ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0028

A High-confidence Model Update Method for Kernel Correlation Filter Trackers

Lin Zhang[1, a], Zeng Xin[1, b], Zhongqiang Luo[1, 2, c, *], Xingzhong Xiong[1, d]

[1]School of Automation and Information Engineering, Sichuan University of Science and Engineering, Yibin, China

[2]Key laboratory of Higher Education of Sichuan Province for Enterprise Informationalization and Internet of Things, Sichuan University of Science and Engineering, Yibin, China

[a]1562204902@qq.com, [b]zincedu@qq.com, [c]luozhongqiang@suse.edu.cn, [d]xzxion@suse.edu.cn

Abstract

The correlation filter trackers have limited ability to represent the appearance model of the target, cannot effectively describe the appearance change. Besides, the model update strategy is unreliable and easily leads to model drift. To solve the above problems, we propose a high-confidence model update method for kernel correlation filter trackers. First, we assign different weights to each feature and perform weighted feature fusion to improve the overall tracking performance. Secondly, in the model update stage, we propose a reliable model update strategy that can avoid the problem of model occlusion during the model update process. We conduct comparative experiments on the OTB-2013 and OTB-2015 datasets. The experimental results prove that the tracking performance of this method is better than other trackers. It has strong robustness to deformation and occlusion, and can effectively improve processing efficiency.

Keywords

Kernel correlation filter; High-confidence; Model update; Feature fusion; Model occlusion.

1. INTRODUCTION

Visual object tracking is a basic research field in computer vision and video processing. It is a method of tracking specific objects from a video or frame sequence and widely used in intelligent monitoring, human-computer interaction, visual navigation, medical diagnosis, and so on. Recently, with the rapid development of artificial intelligence technology, the problem of moving target tracking has received more and more attention. However, due to various factors, including partial occlusion, deformation, large-scale changes, lighting, clutter, fast motion, and motion blur, designing a reliable and robust tracking algorithm is still a challenging problem.

According to its working principle, the target tracking algorithm is divided into two types: generative method and discriminant method. The generative method is mainly to model a given target area in the initial frame, and search for the most similar part of the model in the subsequent frames as the predicted target position. The more famous ones are Kalman filter [4], particle filter [5] and mean-shift [6] and other algorithms. The discriminative method is to treat the target tracking problem as a target detection task in each frame. It uses the image features of the tracking target to train a classifier, and uses the trained classifier to find the optimal solution in subsequent frames. In the tracking process, the classifier is continuously updated with the tracking results in each frame. Among them, the most representative kernel correlation filter has attracted wide attention from researchers due to its fast and efficient tracking performance. It uses a cyclic matrix to generate samples to train the classifier, and uses fast Fourier transform to accelerate the algorithm. However, this algorithm uses a single feature to describe the appearance model of the target, and cannot effectively describe the changes in the target's appearance and background information during the movement. In addition, the kernel correlation filter tracking algorithm does not evaluate the reliability of the sample, and its model update rate is fixed, which easily leads to model drift during the tracking process and cannot effectively deal with the occlusion and deformation of the target.

In this paper, we consider the problems mentioned above and propose a high-confidence model update method for kernel correlation filter trackers. The main contributions of our work can be summarized as follows: 1) First, we use multi-feature fusion to represent the target appearance model, and determine the weight coefficient corresponding to each feature through the response value of each feature; 2) In addition, we adopt the method of scale pool to achieve scale adaptation. 3) At the same time, a reliable model update strategy is proposed. The model is updated only when the peak and average peak-to-correlation energy of the response map is greater than a certain threshold. This strategy can effectively avoid the problem of model occlusion during model update.

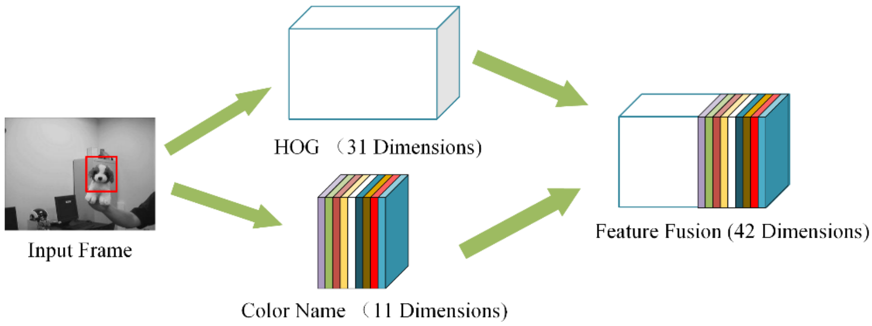

Figure 1. The feature combines the Color Name (11dimensions) feature with the HOG (31dimension) feature.

2. RELATED WORK

In recent years, the correlation filtering target tracking method has been favored by many scholars. Correlation filtering algorithms convert convolution in the time domain into dot products in the frequency domain, which effectively improves the calculation speed. The first to introduce correlation filtering theory in the field of target tracking is the MOSSE tracker proposed by Bolme et al. [7]. This tracker uses grayscale features, and the tracking speed is very fast, up to 669FPS. However, the characterization ability of gray-scale features is not enough to handle situations where the background is complex or the target and the background color are similar. Since then, Henriques [8] et al. improved the MOSSE algorithm by introducing a circulant matrix and a kernel function to improve the accuracy of the algorithm while achieving high-speed tracking. Danelljan et al. [9] extended the single-channel gray feature to the 31-dimensional Histogram of Oriented Gradient (HOG) feature, so that the surface texture feature and contour shape of the target can be well described by the HOG feature and further enhance tracking performance. Danelljan et al. [10] also used the color attributes of the target object to learn adaptive correlation filters by mapping multi-channel features to the Gaussian kernel space to reduce the impact of illumination and occlusion on color distortion. With the deepening of research, people gradually discovered that there is a performance bottleneck in tracking accuracy and speed using separate features. Bertinetto et al. [11] proposed a Staple tracker, which combines color features and HOG features for real-time tracking. Danelljan et al.[12] extended the feature maps of different resolutions to the continuous spatial domain of the same period through interpolation, which can accurately perform sub-pixel positioning. [13]

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0028

reduced the feature dimensions of HOG, CN and CNN to varying degrees, which reduced the number of parameters in the model and effectively reduced the computational complexity. In order to solve the problem that the tracker cannot track the target adaptively, [14-17] introduced multi-scale adaptive estimation of target scale. Liu et al. proposed the RPAC [18] algorithm on the basis of KCF, which decomposes the target into local targets, uses multiple KCF trackers to track the local targets, and estimates the change of target scale by calculating the change in the maximum response score in each response graph. Danelljan et al. [19] proposed a discriminant scale space tracker, which divides the tracking task into translation estimation and scale estimation, and uses position filters and scale filters to perform target positioning and scale estimation respectively. Correlation filter tracking algorithms have undesirable boundary effects due to the periodic assumption of training samples. [20]-[22] introduced regularization terms to eliminate the effects of boundary effects. In addition, Dai et al. [23] proposed an adaptive spatial regularization correlation filter (ASRCF) tracking algorithm. The adaptive spatial regularization term in the algorithm solves the boundary effect and also obtains the spatial regularization weight that establishes a connection with the target. Trackers that incorporate depth features have extremely slow tracking speed due to increased computational complexity. Moreover, these trackers update the target model by interpolation update method without evaluating the reliability of the sample. Therefore, we propose a high-confidence model update strategy to improve target tracking performance. manuscripts must be in English, also the table and figure texts, otherwise we cannot publish your paper. Please keep a second copy of your manuscript in your office.

3. THE PROPOSED METHOD

3.1. Baseline Tracker

The kernel correlation filter achieves high speed and accuracy in object tracking. The main objective of training the classifier $f(x_i)$ is to find a function $f(x_i)=w^T x_i$, that minimizes the squared error over samples x_i and their regression targets y_i. We define the circulant matrix $C(\mathrm{x}_i)$ is generated by cyclic shifts of the base sample x, i.e., $P(x_i)=[x_i, x_1, x_2, \cdots x_{i-1}], i=1,\cdots,N$, where P is a permutation matrix. The tracker uses the cyclic shifts to collect training samples x_i. The objective function can be formulated as follows:

$$\min_{w}\sum_{i}[w^T x_i - y_i]^2 + \lambda\|w\|^2 \tag{1}$$

Where λ is a regularization parameter, which can avoid overfitting of the classifier; f is a nonlinear filter, and w is a filter coefficient. By taking the partial derivative of the above formula, a closed solution is obtained

$$w=(X^T X+\lambda I)^{-1} X^T y \tag{2}$$

where the circulant matrix $X=[x, Px, \cdots, P_{n-1}x]^T$, and I denotes the identity matrix， the circulant matrix can avoid the inversion process according to the diagonalization property of the Fourier space. The kernel function $k(x,x')=\varphi^T(x)\varphi(x')$ is introduced to map samples in low-dimensional space to high-dimensional kernel space. In this way, the solution of w is transformed into the kernelized form of ridge regression

World Scientific Research Journal Volume 7 Issue 6, 2021
ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0028

$$\alpha = (K + \lambda I)^{-1} y \tag{3}$$

Where K represents the kernel space matrix, and a represents the filter coefficients. Using the property of diagonalization, the inversion operation is converted into frequency domain operation, and the discrete Fourier transform is performed on both sides of the formula to obtain

$$\hat{\alpha} = \frac{\hat{y}}{\hat{k}^{xx} + \lambda} \tag{4}$$

Where $\wedge$ represents the Fourier transform of the variable, and $\hat{k}^{xx}$ represents the self-kernel correlation of the sample x in the Fourier domain, which is the first row of k. Since the extracted features contain multiple channels, the Gaussian kernel function is used to convert a single channel to multiple channels. The calculation method is as follows

$$k^{xx'} = \exp(-\frac{1}{\sigma^2}(\| x \|^2 + \| x' \|^2 \text{-} 2F^{-1}(\hat{x}^* \text{ Z } \hat{x}^*))) \tag{5}$$

Where σ denotes the Gaussian kernel bandwidth. Only element dot product and discrete Fourier transform are performed, which greatly improves the calculation speed. The response calculation formula is as follows

$$\hat{f}(z) = \hat{k}^{xz} \text{ Z } \hat{\alpha} \tag{6}$$

Where k^{xz} represents the kernel correlation between the training sample and the test sample. The position corresponding to the maximum value of the response function is the position of the detection target.

In order to adapt to changes in the appearance of the target, linear interpolation is used to update the target sample feature x and filter coefficient a, the update method is as follows

$$\begin{aligned} \hat{x}_t &= (1-\eta)\hat{x}_{t-1} + \eta\hat{x} \\ \hat{\alpha}_t &= (1-\eta)\hat{\alpha}_{t-1} + \eta\hat{\alpha} \end{aligned} \tag{7}$$

Where t and t-1denote the current frame and the previous frame respectively, and η is the learning rate.

3.2. Multiple Feature Integration and Adaptive Scale

Some models are not effective in tracking deformed targets, but using color features to learn targets can well deal with the problems of deformation and motion blur of tracking targets. However, when the lighting conditions change, the color features are relatively weak. At this time, using the HOG feature can track the target of the light change well. We use the complementary characteristics between the features to fuse the 31-dimensional HOG feature and the 11-dimensional color naming feature. The feature fusion is shown in Figure 1. There are many multi-feature fusion strategies, such as fusion methods that directly add or multiply multiple features, and assign different fixed weights to each feature for weighted fusion. This

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0028

paper obtains the response value corresponding to each feature and then determines the weight coefficient corresponding to each feature

$$\lambda_i = \frac{Q_i}{\sum Q_i} \tag{8}$$

Where λ_i is the weight coefficient corresponding to each feature, and Q_i is the maximum response value of the i-th feature. In order to prevent overfitting of weight coefficients caused by simple linear weighting, a penalty factor is introduced

$$\begin{aligned} \lambda_i^{'} &= \frac{Q_i}{1/(20Q) + \sum Q_j} \\ \lambda^{''} &= \frac{\lambda_i^{'}}{\sum \lambda_i^{'}} \end{aligned} \tag{9}$$

Where $\lambda^{''}$ indicates the weighting factor of the i-th feature after correction, from which the fusion response of HOG and CN can be

$$f = \lambda^{''} f_{CN} + (1-\lambda^{''}) f_{HOG} \tag{10}$$

Obviously, when $\lambda^{''}$ it is 1, it means that only CN features are used; when $\lambda^{''}$ is 0, it means that only HOG features are used.

Scale variation is one of the severe challenges of target tracking. The scale pooling is used to estimate the scale variation of the target. We fix the template size as $S_T = (s_x, s_y)$, and define a scaling pool $S = \{t_1, t_2, \ldots, t_k\}$. Suppose the window size of the target in the original image space is s_t. For the current frame, we sample k sizes in $\{t_i s_t \mid t_i \in S\}$ to find a suitable target. The scale pool contains nine scales, with variable values ranging from 0.985 to 1.025. At the same time, bilinear interpolation is used to adjust the size of the image block, the final response is calculated by

$$\arg\max F^{-1} \hat{f}(z^{t_i}) \tag{11}$$

Where z^{t_i} is the sample patch of size $t_i s_t$, and its size has been adjusted to s_T.

3.3. Model Update Strategy

Due to the existence of interference information in the environment, and with the passage of time, the target appearance model will change greatly, and finally the tracking model will drift. In order to prevent the occurrence of model drift, it is necessary to adopt a reasonable update strategy for the target appearance template and parameters. Two evaluation indicators are introduced, the first is the maximum response score $F_{\max}$, it can be defined as

$$F_{\max} = \max F(s, y; w) \tag{12}$$

World Scientific Research Journal Volume 7 Issue 6, 2021
ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0028

Algorithm 1 Our proposed tracking algorithm
Input: Image I_t The dual space coefficient α ; Previous object position P_{t-1} and scale S_{t-1} ; The regenerated template in last frame $\hat{x}_{t-1}$; Output Estimated object position P_t and scale S_t ; The updated target template $\hat{x}_t$; The updated dual space coefficient α ;
Position and Scale estimation
1: Sample the new patch z based on size $t_i s_t$ and resize it to s_T 2: Extract different features and conduct feature fusion by Eq. 8 and 9-10 3: calculate the response with Eq.5 and 6. 4: Get final position and size according to Equation 11. 5: Get target template $\hat{x}_t$ based on new position P_t and size S_t , and calculate new α with Equation 4. 6: calculate the APCE and $F_{\max}$
Model updating
7: if $F_{\max}$ and APCE satisfy the update condition, then 8: update $\hat{x}_{t-1}$ and α with $\hat{x}_t$ and new a by Eq.7 else 9: target position and scale model are not updating. until end of video sequence.

The second is the average peak-to-correlation energy measure, which represents the stability of the current tracking situation and can better reflect the credibility of the tracking, it can be defined as

$$APCE = \frac{|F_{\max} - F_{\min}|^2}{mean(\sum (F_{w,h} - F_{\min})^2)} \tag{13}$$

Where $F_{\max}$, $F_{\min}$ and $F_{w,h}$ denote the maximum, minimum and the w-th row h-th column elements of F (s, y; w). Different from the previous method of updating model parameters through interpolation, we use APCE combined with the maximum response score to jointly decide the update method, the decision formula is as follows

$$K(t) = \begin{cases} true, if \text{APCE(t)} \geq \beta_1 APCE_{\text{mean}} \ \& \ \text{F}_{\max} \geq \beta_2 F_{\max\_mean} \\ false, \text{otherwise} \end{cases} \tag{14}$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0028

Where APCE(t) represents the APCE of the correlation response map of the t-th video frame, and $APCE_{mean}$ is the average APCE of the historical frame of the past 5 frames, and β is the set threshold parameter. When the APCE and the maximum response value of the current frame are higher than the average value of the previous 5 frames respectively, the current position and scale of the target are updated, and other conditions are not updated. The overall algorithm is summarized as Algorithm 1.

4. EXPERIMENTAL RESULTS

In this section, we first introduced the experimental setup. In order to evaluate the performance of the proposed tracker, we compared the proposed tracker with the most relevant 7 trackers in 51 video sequences of OTB-2013 and 100 video sequences of OTB-2015, they are KCF [9], DSST [15], CN [10], CSK [8], Struck [24], TLD [25], ALSA [26], respectively. Our target tracking algorithm is implemented in MATLAB2018 on a PC with Intel i7–10750 CPU (2.6 GHz) and 16 GB memory.

4.1. Experimental Setup

Dataset and evaluation criteria: The data sets OTB-2013 and OTB-2015 contain 51 and 100 video sequences, respectively, and are composed of different video sequences with 11 types of marker attributes, namely, Illumination Variation (IV), Scale Variation (SV), Occlusion (OCC), Deformation (DEF), Motion Blur (MB), Fast Motion (FM), In-Plane Rotation (IPR), Out-of-Plane Rotation (OPR), Out of View (OV), Background Clutter (BC), Low Resolution (LR). In order to evaluate the performance of various tracking algorithms, we use three indicators for measurement. The first is the center position error (CLE), which refers to the Euclidean distance between the target center predicted by the tracker and the manually marked target center. The smaller the value, the better the tracking effect. The second is distance accuracy, which means the relative number of frames when the center position error is less than a certain threshold. The larger the value, the better. The threshold is usually set to 20 pixels. The third type is the overlap rate accuracy, which represents the percentage of all video frames that have been successfully tracked. If the overlap rate of the tracking frame exceeds a certain threshold, the video frame is considered to be tracked successfully, and the threshold is usually set to 0.5. In addition, our results also use the precision plot and success plot under the one-pass evaluation standard to evaluate the overall performance of the tracker.

Table 1. Parameters of our tracker

Our tracker	Padding	η	σ	β_1	β_2
Parameters	1.5	0.01	0.1	0.45	0.6

Features and parameter settings: in this experiment, we use the default parameters of the KCF tracker, and use 31-dimensional HOG features with 5 pixels window sizes and 9 directions, and 11-dimensional Color Name features. In addition, the compromise parameter λ is set to 0.01 to balance the regression error and the regularization term. In practical applications, the appearance model is not very sensitive to λ. If the parameters satisfy that λ is less than 0.02 and greater than 0, almost the same tracking performance can be obtained. The padding of the search area in the correlation filter is set to 1.5, and the learning rate is set to 0.01. In order to achieve scale adaptation, 9 scales are used, and the scale pool is set to {0.985, 0.99, 0.995, 1.0, 1.005, 1.01, 1.015, 1.02, 1.025}, which will obtain better scale adaptation, but will increase computational complexity degree. Other parameter settings are shown in Table 1.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0028

4.2. Overall Performance

Figure 2 and Figure 3 show the accuracy and success rate maps obtained by the tracker in the OTB2013 and OTB2015 datasets, respectively. It can be seen that the distance accuracy of our proposed tracker is 75.7%, and the overlap rate is 70.1%, which is completely better than the other seven discriminative trackers and obtains good tracking performance. In terms of accuracy, our tracker has no obvious advantage over the performance of KCF tracker, but in terms of success rate, our tracker has a performance increase of 7.8% compared to KCF tracker.

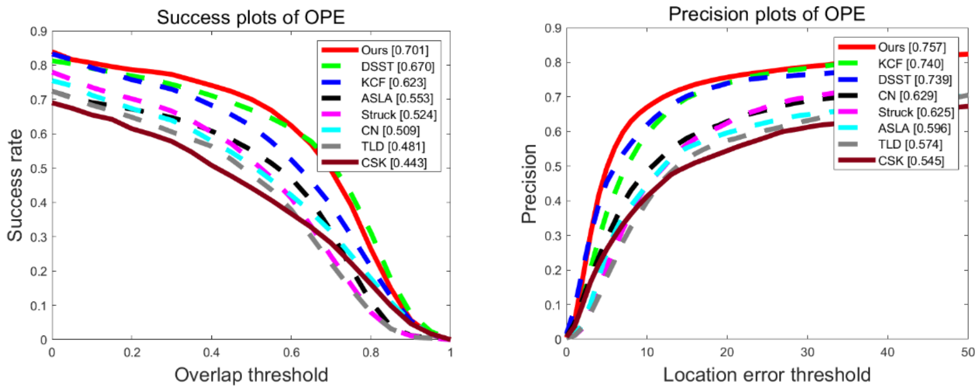

Figure 2. Success and precision plots of OPE comparison for our proposed tracker with 7 other popular trackers on the OTB2013 database

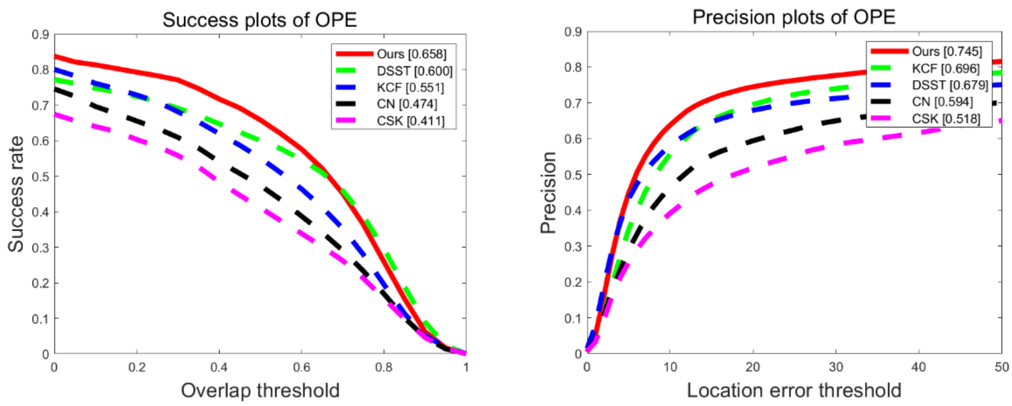

Figure 3. Success and precision plots of OPE comparison for our proposed tracker with 4 other popular trackers on the OTB2015 database

Due to the poor tracking performance of classic discriminant trackers such as Struck, TLD, ASLA, in OTB-2015 we focus on the tracking performance of more advanced correlation filter trackers. It can be seen from Figure 3 that our tracker ranks first with a range accuracy of 74.5% and an overlap rate of 65.8% among many trackers, achieving the best performance. Compared with the KCF tracker, our tracker is 4.9% higher in accuracy. From the perspective of the success rate, the performance is improved by 10.7%. It can be seen from the above analysis that our tracker is superior to other trackers in both data sets and achieves the best performance.

4.3. Attribute-based Performance

Table 2 shows the comparison results of the attribute-based success rate of our proposed tracker on the OTB2013 data set with the other seven trackers. We arrange the trackers from left to right according to the overall performance of the trackers. In each attribute, red

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0028

represents the best performance among all trackers, and green represents the second best performance.

It can be seen from the data in the table that our proposed method achieves the best performance among the 6 attributes, the second best performance among the other 4 attributes, and the third among the IPR (in-plane rotation) attributes. In the OCC and SV attributes, our tracker ranked first with a success rate of 73.0% and 70.6%. Compared with the KCF tracker, the performance increased by 11% and 17.7%, respectively. This shows that our tracker is blocking and There is stronger adaptability in the scale conversion scene.

Table 2. Attribute-based success rate comparison of our proposed tracker with the other 7 popular trackers on the OTB2013 datasets.

Attribute	CSK	TLD	CN	Struck	ASLA	KCF	DSST	Ours
LR	0.397	0.298	0.397	0.384	0.307	0.357	0.497	0.490
IPR	0.457	0.441	0.550	0.463	0.523	0.615	0.697	0.613
OPR	0.439	0.476	0.502	0.478	0.501	0.608	0.642	0.646
SV	0.352	0.442	0.421	0.437	0.594	0.479	0.640	0.656
OCC	0.404	0.473	0.479	0.494	0.496	0.619	0.645	0.730
DEF	0.370	0.455	0.512	0.489	0.513	0.671	0.630	0.706
BC	0.491	0.352	0.534	0.502	0.520	0.672	0.627	0.639
IV	0.388	0.413	0.449	0.458	0.520	0.582	0.681	0.595
MB	0.336	0.523	0.479	0.510	0.301	0.595	0.528	0.591
FM	0.380	0.505	0.436	0.543	0.292	0.557	0.503	0.599
OV	0.410	0.534	0.461	0.491	0.459	0.650	0.512	0.696

Table 3. Attribute-based precision rate comparison of our proposed tracker with the other 7 popular trackers on the OTB2013 datasets.

Attribute	CSK	TLD	CN	Struck	ASLA	KCF	DSST	Ours
LR	0.411	0.328	0.405	0.517	0.295	0.381	0.497	0.517
IPR	0.547	0.547	0.675	0.556	0.586	0.677	0.768	0.725
OPR	0.540	0.573	0.645	0.579	0.563	0.729	0.735	0.719
SV	0.503	0.565	0.598	0.610	0.622	0.679	0.738	0.752
OCC	0.500	0.520	0.619	0.578	0.512	0.749	0.706	0.785
DEF	0.481	0.517	0.575	0.542	0.536	0.728	0.730	0.668
BC	0.585	0.393	0.629	0.530	0.541	0.664	0.694	0.735
IV	0.481	0.517	0.575	0.542	0.536	0.638	0.730	0.728
MB	0.342	0.576	0.550	0.560	0.288	0.650	0.544	0.595
FM	0.381	0.537	0.480	0.541	0.277	0.602	0.513	0.628
OV	0.379	0.524	0.434	0.401	0.373	0.650	0.511	0.690

Table 3 shows the attribute-based accuracy comparison results of our proposed tracker on the OTB2013 data set and the other seven trackers. Similarly, we sort the trackers from left to right according to their overall performance, with red representing optimal performance and green representing suboptimal performance. As can be seen from the table, our tracker ranks first in the performance of 6 attributes such as LR, SV, OCC, BC, FM, OV, and the second best performance among the three attributes of IPR, MB, and IV. In OPR and DEF attributes, our tracker is 1% and 6% lower than KCF.

4.4. Qualitative Analysis

Figure 4 shows the qualitative comparison of 5 trackers under 5 challenging video sequences. The 5 trackers are Ours, KCF, DSST, CN, CSK, and the challenge video sequences are coke, couple, deer, fish, soccer_1. All in all, our tracker can show better tracking performance in these scenarios. In the couple sequence, our tracker shows better tracking performance when the target suffers deformation and fast movement. In the soccer sequence, when the target is occluded and deformed, some trackers lose track of the target after a period of time, but our tracker still maintains good tracking performance. It can be seen that our tracker has strong processing capabilities for occluded and deformed targets.

Ours DSST KCF CSK CN

Figure 4. Qualitative comparison results of our method and state-of-the-art trackers on the typically challenging sequences (from top to down are coke, couple, deer, fish, soccer_1, respectively).

5. CONCLUSION

In this paper, a high-confidence model update method for kernel correlation filter trackers is proposed to address the problem of unreliable model update strategy and model drift. To begin with, we perform linear weighted fusion on the extracted features, which effectively improves the overall tracking performance. Besides that, a reliable model update strategy combining APCE and the Maximum response score is proposed, which effectively avoid the occlusion problem in model update. In the OTB2013 and OTB2015 datasets, our tracker has better processing capabilities for occlusion and deformation. A large number of experimental results show that our proposed tracker is superior to multiple advanced trackers in terms of accuracy and success rate.

ACKNOWLEDGMENTS

This work was supported in part by the Natural Science Foundation of China under Grant 61801319, in part by Sichuan Science and Technology Program under Grant 2020JDJQ0061,

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0028

2019YJ0476 and 2020YFSY0027, in part by the Education Agency Project of Sichuan Province under Grant 18ZB0419, in part by the Sichuan University of Science and Engineering Talent Introduction Project under Grant 2020RC33, in part by the Major Frontier Project of Sichuan Science and Technology Plan under Grant 2018JY0512, in part by Outstanding Young Science and Technology Talent Project of Sichuan Provincial Department of Science and Technology under Grant 2020JDJQ0075, in part by Graduate/Undergraduate Innovation Foundation of Sichuan University of Science and Engineering under Grant y2020017.

REFERENCES

[1] Zhang Shunli, Lu Wei, Xing Weiwei, Zhang Li. Learning Scale-Adaptive Tight Correlation Filter for Object Tracking[J]. IEEE transactions on cybernetics, 2020, 50(1): 1-14.

[2] X. Zhang, G. S. Xia, Q. Lu. Visual object tracking by correlation filters and online learning[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 140: 77-89.

[3] Xingchen Zhang, Ping Ye, Henry Leung, Ke Gong. Object fusion tracking based on visible and infrared images: A comprehensive review[J]. Information Fusion, Volume 63, 2020: 166-187.

[4] Xueming Zhai, Jilei Jia. Research on Object Tracking and Target Recognition Based on Kalman Filter and YOLOV3[J]. International Core Journal of Engineering, 2020, 6(11): 1-12.

[5] Jian Chen, Yanming Lin, Detian Huang, Jian Zhang. Robust tracking algorithm for infrared target via correlation filter and particle filter[J]. Infrared Physics and Technology, 2020, 111: 1-13.

[6] Irene Anindaputri Iswanto, Tan William Choa, Bin Li. Object Tracking Based on Meanshift and Particle-Kalman Filter Algorithm with Multi Features[J]. Procedia Computer Science, 2019, 157: 521-529.

[7] Henriques J F, CASEIRO R, MARTINS P, et al. High-speed tracking with kernelized correlation filters[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015: 583-596.

[8] Wang M, Liu Y, Huang Z. Large Margin Object Tracking with Circulant Feature Maps[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017: 4021-4029.

[9] Martin, Danelljan, Gustav, et al. Discriminative Scale Space Tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(8):1561-1575.

[10] Danelljan M, Shahbaz Khan F, Felsberg M. Adaptive color attributes for real-time visual tracking[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 1090-1097.

[11] Henriques J F, CASEIRO R, MARTINS P, et al. Exploiting the circulant structure of tracking by detection with kernels[C]. European conference on computer vision, 2012: 702-715.

[12] S. Hare, A. Saffari, and P. H. S. Torr. Struck: Structured output tracking with kernels[C]. Proc. IEEE, 2011: 263-270.

[13] Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection[J]. IEEE Transactions on Software Engineering, 2011, 34(7):1409-1422.

[14] X. Jia, H. Lu, and M.-H. Yang. Visual tracking via adaptive structural local sparse appearance model[C]. Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) IEEE, 2012: 1822-1829.