDCF algorithm solves the boundary effect problem of the traditional correlation filter tracker by introducing the spatial regularization method, which penalizes the correlation filter coefficients according to the spatial location. This algorithm not only achieves the expansion of the target search region but also effectively suppresses the response of the background region. The algorithm provides excellent tracking performance even in more severe disturbances in the tracking environment.

The optimization function of the problem studied by the SRDCF algorithm is as follows:

$$E(f) = \frac{1}{2} \sum_{i=1}^{L} x_i \ast f_t - y \|^2 + \frac{\lambda}{2} \sum_{i=1}^{L} \| \omega \circ f_l \|^2,$$

where $\circ$ denotes the matrix pairwise multiplication operation. The second term is the regularization term, and $\omega$ is the spatial weight coefficient. SRDCF takes the sample $x$ as large as possible to retain more background information about the target and then penalizes the samples farther away from the center of the target by spatial weight coefficients. The closer the center of the sample, the smaller the penalty of the correlation filter, and the closer the edge of the sample, the larger the penalty will be to suppress the boundary effect problem caused by circular shifting. The principle of SRDCF is shown in Figure 1 [20].

The above figure shows that the SRDCF has a negative Gaussian shape, which is almost constant for different targets and is fixed during tracking, making the spatial regularization weights not adaptively updated. It needs to distinguish the real target from the background accurately. During the tracking process, when the appearance of the target undergoes an enormous irregular transformation, the Gaussian-like weights of SRDCF introduce bias at this time. In (3), SRDCF is solved iteratively using the Gaussian–Seidel method. It makes the tracking speed of SRDCF very slow, only about 5 fps, which cannot meet the requirement of real-time tracking.

3. Improved Algorithm Model

To address the problem of nonupdating spatial regularization weights in the SRDCF algorithm, this paper proposes an adaptive spatial regularization model for multifeature fusion to improve the original algorithm framework. This model can update the spatial regularization weights as the target changes to more accurately identify the target region and suppress the background region [21]. Since this algorithm uses Gaussian–Seidel iterative method in solving the filter, it leads to a slow tracking speed and cannot track some fast-moving targets in real time. In this paper, the ADMM is used to replace the Gauss–Seidel method to iteratively solve the filter, which not only improves the tracking speed but also further expands the target search range to 25 times the target area (padding = 5), ensuring that the target can be tracked accurately in real time in practical application scenarios. The algorithm in this chapter chooses to fuse HOG, CN, and LBP features as input features in the feature extraction module to train a robust tracker.

In the tracking process, the target’s position is initialized in the first frame, and the target’s CN features, HOG features, and LBP features are extracted, respectively. Then, the normalized weighted fusion is performed according to the response maximum of each feature, and the fused features are effectively correlated with the filter template. Among them, the position with the most significant response score is the best position of the target. In order to speed up the operation, this paper uses the ADMM method to optimize the filter. The overall framework diagram of the algorithm in this paper is shown in Figure 2.

3.1. Feature Extraction. Feature extraction is one of the most critical steps in the target tracking process. HOG and CN features are the most commonly used features in computer vision target tracking. The histogram of the oriented gradient [16] feature mainly describes the edge information of the target. However, it is difficult to cope with complex conditions such as deformation, image blurring, and rotation. CN features [29] are better for tracking motion blur, light intensity changes, background confusion, etc. However, its performance will be degraded when encountering similar color distractions. Among the correlated filtered target tracking algorithms mentioned in this paper, some algorithms use HOG features alone to describe information
about the target, and some trackers use a fusion of both HOG and CN features. The SRDCF algorithm uses a joint fusion of HOG and CN features to achieve complementarity among the different features. Better tracking can be achieved if a small part of the target is occluded or under relatively stable lighting conditions. However, when the target is obscured in a large area or rotated to a large extent, the SRDCF algorithm will lose the target tracking. In order to cope with the performance degradation of the SRDCF algorithm when rotation occurs, the algorithm in this paper improves the target tracking accuracy by fusing three features, HOG + CN + LBP.

The texture information describes the essential structural information of the object surface and the connection with the external environment. The texture feature can improve the problem that the local information of the target edge is masked by fast target motion.

LBP [30] can extract local texture features from images. It was first proposed by Ojala et al. As a typical texture feature, LBP can effectively cope with some complex conditions, such as illumination changes and rotations, and the amount of calculation is small. Compared with HOG features, LBP features are more time-sensitive. However, it is easily disturbed by noise and is more sensitive to directional information. Therefore, in this paper, we consider introducing LBP features and fusing three features of shape, color, and texture as the input features of the algorithm to achieve complementary advantages among different features. The hybrid features enhance the robustness by exploiting the complementarity among multiple features so that the complementary features are still valid even when one feature is unreliable. Moreover, this complementary approach can accomplish tracking in most cases. Figure 3 shows the visualization results of the above three features in the Man video sequence.

3.2. Adaptive Spatial Regularization Correlation Filtering Algorithm Model. In order to solve the error problem introduced by SRDCF’s Gaussian-like weights due to significant changes in target appearance, this section introduces a spatial regularization term to allow the model to learn the actual contours of the target appearance adaptively. Moreover, by adding the spatial regularization term, the spatial
regularization weight can be updated adaptively, so that the filter can adapt to the target transformation well, so as to improve the accuracy of target tracking.

Our proposed objective function is as follows:

\[ E(f, w) = \frac{1}{2} \sum_{i} |x_i - f_i|^2 + \frac{1}{2} \|w \|_2^2 + \frac{\beta}{2} \|w - w_c\|^2. \]  

(4)

where \( w_c \) is the current reference weight penalty. \( \lambda \) and \( \beta \) are regular term coefficients. In the formula, except for the first term, a most minor squares term, the other two terms are spatial regularization terms. The second term imposes a spatial constraint weight on the correlation filter \( f \). The third is an adaptive spatial regularization term, which is introduced to make the adaptive spatial weight \( w \) as similar as possible to the reference weight \( w_c \). This constraint introduces a priori information about \( w \) and can learn a more accurate spatial penalty coefficient when the target changes, avoiding model degradation.

3.3. Model Optimization. (3) uses the Gaussian–Seidel method to solve the filter, but this method requires high convergence conditions, and the SRDCF algorithm is solved with its tracks very slowly. The ADMM method uses the idea of divide-and-conquer to divide a complex computational problem into several simple subproblems, effectively reducing the cost and the computational effort. Since (4) is a convex function that satisfies the necessary conditions for the iteration of the ADMM algorithm, the method is adopted to optimize the proposed model. Let \( \lambda = 1 \), and set \( \delta \) as the step parameter. We introduce an auxiliary variable, \( g \), setting the condition as \( f = g \). Its augmented Lagrangian function is shown in equation (5). Among them, the following equations are derived concerning the literature [21, 23]:

\[ L(f, w, g, \gamma) = \frac{1}{2} \sum_{i} |x_i - f_i - y|^2 + \frac{1}{2} \|w \|_2^2 + \frac{\beta}{2} \|w - w_c\|^2 + \frac{\delta}{2} \sum_{i} \|f_i - g_i + \gamma_i\|^2. \]  

(5)

where \( \gamma \) is the Lagrangian multiplier and \( \beta \) is the penalty factor. The problem is solved by ADMM, as shown in (6), and the problem is converted into three substructures:

\[
\begin{align*}
    f^* &= \arg\min_f \left( \frac{1}{2} \sum_{i} |x_i - f_i|^2 + \frac{\delta}{2} \sum_{i} \|f_i - g_i + \gamma_i\|^2 \right), \\
    g^* &= \arg\min_g \left( \frac{1}{2} \sum_{i} \|w \|_2^2 + \frac{\delta}{2} \sum_{i} \|f_i - g_i + \gamma_i\|^2 \right), \\
    w^* &= \arg\min_w \left( \frac{1}{2} \sum_{i} \|w \|_2^2 + \frac{\delta}{2} \sum_{i} \|w - w_c\|^2 \right).
\end{align*}
\]

(6)

The solution to the first subproblem \( f^* \) is as follows:

First, use Parseval’s theorem to convert to the Fourier domain as follows:

\[ E(\tilde{f}) = \arg\min_f \left( \|\sum_{i=1}^{L} \tilde{x}_i * \tilde{f}_i - \tilde{y}\|^2 + \delta \|\tilde{f} - \tilde{g} + \tilde{\gamma}\|^2 \right). \]  

(7)

Among them, the superscript \( \wedge \) represents its representation in the Fourier domain, and \( \tilde{f} \) is the discrete Fourier transform of \( f \). Vectorizing formula (7), we get:

\[ E(V_j(\tilde{f})) = \arg\min_{V_j(f)} \|V_j(\tilde{x})^T V_j(\tilde{f}) - \tilde{y}\|^2 + \delta \|V_j(\tilde{f}) - V_j(\tilde{g}) + V_j(\tilde{\gamma})\|^2. \]  

(8)
where the superscript \( T \) represents the transpose operation, and \( V \) represents the vectorization operation. Next, the objective function is derived so that the derivative is zero and the minimum value is found, as shown in the following equation:

\[
V_j(\tilde{f}) = \frac{V_j(\tilde{x})\tilde{y}_j + \delta V_j(\tilde{g}) - \delta V_j(\tilde{\gamma})}{V_j(\tilde{x})V_j(\tilde{x})^T + \delta I}.
\]  

(9)

Simplify using the Sherman–Morrison formula to

\[
V_j(\tilde{f}) = \frac{1}{\delta} \left( I - \frac{V_j(\tilde{x})V_j(\tilde{x})^T}{V_j(\tilde{x})^T V_j(\tilde{x}) + \delta} \right) P.
\]  

(10)

Among them, the representation of the vector \( P \) is

\[
P = V_j(\tilde{x})\tilde{y}_j + \delta V_j(\tilde{g}) - \delta V_j(\tilde{\gamma}).
\]  

(11)

The solution to the second subproblem of the above equation is as follows:

\[
\arg\min_w \sum_{l=1}^{L} \| w \odot g_l \|^2 + \delta \| f - g + \gamma \|^2.
\]  

(12)

Taking the derivative concerning \( g \) so that the derivative is 0, we get:

\[
g = \left( W^T W + \delta I \right)^{-1} (\delta f + \delta \gamma).
\]  

(13)

The spatial regularization constraint of the above equation is the adaptive spatial weight penalty term proposed in this paper, which has a more prepared representation of the profile information of the target.

The updated scheme of Lagrange multipliers is as follows:

\[
\hat{\gamma}^{i+1} = \tilde{\gamma}^{i} + \hat{\delta}^{i+1} - \tilde{f}^{i+1}.
\]  

(15)

For the choice of step parameters, \( \delta^{i+1} = \min(\delta^{max}, \rho \delta^{i}) \),

(16)

where \( \delta^{max} \) denotes the maximum value of \( \delta \), and \( \rho \) denotes the scale factor.

4. Experimental Results and Analysis

4.1. Experimental Environment and Configuration. In this paper, the adaptive space regularization correlation filtering target tracking algorithm is implemented in MATLAB. The specific experimental platform is as follows.

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 8G 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 consistent with the benchmark algorithm SRDCF, the initial sample size is set to \( M = N = 50 \), the cell of HOG is \( 4 \times 4 \), padding = 5, and the target search range is expanded to 25 times the target area. The learning rate is set to 0.014, the parameter of ADMM is 1, and the number of iterations of the alternating direction multiplier method is 2. The initial value of the step size parameter is \( \delta_0 = 1 \), and the maximum value is \( 10^{-5} \). The scale factor \( \rho = 0.12 \), the spatial regularization term parameter \( \lambda = 1 \), and the adaptive spatial regularization term hyperparameter \( \beta = 0.01 \).

4.2. Evaluation Indicators. In order to compare the tracking effect of this paper’s algorithm with several other algorithms, the OTB-2015 dataset and OTB-2013 dataset are used for testing. Among them, the OTB-2015 dataset contains 100 video sequences, and 11 tracking difficulties are designed. And the OTB-2013 dataset contains 51 video sequences, also involving video sequences with 11 complex attributes. The OTB dataset mainly uses two metrics of accuracy and success rate to evaluate the robustness of the tracker and uses frames per second (fps) to evaluate the efficiency of the tracker. Accuracy \([31]\) is defined as the percentage of the center location error (CLE) between the ground truth of the target and the predicted target location relative to the total number of frames less than a certain threshold, where the threshold is typically set to 20 pixels. 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 \((x_c, y_c)\):

\[
CLE = \sqrt{(x_b - x_c)^2 + (y_b - y_c)^2}.
\]  

(17)

The success rate is defined as the degree of overlap between the predicted target frame and the actual frame \([31]\). The threshold value is usually set to 0.5, and when less than 0.5 indicates tracking failure. There are three forms of robustness estimation for accuracy and success curves, namely one-pass evolution (OPE), temporal robustness evolution (TRE), and spatial robustness evolution (SRE). Usually, most of the algorithms use OPE as the evaluation.

4.3. Experiments on the OTB-2015

4.3.1. Quantitative Analysis

(1) Analysis of the Overall Experimental Results. The benchmarked algorithm of this paper’s algorithm is the SRDCF algorithm. In order to be more convincing, the tracking algorithm ourSRDCF is compared with several classical correlation filtering algorithms, such as CSK \([17]\), KCF \([19]\), SRDCF \([20]\), BACF \([21]\), STRCF\([23]\), TLD \([32]\), and SAMF \([33]\). The experimental comparison results are as follows.

From Figure 4, it can be seen that among the compared algorithms, the STRCF algorithm performs well in both tracking accuracy and success rate, while the ourSRDCF
algorithm improved in this paper also performs well in performance, with tracking accuracy and success rate reaching 0.86 and 0.819, respectively, which are 2.2% and 3.8% higher than the benchmark SRDCF algorithm, proving that the proposed algorithm outperforms the SRDCF algorithm overall. The number of frames per second processed by different algorithms is shown in Table 1.

Table 2 is compiled from Figure 4; the top three best-performing algorithms are shown in bolded black, blue, and green fonts. As seen from Table 2, in terms of accuracy, STRCF performs best with 89.2%, our RDF comes second, and the BACF algorithm ranks third. The improved algorithm also outperforms its benchmark method in terms of success rate.

(2) Experiment Results and Analysis of Each Attribute. There are 100 video sequences with 11 attributes in the test set, containing a variety of common practical and complex scenarios and tracking difficulties. The adaptive spatial regularization method for multifeature fusion proposed in this paper is tested on these 11 different attributes and compared with the SRDCF algorithm and several trackers such as BACF, STRCF, SAMF, and KCF. Figures 5 and 6 are the performance evaluation results of this algorithm and several other algorithms in 11 complex attributes.

From Figures 5 and 6, we can see that our method can adapt well to various complex tracking environments and achieve high-quality tracking. The improved algorithm in this paper has a robust performance in tracking results. In the occlusion property, the accuracy of the improved algorithm is slightly lower than that of the SRDCF algorithm, but the success rate is 0.6% higher than it. Overall, the improved algorithm in this paper performs well in terms of performance. The proposed accelerated SRDCF method has dramatically improved speed while expanding the search area, enabling real-time tracking even in complex environments. By incorporating the spatially adaptive regularization term with the feature fusion strategy, it can be seen that the tracking can be sustained even under the occurrence of rotation, scale change, or fast motion.

In order to reflect the tracking performance of the proposed algorithm compared with other algorithms, six video sequences, Couple, football, jumping, Shaking, Singer, and walking, are selected for the analysis of the central position error CLE.

Figure 7 compares the tracking CLE of unique algorithms in six video sequences where the smaller CLE indicates a minor tracking error.

The red ones in Table 3 are the best-performing algorithms; as shown in the above table, the center position error of ourSRDCF in this article is relatively the smallest among the six sequences. It shows that using a feature fusion strategy, ADMM model optimization, and adaptive space regularization can improve the performance of the SRDCF tracker.

4.3.2. Qualitative Analysis. This section will visually demonstrate the comparison between the proposed adaptive spatial regularization correlation filtering algorithm ourSRDCF and some current mainstream correlation filtering target tracking algorithms in some actual complex tracking scenarios. The evaluation algorithms are some of the influential related filtered target tracking algorithms in the field, including SAMF, KCF, TLD, CSK, and the benchmark algorithm SRDCF in this paper. For simplicity, only some video sequences are selected in the OTB dataset for visual display, as shown in the following figures.

Figure 8 shows the visual tracking results of several advanced algorithms in the “bolt” and “car4” video sequences. The “bolt” video sequence suffers from fast target motion and partial occlusion, as seen in frames 25, 192, and 268. Among several comparative algorithms, only the SAMF algorithm and ourSRDCF algorithm proposed in this paper can perform accurate tracking, corresponding to red and blue boxes, respectively, while other algorithms lose track. The successful tracking of the SAMF algorithm is attributed to the use of scale filters, and the successful tracking of ourSRDCF algorithm is attributed to the use of the ADMM method to solve the filters, which improves the real-time tracking performance of the algorithm. Only our SRDCF, SRDCF, and SAMF algorithms are used for robust tracking in the “car4” sequence. Among them, the SRDCF algorithm is successful in tracking because of the use of spatial regularization to improve the performance of the algorithm.
Table 1: The number of picture frames processed by the algorithm per second.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>ourSRDCF</th>
<th>SRDCF</th>
<th>TLD</th>
<th>CSK</th>
<th>SAMF</th>
<th>KCF</th>
<th>BACF</th>
<th>STRCF</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPS</td>
<td>37.741</td>
<td>5.7</td>
<td>31.32</td>
<td>297.765</td>
<td>15.346</td>
<td>169.72</td>
<td>38.97</td>
<td>22.54</td>
</tr>
</tbody>
</table>

The bolded values in 1 represent trackers with the fastest frame rates.

Table 2: Overall performance evaluation table of six algorithms.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>ourRDCF</th>
<th>SRDCF</th>
<th>TLD</th>
<th>CSK</th>
<th>SAMF</th>
<th>KCF</th>
<th>BACF</th>
<th>STRCF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>0.860</td>
<td>0.838</td>
<td>0.608</td>
<td>0.545</td>
<td>0.786</td>
<td>0.737</td>
<td>0.849</td>
<td>0.892</td>
</tr>
<tr>
<td>Success rate</td>
<td>0.819</td>
<td>0.781</td>
<td>0.521</td>
<td>0.433</td>
<td>0.737</td>
<td>0.611</td>
<td>0.822</td>
<td>0.862</td>
</tr>
</tbody>
</table>

The bold values in 2 represent the tracker with the highest tracking accuracy and success rate.

Figure 5: The accuracy curves of the six algorithms on the OTB-2015 dataset.
Figure 9 shows the tracking results in the “football” and “freeman” video sequences. The “football” has background clutter and occlusion challenges. At frames 87 and 285, several algorithms can still successfully locate the target. However, in the case of apparent rotation and background confusion at frames 295 and 362, only the algorithm ourSRDCF of this paper and the benchmark algorithm SRDCF are still able to frame the target accurately. The “freeman” sequence has rotation and partial occlusion, and only the three algorithms of SRDCF, our RDF, and SAMF
are successfully tracked. At frame 137, the SRDCF algorithm in the green frame does not track very well. At the same time, our SRDCF is largely unaffected, mainly because it incorporates LBP features that can effectively cope with large rotations.

Figure 10 shows the comparison results of several algorithms in “shaking” and “motorRolling.” The “shaking” video sequence suffers from a complex scene with a cluttered background and changing illumination. With the change of illumination intensity, only the algorithm in this paper can successfully locate accurately from beginning to end, while several other comparison algorithms have tracking drift. In the “motorRolling” video sequence, the target undergoes complex challenges such as fast movement and out-of-plane rotation. At the 10th frame, several algorithms can locate the target. However, in the subsequent frames 30, 35, and 125, when large out-of-plane rotations occur, only the algorithm in this paper can maintain robust tracking throughout.

![Figure 7: Comparison of center position errors of several tracking algorithms on 6 video sequences.](image)

![Table 3: The center position error of the tracking algorithm on 6 video sequences.](table)

<table>
<thead>
<tr>
<th>Algorithm/sequence</th>
<th>Couple</th>
<th>Football</th>
<th>Jumping</th>
<th>Shaking</th>
<th>Singer</th>
<th>Walking</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAMF</td>
<td>17.70</td>
<td>13.37</td>
<td>26.95</td>
<td>41.85</td>
<td>4.11</td>
<td>2.04</td>
</tr>
<tr>
<td>KCF</td>
<td>47.56</td>
<td>14.60</td>
<td>26.12</td>
<td>112.50</td>
<td>10.76</td>
<td>3.97</td>
</tr>
<tr>
<td>SRDCF</td>
<td>3.97</td>
<td>5.68</td>
<td>4.47</td>
<td>85.44</td>
<td>2.72</td>
<td>2.01</td>
</tr>
<tr>
<td>ourSRDCF</td>
<td>3.84</td>
<td>3.92</td>
<td>3.93</td>
<td>6.70</td>
<td>2.57</td>
<td>1.89</td>
</tr>
</tbody>
</table>

![Figure 8: Comparison of tracking results of six algorithms in “bolt” and “car4” video sequences.](image)
In contrast, several other current mainstream tracking algorithms fail. The main reason why the algorithm in this paper can cope with such complex tracking scenarios is attributed to the use of a multifeature fusion strategy and filter optimization. Moreover, incorporating adaptive spatial regularization enables the spatial regularization weights to be adaptively adjusted as the target profile changes, thus improving the tracking performance of the algorithm.

4.4. Experiments on the OTB-2013. To be more convincing, the overall tracking effect of our improved algorithm ourSRDCF and its benchmark algorithm SRDCF is tested. The test dataset for the experiments is OTB-2013, and 51 video sequences it contains were experimented on, including 11 video sequences with complex attributes. Figure 11 shows the accuracy and success rate comparison results of the two algorithms in the OTB-2013 dataset.

As can be seen from the above figure, among the two algorithms compared, the improved algorithm ourSRDCF in this paper performs well in terms of performance, with tracking accuracy and success rate reaching 82.7% and 80.2%, respectively, which are 3.4% and 6.3% improvement over the benchmark algorithm SRDCF. This shows that the proposed method has a significant improvement in the performance of the tracker.

4.5. Ablation Experiments. To explore the contribution of different components of the adaptive spatial regularization algorithm with improved multifeature fusion in this paper, we conducted an ablation study to verify the effectiveness of critical components in our tracker. Figure 12 shows the ablation experiment of the algorithm in this paper on the OTB-2015 dataset.

The basic framework of the algorithm in Figure 12 is the SRDCF algorithm, where ours1 represents the fusion of HOG features, CN features, and LBP features on this basis. It has no improvement in accuracy compared to the benchmark algorithm SRDCF but has a 3% improvement in success rate. ours2 represents the introduction of the ADMM module to optimize the filter based on ours1, which
improves the accuracy and success rate by 0.7% based on the ours1 method. Furthermore, ourSRDCF represents the adaptive spatial regularization algorithm for multifeature fusion; its success rate and accuracy are 81.9% and 86%, respectively.

5. Conclusions

This paper improves some shortcomings of the benchmark algorithm SRDCF. An adaptive spatial regularization target tracking algorithm is proposed and tested on datasets containing 100 video sequences and 50 video sequences, respectively. The improved algorithm is compared with the benchmark SRDCF algorithm and several other classical algorithms. Experiments prove that the three improvement ideas proposed in this article have a certain degree of promotion in tracking accuracy and success rate. By fusing several features as the input features of the algorithm, the advantages between different features are complemented. Meanwhile, adding the adaptive spatial regularization term can learn more accurate spatial penalty coefficients when the target changes, avoiding model degradation. In addition, the accelerated SRDCF algorithm based on ADMM improves the tracking speed while broadening the search scope, making the speed in this paper reach 37.741 FPS, which is about seven times higher than the benchmark algorithm SRDCF and achieves the tracking real-time effect. By comparing and analyzing the tracking results of several advanced tracking methods on the OTB-2015 dataset, the improved tracker performed well in terms of tracking precision and success rate, reaching 86% and 81.9%, respectively. The experiments in the OTB2013 dataset show that the improved algorithm achieves 82.7% and 80.2% in accuracy and success rate, respectively, and its improvement over the benchmark algorithm SRDCF is 3.4% and 6.3%, respectively. It is again verified that the improved algorithm not only has a significant improvement in tracking speed and accuracy but also can better adapt to the changes of the target, while effectively reducing the computational overhead and achieving the tracking real-time requirements.

By adding an adaptive spatial regularization term, the filter can adapt to changes in the target to a certain extent, effectively highlighting the target region and suppressing the background region. However, when the target changes significantly, it may cause the weight penalty term to change abruptly between adjacent frames. Therefore, the next step can be considered to introduce temporal regularization to establish the connection between adjacent frames filters to slow down the drastic changes occurring between adjacent frames, further enhance the discriminative ability and tracking stability of the filter, and can effectively alleviate the
problem of filter model degradation over time. The paper only uses standard manual features, which have good robustness in correlation filtering target tracking. However, they cannot precisely characterize one feature of the image, and it is challenging to represent the exclusive feature of the target. When using only standard features, the semantic information of the target is weakened, which reduces the tracking robustness of the algorithm when facing complex scenes. In contrast, depth features contain more stable semantic information, so combining both manual and depth features can enhance the algorithm’s performance.

Data Availability

A publicly available dataset was analyzed in this study. Our dataset can be obtained from https://cvlab.hanyang.ac.kr/tracker_benchmark/datasets.html, August 25, 2021.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work has been supported by the National Natural Science Foundation of China (62166042, U2003207), Natural Science Foundation of Xinjiang, China (2021D01C076), and Strengthening Plan of National De-fense Science and Technology Foundation of China (2021-JCJQ-JJ-0059).

References


