ISSN: 2414-1895 DOI: 10.6919/ICJE.202204_8(4).0068

A Survey of Target Tracking Algorithms based on Correlation Filtering

Lin Zhang

School of Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science and Engineering, Yibin 644000, China

Abstract

Visual object tracking plays an important role in many computer vision applications. Discriminative target tracking method based on correlation filtering(CF) theory has become a research hotspot in the field of target tracking due to its robustness and efficiency. Discriminative correlation filter greatly improves tracking robustness by introducing feature representation, nonlinear kernel, scale estimation, spatio-temporal regularization and continuous convolution. This paper first introduces the basic CF theory and the basic framework of target tracking. Secondly, CF-based trackers are summarized by category. Thirdly, using the target tracking benchmark database (OTB-2013) video sequence to conduct algorithm comparison experiments, analyze and compare the performance of 9 typical different CF trackers in recent years. Finally, according to the current research status, it points out the possible future development trend of CF. Although target tracking based on correlation filter has been widely used in the field of tracking and has made some progress, target tracking is still a huge challenge due to the impact of complex scenes and dramatic changes in the appearance of the target itself. It is of great significance for the development of target tracking to study the correlation filtering tracking algorithm with high efficiency and robustness.

Keywords

Correlation Filtering; Target Tracking; Feature Representation; Scale Estimation.

1. Introduction

Visual object tracking is a basic research field in the field of computer vision and video processing. It is a method of tracking specific objects from a video or frame sequence [1]. Recently, with the rapid development of artificial intelligence technology, the problem of moving target tracking has received more and more attention. According to its working principle, the target tracking algorithm can be divid-ed into two types: generative method and discriminant method [2]. CF-based tracker is a discriminative tracking method, which takes the target model and background information into account at the same time, and extracts the target model by comparing the difference between the target model and the back-ground information, thereby obtaining the target position in the current frame.

Early target tracking focused on generative model tracking algorithm [3] .The generative method is mainly to model a given target area in the initial frame, and search for the most similar part of the mod-el in subseq quent frames to be the predicted target position, such as kalman filter [4], particle filter [5] and mean-shift [6] and other algorithms. The tracking accuracy of this type of method is low, because the background information is not taken into account, it is easy to fail to track under the interference of target occlusion, lighting changes, and motion blur. And the tracking algorithm has a slow execution speed (about 10 frames per second) and cannot meet the real-time requirements. Discriminative method is to treat the target tracking problem as a target detection task in each frame,

ISSN: 2414-1895 DOI: 10.6919/ICJE.202204_8(4).0068

and use the image characteristics of the tracking target to train a classifier. The target area in the image is used as a positive sample, and the background area is used as a negative sample. In the subsequent frames, the trained classifier is used to find the optimal solution, and the tracking result in each frame is continuously used to update the classifier during the tracking process.The typical discriminant method is CF-based trackers, which has faster tracking speed and better tracking results, and can meet real-time requirements. There are three reasons for the popularity of CF. Firstly, it is very efficient to calculate the spatial correlation in the Fourier domain as the element product, so the tracking algorithm based on the correlation filter achieves a high tracking speed. Secondly, the CF naturally takes the surrounding contextual information into account, thus providing a stronger discrimination capability than the appearance model based only on the target object. Thirdly, the learning CF can be expressed as a ridge regression problem [7], in which the cyclically shifted input features are returned to soft labels, which is different from the existing tracking methods through detection [8], [9]. In this method, binary samples are drawn around the estimated position for incremental learning of the classifier to distinguish the target object from the background. Therefore, CF-based trackers will not suffer from the ambiguity of assigning positive and negative labels to spatially correlated samples. 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.

The paper is organized as follows. In Section 2, we introduce the principle of CF trackers. Section 3 discusses the development and classification of correlation filter tracking algorithms. The state-of-the-art target tracking algorithm performs performance evaluation in Section 4. We conclude the paper in Section 5.

2. Correlation Filtering Theory

The concept of correlation filter first appeared in the field of communication, mainly used to measure the degree of correlation between signals and signals. The greater the degree of correlation between the two signals, the greater the similarity between the two signals. In view of this idea, the basic idea based on CF is to train an optimal filter template to maximize its response in the target area, which can be expressed in the time domain as:

$$g = x \otimes h \tag{1}$$

where x denotes the input data, which can be image features or original pixels, h denotes the correlation filter, and y is the correlation output, denotes convolution operation. Because convolution calculation is very time-consuming, in order to speed up the calculation, Fourier transform is introduced, which can be converted to:

$$G = X \odot H^* \tag{2}$$

Where X, H, G are the Fourier transform of x, h, g respectively,$\odot$denotes element-wise, H* is the conjugate representation of H. Through the above formula we can obtain:

$$H^* = \frac{G}{X} \tag{3}$$

H* is the Fourier transform of the filter template obtained by training. Assuming that y is the desired output, for the new target appearance z, the correlation filter h needs to satisfy:

ISSN: 2414-1895 DOI: 10.6919/ICJE.202204_8(4).0068

$$y = (H^* \odot Z) \tag{4}$$

where y is the expected output in the time domain, $F^{-1}(\cdot)$ represents the inverse Fourier transform. FFT effectively reduces the computational cost. For a picture of n*n size, the complexity of the circular convolution operation is $O(n^4)$,while the FFT only needs $O(n^2 \log n)$. So far, formulas (1) to (4) have built a basic bridge for the application of correlation filter in the field of target tracking.

3. The Development of Correlation Filter Tracking Algorithm

3.1 Trackers based on Feature Selection

The first to introduce correlation filter theory in the field of target tracking is the MOSSE[10] proposed by Bolme et al. This tracker uses grayscale features and has a fast tracking speed of up to 669FPS. However, the representation ability of grayscale features is not enough to handle situations where the background is complex or the target and the background color are similar. After that, Henriques et al. [11] improved the MOSSE algorithm by introducing a circulant matrix and kernel function to improve the accuracy of the algorithm while achieving high-speed tracking. KCF [12] algorithm extended the single-channel grayscale 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 captured by the HOG feature. Danelljan et al. [13] 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 are performance bottlenecks in tracking accuracy and tracking speed when using individual features. Therefore, multi-feature fusion has become a research hotspot in target tracking. Bertinetto et al. [14] proposed Staple tracker, which combines color features and HOG features for real-time tracking. SAMF[15] simultaneously fuses the original image gray information, color attributes and HOG features to improve the robustness of the tracker in complex environments. Liang et al.[16] introduces a new tracking method via extracting and evaluating multi-features for both target region and its adjacent surroundings.

In recent years, many correlated filter trackers based on deep convolution features have developed rapidly. The HCF proposed by Ma et al. [17] uses the trained VGG-19 as a feature extractor to model the extracted specific three-layer convolutional features instead of the original HOG features, which effectively improves the target tracking performance. Danelljan et al. [18] proposed the DeepSRDCF algorithm, which replaced the hand-crafted features with convolution features on the basis of SRDCF, a single convolution layer is used for modeling, and the output depth features are used for target tracking, which can better distinguish the target from the background. C-COT [19] 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. ECO [20] realized the fusion of traditional manual features and convolutional features. Through factorization operations, the feature dimensions of HOG, CN and CNN were reduced to varying degrees, reducing the number of parameters in the model. Bertinetto [21] proposed that the Siamese-FC framework completes the tracking task by measuring the similarity between the target area and the candidate area. Valmadre [22] introduced the CF on the basis of Siamese-FC and proposed the CFNet algorithm, which realized end-to-end learning through back propagation. HC-Siam [23] use the convolutional features of each layer to compare the correlation, and identify the location of the tracking object according to the maximum correlation.

The CF-based trackers has experienced the transformation from single feature to multi-feature fusion, from artificial feature to deep feature. How to match the appropriate features for the tracking task and which method to select for feature fusion is the key to improving the tracking performance.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202204_8(4).0068

3.2 Scale Adaptive Trackers

Early trackers could not track targets adaptively, and easily introduced background interference and edge information loss. There are three ways to improve this problem: the method based on scaling pool, part/patch model and keypoints.

The scaling pool method is to scale the original target object to different scales. Danelljan et al. [24] proposed DSST, which divided the tracking task into translation estimation and scale estimation, and used the position filter and scale filter to perform target positioning and scale estimation respectively, and the scale corresponding to the maximum response was the optimal scale. fDSST [25] takes into account the problem of computational complexity, and uses dimensionality reduction operations and QR decomposition to reduce the amount of calculation. As an extension, Lu et al. [26] proposed a robust scale and rotation estimation method based on the the kernelized correlation filter and Fourier-Mellin transform.

The block method is to divide the target into several small blocks and alculate the distance between the blocks in the current frame, and estimate the scale change of the target by judging whether the distance between the blocks meets the set threshold. Liu et al. [27] proposed RPAC on the basis of KCF, which decomposes the target into partial targets, and estimates the change of the target scale by calculating the change in the maximum response score in each response map. RPT [28] estimates the change of target scale by recording the relative position changes of different sub-blocks, and uses Gaussian filter to smooth the scale. DPCF [29] uses a global filter and the coupling of multiple component filters to co-process local occlusion and scale changes.

3.3 Trackers for Solving Boundary Effects

Correlation filter trackers produce unwanted boundary effects due to the periodic assumption of training samples, which seriously affects tracking performance. [30]-[32] introduced regularization terms to eliminate the influence of boundary effects. Danelljan et al. proposed the SRDCF[33], which introduces a spatial regularization term into the DCF framework, and determines the penalty coefficient of the filter according to the spatial location, which weakens the interference of background information. In addition, Galoogahi et al. [34] used a larger training area and a smaller filter size to learn CF from the tailored samples, which significantly increased the number of samples that are not contaminated by boundary effects. Alan et al. [35] proposed CSR-DCF based on channel reliability and spatial confidence, which reduces unnecessary boundary effects. He et al. [36] proposed an online adaptive learning method for spatio-temporal regularization, introducing spatial local response graph changes as spatial regularization, so that DCF focuses on the learning of the trusted part of the object. Hamed et al. [37] proposed BACF to increase the number of samples and narrow the search area by expanding the cyclic sampling area, reducing the interference of background information. However, the spatial regular weight does not establish a connection with the target. In tracking scenarios such as deformation and rotation, the algorithm may not be able to obtain reliable filter penalty coefficients. Dai et al. [38] proposed ASRCF tracker , the adaptive spatial regularization term solves the boundary effect and also obtains the spatial regularization weight that establishes a connection with the target. Li et al. introduced STRCF model [39], in the tracking process, only the information of the previous frame is used, and the target can be successfully tracked in the presence of occlusion, and at the same time, it can well adapt to larger appearance changes. Since then, a large number of improved versions of the STRCF trackers have appeared. Elayaperumal et al. [40] proposed a novel sparse context-aware spatio-temporal CFs, which uses context information to accurately locate the target, and introduces multi-scale sparse spatio-temporal constraints in the target model. Zhao et al. [41] used an improved cyclic shift operation to collect training samples, which is more robust to occlusion and rapid movement.

4. Experiments and Discussions

In this section,we introduce the performance evaluation method of the target tracking algorithm, then compares and analyzes the target tracking algorithm on the data set OTB-50.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202204_8(4).0068

4.1 Video Dataset

The test video sequences in this article are all from the target tracking benchmark database OTB-50, which composed of 50 different video sequences with 11 types of marker attributes, namely, illumination viriation (IV), scale viriation (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). The attribute marking situation of the video sequence is shown in Table 1.

Table 1. Tag attribute of the video sequences on OTB-50

Sequence	attribute	Sequence	attribute
Basketball	IV OCC DEF OPR BC	Human4	IV SV OCC DEF
Biker	SV OCC MB FM OPR OV	Human6	SV OCC DEF FM OPR OV
Bird1	FM OV DEF	Human9	IV SV DEF MB FM
BlurBody	SV DEF MB FM IPR	Ironman	IV SV OCC MB FM IPR OPR
BlurCar2	SV MB FM	Jump	SV OCC DEF MB FM IPR OPR
BlurFace	MB FM IPR	Jumping	MB FM
BlurOwL	SV MB FM IPR	Liquor	IV SV OCC BM FM OPR OV BC
Bolt	OCC DEF IPR OPR	Matrix	IV SV OCC FM IPR OPR BC
Box	IV SV OCC MB IPR OPR	MotorRolling	IV SV MB FM IPR BC LR
Car1	IV SV MB FM BC LR	Panda	SV OCC DEF IPR OPR OV LR
Car4	IV SV	RedTeam	SV OCC IPR OPR LR
CarDark	IV BC	Shaking	IV SV IPR OPR BC
Carscale	SV OCC FM IPR OPR	Singer2	IV DEF IPR OPR BC
ClifBar	SV OCC MB FM IPR OV	Skating1	IV SV OCC DEF OPR BC
Couple	SV DEF FM OPR BC	Skating2-1	SV OCC DEF FM OPR
Crowds	IV DEF BC	Skating2-2	SV OCC DEF FM OPR
David	IV SV OCC DEF MB JPR OPR	Skiing	IV SV DEF IPR OPR
Deer	MB FM IPR BC LR	Soccer	IV SV OCC MB FM IPR OPR
Diving	SV DEF IPR	Surfer	SV FM IPR OPR LR
DragonBaby	SV OCC MB FM IPR OPR OV	Sylvester	IV IPR OPR
Dudek	SV OCC DEF FM IPR OPR OV	Tiger2	IV OCC DEF MB FM IPR OPR OV
Football	OCC IPR OPR BC	Trellis	IV SV IPR OPR BC
Freeman4	IV SV OCC DEF	Walking	SV OCC DEF
Girl	SV OCC IPR OPR	Walking2	SV OCC LR
Human3	SV OCC DEF OPR BC	woman	IV SV OCC DEF MB FM OPR

ISSN: 2414-1895 DOI: 10.6919/ICJE.202204_8(4).0068

4.2 Evaluation Criterion

In order to evaluate the performance of various tracking algorithms, we used three indicators for measurement. The first is accuracy, which represents the relative number of frames where the center position error (CLE) in the video sequence is less than a certain threshold. The larger the value, the better. The threshold is usually set to 20. The center position error refers to the Euclidean distance between the target center predicted by the tracker and the manually labeled target center. The smaller the value, the better the tracking effect. The second is the success rate, which represents the percentage of all video frames that are successfully tracked. If the tracking frame overlap rate exceeds a certain threshold, the video frame is considered to be successfully tracked, and the threshold is usually set to 0.5. The overlap rate is the degree of overlap between the tracking frame of the indicator note and the predicted tracking frame. The calculation formula is as follows:

$$VOR = \frac{Area(B_T \cap B_R)}{Area(B_T \cup B_R)} \tag{5}$$

where BT and BR respectively represent the predicted target frame and the labeled target frame, and operators U and ∩ represent overlapping area and total coverage area respectively. The third is the speed of the tracking algorithm, which is mainly used to characterize the real-time performance of the algorithm through frame per second (FPS). Normally, the tracking algorithm can meet the real-time tracking requirements when the tracking algorithm reaches 25FPS.

4.3 Experiment and Discussion

There are nine comparison algorithms used in the experiment, namely KCF, CSK, SAMF, fDSST, SRDCF, BACF, STRCF, ECO, ARCF. Table 2 shows the comparison results of these 9 representative trackers. It can be seen from the table that the HOG feature is the most used feature, and most of the improved methods use the original image grayscale, HOG, color feature and deep feature for fusion. Classical algorithms (such as KCF, CSK) use a single feature to characterize the target appearance model, and do not consider target scale changes. Since then, most of the improved algorithms not only use multi-feature fusion to describe the target, but also introduce the target scale adaptive strategy to further improve the target tracking performance.

Table 2. Details of the tracking algorithm

CF tracker	feature	scale
CSK	Raw Pixel	NO
KCF	HOG	NO
SAMF	Raw Pixel +HOG+CN	YES
SRDCF	Raw Pixel +HOG+CN	YES
fDSST	Raw Pixel +HOG	YES
ECO	HOG+CN+CNN	YES
BACF	HOG	YES
STRCF	HOG+CN+CNN	YES
ARCF	HOG+CN+CNN	YES

4.3.1 Overall Performance Analysis

To evaluate the performance of different tracking algorithms, we compare nine representative advanced trackers under the OPE evaluation standard. Figure 1 shows the comparison results of the distance accuracy and success rate of the nine tracking algorithms on the OTB-2013 dataset.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202204_8(4).0068

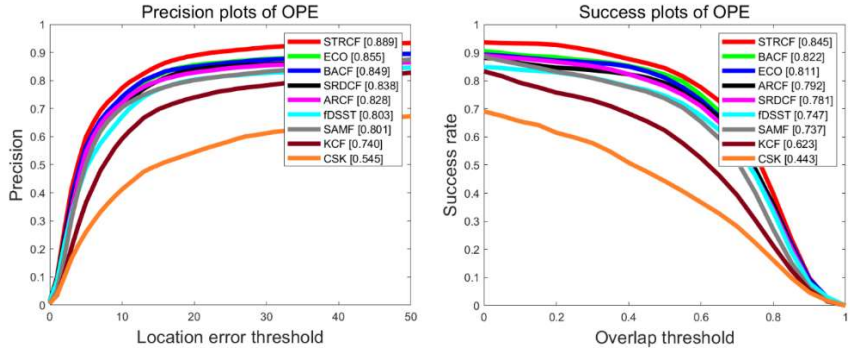

Figure 1. Comparison of distance accuracy and success rate on OTB-2013 dataset

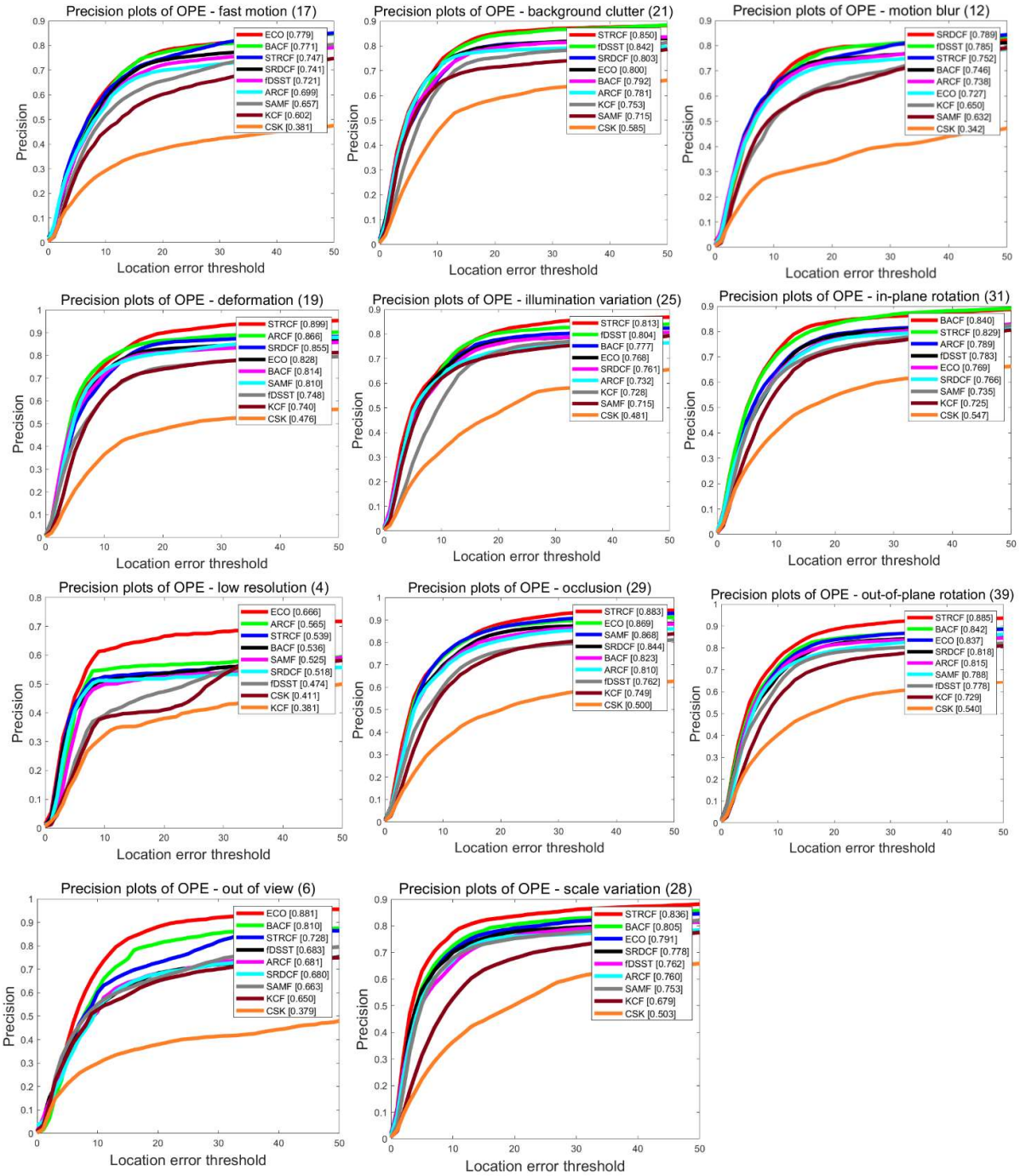

Figure 2. Comparison of distance accuracy over 11 challenge attributes

ISSN: 2414-1895 DOI: 10.6919/ICJE.202204_8(4).0068

Among the trackers in the literature, STRCF achieve the best results with 81.3% precision and 61.3% success rate. ECO and BACF ranked second with 0.855 accuracy and 0.822 success rate respectively. Regarding the problem that the periodic hypothesis of the discriminant correlation filter leads to unnecessary boundary effects, since the STRCF tracker integrates HOG, Color Naming and CNN features, and introduces a temporal regularization term based on the spatial penalty model, compared with the SRDCF algorithm, the distance accuracy is improved by 5.1%, and the overlap success rate is increased by 6.4%. BACF also processed the boundary effect, expanded the search area and used real negative samples, and achieved better perfor mance than SRDCF. It is clear that the sample quality has a great influence on the performance of the tracker.

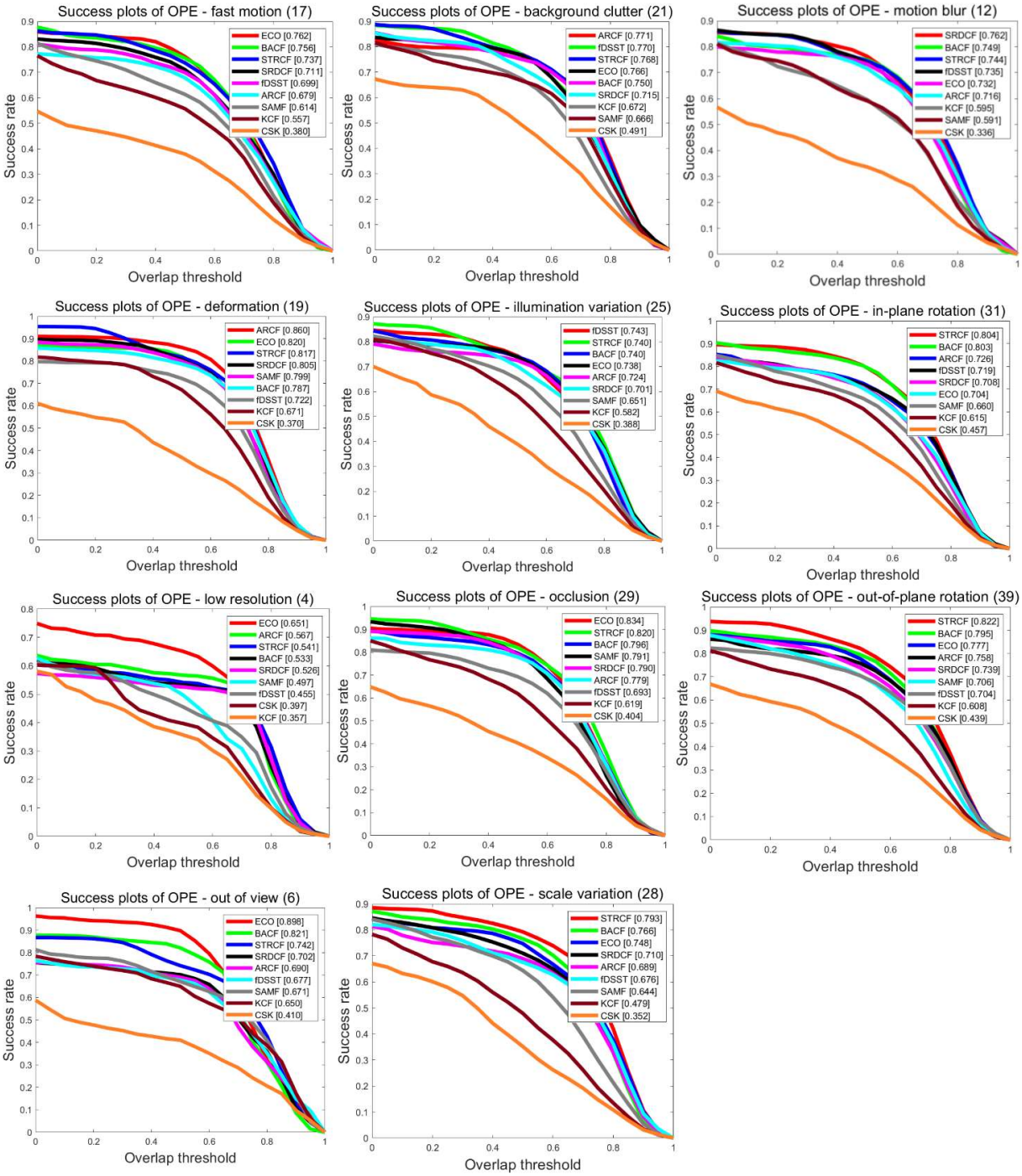

Figure 3. Comparison of success rate over 11 challenge attributes

ISSN: 2414-1895 DOI: 10.6919/ICJE.202204_8(4).0068

Classical tracker KCF accuracy rate is 0.74, success rate is 0.623, fDSST introduces scale transformation on the basis of KCF, the accuracy rate and success rate are increased by 6.3% and 12.4% respectively. The speeds of CSK and KCF algorithms in the classic algorithms are all above 100FPS, and the subsequent various CF have effectively improved the algorithm performance from different optimization perspectives. However, it affects the computing efficiency to a certain extent, making most trackers improve tracking performance at the expense of tracking speed.

4.3.2 Attribute-based Performance Evaluation

We uses 11 challenge attributes on the OTB-2013 dataset to evaluate the performance of the trackers. Figure 2 and Figure 3 shows the distance accuracy and success overlap rate of the trackers under each attribute. In the distance accuracy comparison, ECO obtains the best performance in the challenges of fast motion and low resolution. In the case of light changes, STRCF showed the best performance, fDSST and BACF were sub-optimal. For the scale variation scenarios discussed in this article, STRCF, BACF and ECO all show strong advantages. Compared with KCF and other trackers with scale viriation, the performance is significantly improved. In the occlusion challenge, the trackers usually use the interference background as the target information when observing the target, which leads to tracking drift. STRCF and ECO trackers have shown excellent performance. By measuring the difference of correlation filters between adjacent frames, it promotes the adaptability of the rapid changes of the learning target and improves the robustness when the target is partially occluded. Out of view is one of the most severe challenges for most trackers, the ECO using deep features ranks first with an accuracy rate of 0.898 and a success rate of 0.881. In rotation and deformation, STRCF and BACF performed best, and ARCF had the second best performance in deformation. It can be seen from the analysis that it is obvious that the best performance under each challenge attribute is the regularized CF.

5. Conclusions and Prospects

In this paper, we discusses the current research status of correlation filter target tracking, and compare the current advanced target tracking algorithms in OTB-50 video sequences. Obviously, it can be concluded from the experimental results that CF perform well under various attributes. However, due to the complexity of the actual scene and the uncertainty of the appearance of the target, the existing algorithms still cannot meet the demand. It is of great significance to study efficient and robust correlation filtering algorithms.

Future research can focus on balancing the relationship between tracking performance and real-time performance. In addition, searching for a suitable occlusion evaluation mechanism to improve the robustness of the target occlusion environment to achieve long-term tracking effect is also a hot issue of current research.The correlation filters does not evaluate the reliability of the samples, when using damaged samples to train filter parameters, it will cause deviations in subsequent target positioning. Therefore, it is particularly important to develop a reliable model update mechanism.

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 and 2020YFSY0027, 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 Graduate Innovation Fund of Sichuan University of Science and Engineering under Grant y2020018, in part by Undergraduate Student Innovation Fund of Sichuan University of Science and Engineering under Grant cx2020161.

References

[1] Yuan G L, Xue M G. Visual tracking based on sparse dense structure representation and online robust dictionary learning[J]. Journal of Electronics & Information Technology, 2015, 37(3): 536-544.

[2] Zhang K, Liu Q, Wu Y, et al. Robust Visual Tracking via Convolutional Networks[J]. IEEE Transactions on Image Processing. 2015, 25(4): 1779-1792.

[3] Ross D A, Lim J, Lin R S, et al. Incremental learning for robust visual tracking[J]. International Journal of Computer Vision, 2008, 7(1-3): 125-141.

[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): 905-911.

[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: 103516.

[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] Yin Rong,Liu Yong, Wang Weiping, Meng Dan. Sketch Kernel Ridge Regression Using Circulant Matrix: Algorithm and Theory[J]. IEEE transactions on neural networks and learning systems, 2019: 3512-3524.

[8] Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(7): 1409-1422.

[9] Hare Sam, Golodetz Stuart, Saffari Amir, Vineet Vibhav, Cheng Ming-Ming. Struck: Structured Output Tracking with Kernels[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 38(10): 2096-2109.

[10] Bolme D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters[C]. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2010: 2544-2550.

[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] 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, 37(3): 583-596.

[13] 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.

[14] BERTINETTO L, VALMADRE J, GOLODETZ S, et al. Staple: Complementary learners for real-time tracking[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 1401-1409.

[15] LI Y, ZHU J. A scale adaptive kernel correlation filter tracker with feature integration[C]. European conference on computer vision. Springer, Cham, 2014: 254-265.

[16] Yun Liang, Jian Zhang, Mei-hua Wang, Chen Lin, Jun Xiao. Multi-features guided robust visual tracking[J]. Multimedia Tools and Applications, 2021, 80(11): 16367-16395.

[17] Ma C, Huang J B, Yang X, et al. Hierarchical convolutional features for visual tracking[C]. Proceedings of the IEEE international conference on computer vision. 2015: 3074-3082.

[18] Danelljan M, Hager G, Khan F S, et al. Convolutional features for correlation filter based visual tracking[C]. Proceedings of IEEE International Conference on Computer Vision Workshop. Santiago, Chile: IEEE, 2015: 621-629.

[19] Martin Danelljan, Andreas Robinson, Fahad Khan, Michael Felsberg. Beyond Correlation Filters: Learning Continuous Convolution Operators for Visual Tracking[C]. European conference on computer vision, 2016: 472-488.

[20] Danelljan M, Bhat G, et al. Efficient convolution operators for tracking[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6638-6646.

[21] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional siamese networks for object tracking[C]. European conference on computer vision. Springer, Cham, 2016: 850-865.

[22] Valmadre J, Bertinetto L, Henriques J, et al. End-to-end representation learning for correlation filter based tracking[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2805-2813.

[23] Yu Meng, Zaixu Deng, Kun Zhao, Yan Xu, Hao Liu. Hierarchical correlation siamese network for real-time object tracking[J]. Applied Intelligence, 2021, 51(6): 3202-3211.

[24] Danelljan M, H¨ager G, Khan F S, Felsberg M. Accurate scale estimation for robust visual tracking. In: Proceedings British Machine Vision Conference. London, England: BMVA Press, 2014. 65.1-65.11.

[25] Danelljan M, Hager G, Khan F S, et al. Discriminative scale space tracking[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(8): 1561-1575.

[26] Huimin Lu, Dan Xiong, Junhao Xiao, Zhiqiang Zheng. Robust long-term object tracking with adaptive scale and rotation estimation[J]. International Journal of Advanced Robotic Systems, 2020, 17(2): 1-14.

[27] Liu Ting, Gang Wang, Qingxiong Yang. Real-time part-based visual tracking via adaptive correlation filters[C]. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2015: 4902-4912.

[28] I Y, ZHU J, HOI S C H. Reliable patch trackers: Robust visual tracking by exploiting reliable patches[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 353-361.

[29] Akin O, Erdem E, Erdem A, et al. Deformable part-based tracking by coupled global and local correlation filters[J]. Journal of Visual Communication and Image Representation, 2016, 38: 763-774.

[30] Chong Sun, Dong Wang, Huchuan Lu, Ming-Hsuan Yang. Learning Spatial-Aware Regressions for Visual Tracking[C]. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2018: 8962-8970.

[31] Mengdan Zhang, Qiang Wang, Junliang Xing, Jin Gao, Peixi Peng, Weiming Hu. Visual Tracking via Spatially Aligned Correlation Filters Network[C]. European conference on computer vision, 2018: 469-485.

[32] Jongwon Choi, Hyung Jin Chang, Tobias Fischer, Sangdoo Yun, Kyuewang Lee, Jiyeoup Jeong, Yiannis Demiris, Jin Young Choi. Context-aware Deep Feature Compression for High-speed Visual Tracking[C]. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2018: 479-488.

[33] Danelljan M, Hager G, Shahbza K F, et al. Learning spatially regularized correlation filters for visual tracking[C]. Proceedings of the IEEE international conference on computer vision. 2015: 4310-4318.

[34] Hamed Kiani Galoogahi, Terence Sim, Simon Lucey. Correlation Filters with Limited Boundaries[C]. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2015: 4630-4638.

[35] Lukezic A, Vojir T, Cehovin Zajc L, et al. Discriminative correlation filter with channel and spatial reliability[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6309-6318.

[36] Zhiqun He, Yingruo Fan, Junfei Zhuang, Yuan Dong, HongLiang Bai. Correlation Filters With Weighted Convolution Responses[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017 1992-2000.

[37] Galoogahi H K, Fagg A, Lucey S. Learning background-aware correlation filters for visual tracking[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 1135-1143.

[38] Kenan Dai, Dong Wang, Huchuan Lu, Chong Sun, Jianhua Li. Visual Tracking via Adaptive Spatially Regularized Correlation Filters[C]. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2019: 4670-4679.

[39] Feng Li, Cheng Tian, Wangmeng Zuo, Lei Zhang, Ming-Hsuan Yang. Learning Spatial-Temporal Regularized Correlation Filters for Visual Tracking[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4904-4913.

[40] Elayaperumal Dinesh, Joo Young Hoon. Visual Object Tracking Using Sparse Context-aware Spatio-temporal Correlation Filter[J]. Journal of Visual Communication and Image Representation, 2020, 70: 102820.

[41] Di Yuan, Xiu Shu, Zhenyu He. Learning temporal regularized correlation filters for high performance online visual object tracking[J]. Journal of Visual Communication and Image Representation, 2020, 72: 102882.