

Research Article

A Visual Tracking Method Based on an Adaptive Overlapping Correlation Filter for Robotic Real-Time Cognitive Imaging

Yihua Lan, Pianpian Ma, Anfeng Xu, and Jinjiang Liu 

School of Computer and Information Technology, Nanyang Normal University, Nanyang 473061, China

Correspondence should be addressed to Jinjiang Liu; jjliunynu@sina.com

Received 6 June 2020; Revised 13 July 2020; Accepted 31 July 2020; Published 24 August 2020

Academic Editor: Yin Zhang

Copyright © 2020 Yihua Lan 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.

Computer vision is a very important research direction in the cognitive computing field. Robots encounter various target-tracking problems with computer vision systems. Robust scale estimation is an important issue in tracking algorithms. Most of the available methods have difficulty addressing even reasonable changes of scale in complex videos. In this paper, we propose a visual tracking method based on robust scale estimation, which uses a discriminant correlation filter based on a time-dependent scale-space filter and an adaptive cross-correlation filter. The tracker uses separate essential filters for sample migration and scale estimation. Furthermore, the built-in scale estimation method can be introduced into other tracking algorithms. We validate the proposed method on the UAV123 dataset. The results of comparison experiments with the traditional correlation filter tracking method demonstrate that the proposed method improves the success rate and tracking accuracy while controlling the computational complexity; its success rate measured by the area under the curve is 0.638, while at a location error precision of 20%, it is 0.649.

1. Introduction

Computer vision is the basis of cognitive imaging and processing for robots. Robots interact with the external environment in real time through computer vision systems. After acquiring accurate control information, they can control themselves through a closed-loop process to achieve various tasks requiring artificial intelligence, including positioning, navigation, formation, cooperation, group intelligence, and even tasks requiring intelligent decisions and collaborative work. Therefore, the study of computer vision—especially real-time moving object tracking—is an important research field in robot cognitive computing. Through many unremitting scholastic efforts, real-time target-tracking technology has made considerable progress. However, due to the complexity of visual tracking systems and the variability of the targets themselves, robust target tracking is difficult.

Despite great progress in recent years, the target-tracking problem remains intractable, mainly due to partial occlusion interference, deformations, motion blur, illumination changes, and cluttered backgrounds in the video image sequences. In particular, scale changes are difficult to esti-

mate during the tracking process. Similar problems also arise in environment identification [1], image reconstruction [2], and intelligent systems [3].

Research on visual tracking algorithms is ongoing and has made great progress in terms of stability and accuracy, but many problems remain to be solved. Because image processing involves the use of large quantities of data, to achieve the goals of better image recognition and tracking, it is necessary to implement more complex algorithms with more computations, which leads to the development of tracking algorithms with higher complexities. This higher algorithm complexity generally results in greater stability and accuracy but degrades real-time performance. Practical applications involving visual tracking usually have stringent real-time tracking and accuracy requirements, but the lag in hardware development makes it difficult for complex tracking algorithms to run in real time on embedded hardware. Therefore, image tracking must balance algorithm complexity with the available hardware computing power to achieve a satisfactory tracking effect.

In 2010, Bolme et al. [4] first introduced the correlation filter in signal processing into target tracking and proposed

the minimum output sum of squared error (MOSSE) filter method, whose discrimination model is based on the least square error, thereby creating a new tracking approach. The principle of this model is simple, it operates rapidly, and it is better at distinguishing the target from the background. Therefore, a number of follow-up studies based on this method have introduced improvements. In 2012, Henriques et al. [5] proposed the circulant structure of tracking-by-detection with kernel (CSK) algorithm based on MOSSE; in this approach, ridge regression was introduced as the loss function using the kernel technique. CSK optimized the training objective and derived the closed-form solution of the correlation filter. Furthermore, it greatly simplified the matrix multiplication operation in the Fourier domain by using the characteristics of the circulant matrix while retaining the speed advantage of MOSSE and achieving an improved tracking effect.

In 2014, Henriques et al. [6] improved the single channel filter of CSK to a multidimensional filter and proposed a new method called the kernel correlation filter (KCF) algorithm that updated the one-dimensional grayscale feature of CSK with the multidimensional histogram of oriented gradient (HOG) feature. Through experiments, this study fully verified that KCF substantially improved tracker performance.

In 2014, Danelljan et al. [7] proposed a new idea for scale-space filtering based on the correlation filtering algorithm that won the 2014 visual object tracking (VOT) competition. The algorithm is simple, with excellent performance and high portability. Compared with MOSSE, KCF, and other algorithms, the algorithm introduces two main contributions: a multifeature fusion mechanism and a relatively fast scale-space filtering optimization method.

More recently, in 2017, Danelljan et al. [8] proposed a new dimension-reduction tracking method, the efficient convolution operators (ECO), which simplifies feature extraction. Simplifying the feature set greatly reduced the calculations. Furthermore, the algorithm also stores historical target features in a manner that simplifies the feature set, which greatly improved tracking robustness.

Although tracking systems based on correlation filtering have made great progress, they have failed to make a breakthrough in efficient scale-space filtering [9] while maintaining tracking robustness. In the past three years, many researchers have investigated various tracking problems using neural networks [10], filter integration [11], filter channelization [12], deep learning [13], and other approaches. These newer studies provided many good ideas for our work.

In this paper, we present a tracking method that effectively estimates target scale by training a scale classifier. After determining the optimal target location, the target scale can be estimated independently. This method improves precision through an efficient scale-space search. The improvements of this article mainly include the following: (1) We propose a spatial filtering prediction technology based on time domain correlation. Using this technology substantially reduces the calculations needed for spatial filtering. (2) We also propose an adaptive overlapped filter strategy that avoids many unnecessary filter update calculations.

The paper is organized as follows. The second chapter mainly introduces the principle and skills of the discriminant correlation filter in the tracking algorithm. These principles and techniques will run through the following algorithm improvements. The third chapter introduces the proposed improvement method. In chapter four, a series of experiments compared with some existing traditional typical tracking algorithms are carried out to demonstrate the efficiency of the proposed method.

2. Learning Discriminant Correlation Filters

Current tracking algorithms are composed of five main parts: motion modeling, feature extraction, observation modeling, template or filter updating, and postprocessing [14]. Motion modeling is used to model the target motion. By predicting the target's position in the next frame, the corresponding target candidate search area can be obtained. Feature extraction uses a series of feature vectors to represent candidate image data while removing redundant features and retaining effective features. Observation modeling uses features extracted from candidate images to determine whether they represent the target or the background. Updating controls the strategy and frequency of model updating and balances model adjustment and tracking migration to maintain tracking accuracy while considering tracking robustness. When a tracking system includes multiple trackers, postprocessing integrates the results of each tracker into a final optimal tracking state.

The MOSSE tracker learns a discriminant correlation filter to locate a new frame's target position [4]. The method uses a series of image patches of the target's appearance (r_1, r_2, \dots, r_t) as training samples. These sample tags are associated with the filter output ($out_1, out_2, \dots, out_t$). The goal is to find a filter that maximizes the response to the target, that is, to satisfy $out = r \odot h$. The optimal correlation filter OCF_t for a time series t is obtained by the sum of the minimum mean square error:

$$\varepsilon = \sum_{i=1}^t \|\text{OCF}_t \odot f_i - out_i\|^2 = \frac{1}{MN} \sum_{i=1}^t \|\bar{H}_t F_i - G_i\|^2. \quad (1)$$

where the functions r_i , out_i , and h_t all have $M \times N$ dimensions. The symbol \odot indicates a cyclic correlation. By $0 = (\partial/\partial H_t) \sum_{i=1}^t \|\bar{H}_t F_i - G_i\|^2$, Equation (1) can be minimized by the following filter model:

$$H_t = \frac{\sum_{i=1}^t \bar{G}_i F_i}{\sum_{i=1}^t \bar{F}_i F_i}. \quad (2)$$

The relevant output, out_i , is constructed by a Gaussian function whose peak value lies at the target position, r_i . In Equation (2), the numerator $\sum_{i=1}^t \bar{G}_i F_i$ and denominator $\sum_{i=1}^t \bar{F}_i F_i$ of H_t are updated separately by the weighted means of the new observations r_t .

Given a new image z , $y = \Psi^{-1}\{\bar{H}_t Z\}$ is used to calculate the correlation score y , where Ψ^{-1} denotes the inverse discrete Fourier transform (DFT) operation. The new target

position can be estimated by the maximum result of the correlation score y . The fast Fourier transform (FFT) is used to implement efficient training and searching.

On this basis, the ridge regression method is introduced into the least square problem of Formula (1) to form the biased estimation regression method. By giving up the unbiasedness of the least square, we can obtain an optimal filter fitting method with better tolerance to ill-conditioned data and more reliable calculations.

Let f represent a rectangular region of the target extracted from the feature image. The dimension of f is expressed as f^j . The goal is to find the optimal correlation filter ocf , which consists of one filter ocf^j for each feature dimension. The optimal correlation filter ocf can be obtained by minimizing the following loss function:

$$\varepsilon = \left\| \sum_{j=1}^p ocf^j \odot f^j - \text{out} \right\|^2 + \lambda \sum_{j=1}^p \|ocf^j\|^2, \quad (3)$$

where out is the relevant output of training sample f and the parameter $\lambda \geq 0$ controls the regularization. The solution to Equation (3) is

$$H^l = \frac{\bar{G}F^l}{\sum_{k=1}^d F^k F^k + \lambda}. \quad (4)$$

Adding the regularization term helps avoid overfitting and ill-conditioned solutions. Equation (4) still involves a matrix inversion process. Samples can be obtained by using the cyclic shift matrix properties to avoid inversion. Assume that $F = \Gamma \text{diag}(\hat{f})\Gamma^H$,

$$H = \Gamma \cdot \text{diag} \left(\frac{\hat{f}}{\hat{f} \odot f \wedge^* + \lambda I} \right) \cdot \Gamma^H \cdot g. \quad (5)$$

Then, the convolution property of a cyclic matrix can be used to obtain the frequency domain display solution. Another advantage of the cyclic matrix method is its selection of positive and negative samples. The tracking algorithm trains the classifier by online learning. In each frame, appropriate positive samples and negative samples are selected for classifier training. Generally, the labels of negative samples are assigned a 0, while the labels of positive samples are assigned a 1. Theoretically, when more samples are collected, the tracker discrimination ability will be stronger. However, due to the time sensitivity of tracking, a modern tracker has to balance the number of samples with the amount of computation. Therefore, the common approach is to select only a small number of samples randomly from each frame. However, an insufficient number of samples adversely affect the tracker's judgment. When the circular matrix method is used, a sample can be used as a "base sample" to "copy" thousands of similar samples through circular displacement, but these samples are related to the base samples only after they are transferred to the Fourier domain, that is, these samples will not cause a calculation increase when they are calculated in

the Fourier domain. Therefore, this method can obtain a nearly unlimited number of samples from an image without introducing large amounts of extra computation.

3. The Proposed Efficient Scale-Space Filtering

In this chapter, Section 3.1 briefly introduces the multiscale tracking method based on the feature pyramid; Section 3.2 gives the basic process of multiscale tracking based on the interactive iteration of position filter and scale filter. On this basis, it is pointed out that the main calculation cost of the algorithm will be greatly increased due to the introduction of a scale filter. In Section 3.3, a precise scale estimation method is introduced, and the search range of the scale level is greatly reduced by filtering and predicting the target scale on the time axis; in Section 3.4, in order to reduce the calculation process and tracking drift further, the improved algorithm changes the iterative calculation process of the position filter and scale filter and proposes an adaptive overlapping correlation filtering method based on deviation prediction on time scale.

3.1. Standard Scale-Space Tracking. An improved tracking method is based on learning a three-dimensional-scale spatial correlation filter. The scale of this filter is fixed at $X \times Y \times Z$, where Z is the number of scales. To update the filter, the feature pyramid of the rectangular region around the target is calculated. To obtain the target position in the new frame, the $X \times Y \times Z$ rectangular cuboid is extracted from the feature pyramid as described above. In theory, the more samples that are collected, the stronger the discriminative ability of the trained tracker is. However, due to the time sensitivity of tracking, the tracker must balance the number of samples with the computational complexity.

3.2. Discriminative Scale-Space Tracking. Two correlation filters (a position filter and a scale filter) are used for target location and scale evaluation. The two filters are relatively independent; therefore, different feature types and feature calculation methods can be selected during their training and testing. Training sample f , which is used to update the scale filter, is obtained by extracting features of different image sizes around an object at the center. Similar to the spatial filter and the scale filter defined in Section 3.1, we can extract a target image at the center position and a size of s , where a represents the scale factor. The iterative process of the scale filter and spatial filter undoubtedly improves the efficiency of the scale-space search. Because the scale parameter is an exponential function, the scale size does not increase linearly: the larger the scale parameter is, the larger the search step is. In contrast, the smaller the scale parameter is, the finer the scale-space searches are. That is, coarse detection is conducted at a larger scale, while fine detection is conducted at a smaller scale.

3.3. Scale-Space Search Based on Time Association. The introduction of scale estimation into the tracker leads to a greater computational cost. Ideally, an accurate scale estimation method should be both robust and efficient simultaneously.

TABLE 1: Analysis of traditional tracking in the UAV123 database without scale estimation (showing some videos with typical tracking problems).

Video name	Tracking failure location	Failure reason description	Video name	Tracking failure location	Failure reason description
Car1	1/3	Scale+background change	Person12	3/5	Shelter
Car3	4/5	Background accumulation	Person14	1/10	Shelter
Car6	1/3	Out of sight	Person16	1/5	Shelter
Car9	2/5	Occlusion+scale change	Person21	1/100	Small target
Car11	1/100	Small target	Person1_s	1/3	Similar target
Car15	1/100	Small target	Truck1	1/2	Illumination change and deformation
Car16	1/5	Rapid deformation+scale change	Truck2	1/3	Shelter
Person4	1/3	Scale change, severe deformation	Wakeboard3	1/3	Scale+deformation+rapid movement
Person6	9/10	Deformation and scale accumulation	Wakeboard5	1/3	Rapid movement
Person9	1/5	Out of sight	Wakeboard8	1/3	Scale change+background learning

In a visual tracking scene, the scale difference between two frames is usually smaller than that of the positional difference because the relative distance between the target and the camera in the tracking process generally does not change dramatically over short interframe periods. For the target-tracking process in a monitoring scene, in which the camera is typically fixed and the target moves, the target scale changes are mainly caused by the advance or retreat of the target. In contrast, during the dynamic tracking process when the camera is in motion but the target remains relatively still, scale changes to the target are mainly caused by changes in the optical parameters of the camera. However, the scale changes produced by these processes have their own specific laws. We use the Kalman filter method to filter and predict the target scale on the time axis, which greatly reduces the search range at the scale level. The adaptive filter is used to control the size of the search interval. When the residual is reduced, the search range is narrowed; conversely, when the residual is larger, the search range is enlarged.

3.4. Adaptive Cross-Correlation Filtering. When scale filtering is added to the traditional spatial filtering target-tracking process, the entire filtering calculation process is multiplied, and the real-time advantage of correlation filtering is significantly reduced. To reduce unnecessary calculations and improve the efficiency of the complete filtering process, we propose an adaptive overlapping correlation filtering method. In this filtering process, the traditional alternating calculations of the position and scale filters in each frame are changed to calculate either only the position filter or only the scale filter in a single frame. The specific filtering process depends on the deviation predictions of the position filter and the scale filter in the time scale. Generally, when the target's position changes rapidly, the scale change is usually small. In contrast, when the scale changes drastically, the position is usually undisturbed. Finally, when both the posi-

tion and the scale changes are large, the position filter and scale filter should be memorized for a brief period instead of being refreshed quickly to reduce the tracking drift.

4. Experiments and Analysis

In this study, we compare our novel and fast scale estimation-based tracking method with existing traditional typical tracking algorithms.

4.1. Analysis of the Test Dataset. The commonly used datasets for target tracking include the OTB and VOT series; however, larger datasets are also currently available: LASOT and TrackingNet. In addition, there is a dataset called UAV123, which is a special scene dataset, all of whose images were acquired by unmanned aerial vehicles (UAVs). These images are characterized by clean backgrounds and numerous perspective changes. The dataset has a total size of approximately 13.5 Gb.

First, we tested the common correlation filtering algorithms on the UAV123 database without scale estimation. Analysis of traditional tracking in the UAV123 database without scale estimation is shown in Table 1. As seen from Table 1, the reasons for target-tracking failures in the image sequence mainly involve background error accumulation, scale changes, rapid scale changes, occlusion, small targets, similar targets, fast movement, illumination changes, and the target moving out of the image area.

By further investigating the above reasons for errors, we find the following:

- (1) A small target is usually related to initialization failure; a small target has few pixels and little feature information; thus, tracking fails quickly either during initialization or close to the start of tracking

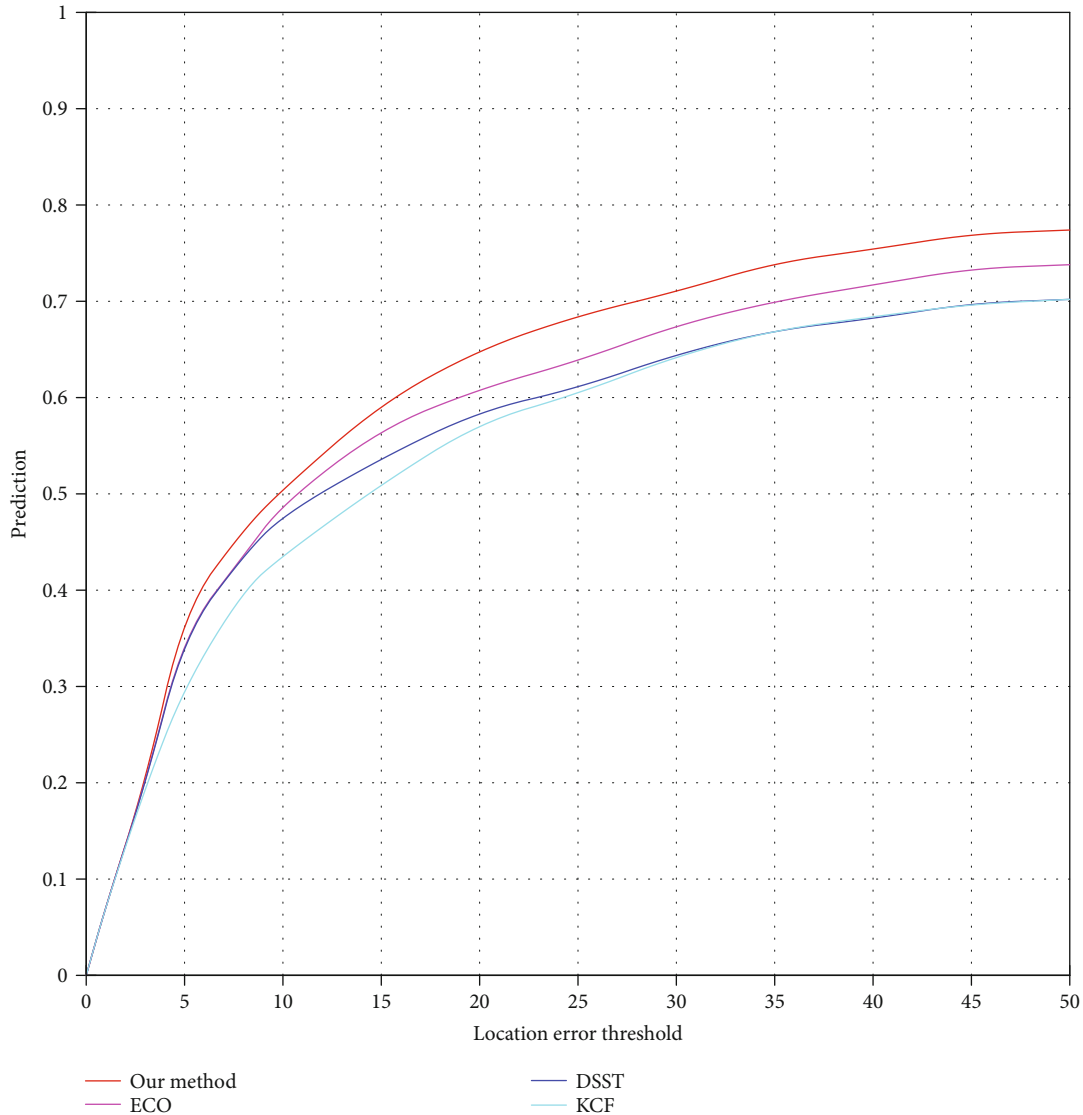


FIGURE 1: Accuracy curves of the compared algorithms on the UAV123 dataset.

- (2) For occlusions and situations in which the target leaves the image area, tracking inevitably fails. This condition needs to be solved by adding a target re-detection module
- (3) Simultaneous rapid movement and target deformation will also lead to target-tracking failure. This is because the size of the search area is limited. When rapid movement occurs, the center of the target is displaced by a large amount, which may cause the target to cross the search boundary. When rapid deformation occurs, the appearance information of the target changes substantially, allowing the filter insufficient learning time; this leads to target-tracking failure
- (4) Correlation filtering is able to resist changes in illumination, but when the target illumination suddenly changes dramatically, such as when the target enters a shadowed area from sunlight, tracking may fail
- (5) Regarding the influence of similar targets, when two similar targets are close together or cross paths, the filter may fail to track the correct target
- (6) Background error accumulation and scale change are also issues. Because the selected target area includes some background information, when the target appearance or scale changes, the filter will continuously learn more background information. Consequently, over time, these errors accumulate, eventually resulting in tracking failure

4.2. *Experimental Configuration.* The performance of the proposed algorithm is verified quantitatively by standard experimental parameters [15]. The experimental results are

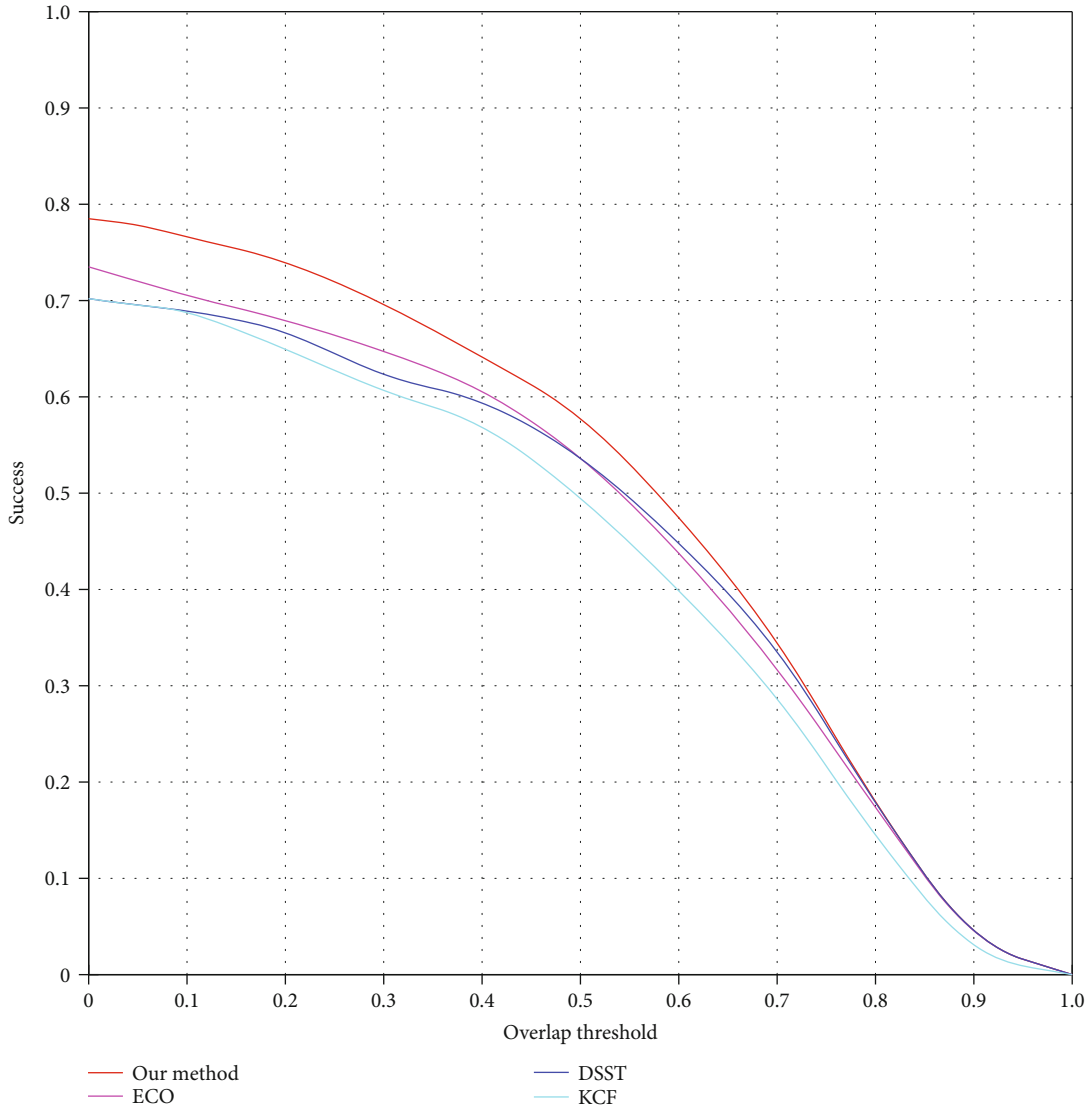


FIGURE 2: Success rate curves of the compared algorithms on the UAV123 dataset.

represented by the metrics of distance accuracy (DP), central position error (CLE), and overlap accuracy (OP). Additionally, the tracker's speed is represented by the median frame rate (FPS) in the video frames. We also report the results in terms of accuracy and through success rate graphs [10].

The DP is the ratio of the number of test video frames whose central position error is less than a threshold number of pixels to the total number of test video frames. This index represents the overall stability of the visual tracking process.

CLE represents the Euclidean distance between the tracking algorithm output target center and the calibrated target center. This index reflects the degree of coincidence between the tracking center and the actual target center, and it is one of the descriptive indexes of tracking accuracy.

In the tracking algorithm, the OP of each video frame is equal to the ratio of the intersection between the output frame area and the calibration frame area to the area of the union between the output frame area and the calibration frame area. Generally, an overlap ratio greater than 0.5 repre-

sents tracking success for the current frame; otherwise, it is considered a tracking failure. This index indicates the overall correctness of the visual tracking process.

FPS reflects the rapidity of the tracking algorithm.

4.3. Comparison with Traditional Tracking Algorithms. We performed this experiment on the UAV123 dataset. This dataset consists of many video sequences depicting a variety of scenes, such as biking, boating, driving a car, groups of people, trucks, individuals, and wake boarding, all of which were acquired using a UAV. The dataset includes changes in target scale, fast-moving targets, small targets, target shape changes, and target occlusions.

The comparison algorithms include our proposed algorithm, the ECO algorithm, the visual tracking via discriminative sparse similarity map (DSST) [16] algorithm, and the KCF algorithm.

The parameters are set as follows: padding: 4; HOG features: cell_size = 6, hog_orientations = 9; compression

TABLE 2: Comparison of algorithm accuracy and speed in the UAV123 dataset.

Methods	DSST	KCF	ECO	Our method
P20	0.582	0.569	0.609	0.649
AUC	0.579	0.565	0.605	0.638
FPS	28	42	24	39

dimension: 10; CN features: cell_size = 4; compression dimension: 3; CG iterations: 5; ideal Gaussian output Sigma: 1/16; learning rate: 0.009; and sample space size: 30.

Figures 1 and 2, respectively, show the tracking accuracy curves and success rate curves of the four algorithms (including our new algorithm) on the UAV123 dataset. As the figures show, the tracking accuracy and success rate of the proposed algorithm are obviously better than those of the other algorithms. This result occurs primarily because the correlation filtering in the new algorithm uses the characteristics of the target area to “match” in the next frame, and these good characteristics make the target and the complex background easier to discriminate. In addition, the new algorithm has high discrimination and robustness capabilities for both multidimensional information features and spatial scale features. Table 2 also shows that in the tracking accuracy P20 data, the new algorithm achieves 0.649, which is significantly higher than the other algorithms. Regarding the success rate shown by the area under the curve (AUC) data, the new algorithm achieves 0.638, which is also significantly higher than the other algorithms. For FPS, the highest score is 42, by KCF, but the new algorithm achieves a score 39, which is not considerably different from the KCF score, showing that the new algorithm does not significantly increase the amount of computation after adding the efficient scale filtering using the strategy described above.

5. Conclusion

In this paper, we present a novel scale estimation algorithm for visual tracking. This article uses a visual tracking algorithm based on a robust scale estimation process, which uses a discriminant correlation filter based on a time-dependent scale-space filter and an adaptive cross-correlation filter. Compared with the traditional filter approach, the proposed tracker provides a better overall performance and improves the computational efficiency. Moreover, because the scale estimation process presented in this paper is independent, it can easily be introduced into any tracking algorithm. Our algorithms show good stability and robustness. In the future work, we plan to study how to address tracking tasks in more complex environments and improve the capability of visual tracking in quickly changing scenes.

Data Availability

The data used to support the findings of this study are available from this website: [https://cemse.kaust.edu.sa/ivul/benchmark-and-simulator-uav-tracking-dataset\(UAV123\)](https://cemse.kaust.edu.sa/ivul/benchmark-and-simulator-uav-tracking-dataset(UAV123)).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is financially supported by the Cultivation Program for Youth Backbone Teachers in Colleges and Universities of Henan Province (Grant No. 2019GGJS184) and the Key Technologies R&D Program of Henan Province (Grant No. 182102310752).

References

- [1] Y. Sakai, H. Lu, J.-K. Tan, and H. Kim, “Recognition of surrounding environment from electric wheelchair videos based on modified YOLOv2,” *Future Generation Computer Systems*, vol. 92, pp. 157–161, 2019.
- [2] H. Lu, Y. Li, T. Uemura, H. Kim, and S. Serikawa, “Low illumination underwater light field images reconstruction using deep convolutional neural networks,” *Future Generation Computer Systems*, vol. 82, pp. 142–148, 2018.
- [3] Y. Zhang, R. Gravina, H. Lu, M. Villari, and G. Fortino, “Pea: parallel electrocardiogram-based authentication for smart healthcare systems,” *Journal of Network and Computer Applications*, vol. 117, pp. 10–16, 2018.
- [4] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, “Visual object tracking using adaptive correlation filters,” in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2544–2550, San Francisco, CA, 2010, IEEE.
- [5] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, “Exploiting the circulant structure of tracking-by-detection with kernels,” in *European conference on computer vision*, pp. 702–715, Florence, Italy, 2012, Springer.
- [6] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, “High-speed tracking with kernelized correlation filters,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583–596, 2015.
- [7] M. Danelljan, G. Häger, F. Khan, and M. Felsberg, “Accurate scale estimation for robust visual tracking,” in *British Machine Vision Conference*, Nottingham, 2014BMVA Press.
- [8] M. Danelljan, G. Bhat, F. S. Khan, and M. Felsberg, “ECO: efficient convolution operators for tracking,” in *IEEE Conference on Computer Vision & Pattern Recognition*, Los Alamitos, CA, USA, 2017.
- [9] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, “Discriminative scale space tracking,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 8, pp. 1561–1575, 2017.
- [10] Z. Zhu, G. Huang, W. Zou, D. Du, and C. Huang, “Uct: learning unified convolutional networks for real-time visual tracking,” in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 1973–1982, Venice, 2017.
- [11] B. Uzcent and Y. Seo, “EnKCF: Ensemble of kernelized correlation filters for high-speed object tracking,” in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1133–1141, Lake Tahoe, NV, USA, 2018, IEEE.
- [12] A. Lukezic, T. Vojir, L. Cehovin Zajc, J. Matas, and M. Kristan, “Discriminative correlation filter with channel and spatial reliability,” in *Proceedings of the IEEE Conference on Computer*

Vision and Pattern Recognition, pp. 6309–6318, Honolulu, HI, USA, 2017.

- [13] J. Valmadre, L. Bertinetto, J. Henriques, A. Vedaldi, and P. H. Torr, “End-to-end representation learning for correlation filter based tracking,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2805–2813, Honolulu, HI, USA, 2017.
- [14] R. Wenbi, R. Chunyang, and X. Qiang, “Correlation filtering target tracking based on color and part spatial relation constraints,” in *Proceedings of the 2nd International Conference on Intelligent Information Processing -IIP'17*, Bangkok, Thailand, 2017.
- [15] K. Zhang, Q. Liu, Y. Wu, and M.-H. Yang, “Robust visual tracking via convolutional networks without training,” *IEEE Transactions on Image Processing*, vol. 25, no. 4, pp. 1–1792, 2016.
- [16] B. Zhuang, H. Lu, Z. Xiao, and D. Wang, “Visual tracking via discriminative sparse similarity map,” *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1872–1881, 2014.