IEICE TRANS. INF. & SYST., VOL.E106–D, NO.5 MAY 2023

PAPER Special Section on Deep Learning Technologies: Architecture, Optimization, Techniques, and Applications

An Improved Real-Time Object Tracking Algorithm Based on Deep Learning Features

Xianyu WANG[†,††], **Cong LI**[††], **Heyi LI**[†††], **Rui ZHANG**[††††], **Zhifeng LIANG**[††††], *and* **Hai WANG**[†††a)], ***Nonmembers***

SUMMARY Visual object tracking is always a challenging task in computer vision. During the tracking, the shape and appearance of the target may change greatly, and because of the lack of sufficient training samples, most of the online learning tracking algorithms will have performance bottlenecks. In this paper, an improved real-time algorithm based on deep learning features is proposed, which combines multi-feature fusion, multi-scale estimation, adaptive updating of target model and re-detection after target loss. The effectiveness and advantages of the proposed algorithm are proved by a large number of comparative experiments with other excellent algorithms on large benchmark datasets.

key words: *object tracking, feature fusion, deep learning, model update, re-detection*

1. Introduction

Object tracking has very important applications and broad development prospects in the field of computer vision, such as: automatic driving, military, video surveillance and so on [1]. In recent years, with the advancement of computer vision, many excellent target tracking algorithms have been proposed, and many of them have been successfully applied in our real life. However, due to the complexity of the target environment, the efficiency and accuracy of the target tracking algorithm has been affected.

In this paper, we mainly study the problems caused by background interference, occlusion and out of view in the task of target tracking in complex scenes, and propose an adaptive model update and lost re-detection correlation tracking (MURCT) algorithm. In the feature extraction [2] stage, the HOG and CN manual features and the depth features extracted by the convolutional neural network are adaptively fused to achieve the effect of complementary learning. Meanwhile, we use the EdgeBox algorithm which is used to extract candidate target boxes in target detection, combined with the scale pyramid, and selects the optimal solution as the result of scale estimation to improve the fitting ability of the target. The adaptive model update module detects whether the target is occluded or lost according to a discrimination mechanism, and stops updating the model when the target is occluded, so as to prevent model pollution. The lost re-detection module is used for activating the re-detection module to retrieve the target position when the target is judged to be lost. And, the proposed algorithm was tested and compared on the OTB dataset [3], [4] and the VOT2016 dataset [5], which proved the effectiveness of the algorithm.

The remaining sections of the paper are organized as follows. Some previous attempts to solve the object tracking problem are presented in Sect. 2. Section 3 describes the process of our proposed MURCT tracking method. Section 4 shows the experimental evaluation results on the test dataset, and finally conclusions are presented in Sect. 5.

2. Related Work

The concept of object tracking was first proposed by Wax et al. [6] in 1955. In the following decades, many excellent object tracking algorithms were proposed by researchers from various countries. In recent years, tracking algorithms based on correlation filters and deep learning [7], [8] have gradually become the most prominent research directions in this field.

Bertinetto et al. [9] proposed a Fully-Convolutional Siamese Networks (SiamFC) for object tracking, in which two fully convolutional networks with the same structure are used to extract the features of the target region. Li et al. [10] introduced the Region Proposal Network (RPN) in object detection into SiamFC, and divided object tracking into two subtasks: regression and classification. In 2010, Bolme et al. [11] introduced correlation filters algorithms to the field of object tracking for the first time, and proposed the Minimum Output Sum of Squared Error (MOSSE) algorithm. In 2014, Henriques et al. [12] proposed a high-speed tracking algorithm that uses Kernelized Correlation Filters (KCF) to train discriminant classifiers. Ma et al. [13] proposed a Long-term Correlation Tracking (LCT) algorithm that uses two filters to learn the long-term and short-term appearance of the tracked target. Danelljan et al. [14] proposed a Continuous Convolution Operators (CCOT) tracking algorithm. In the second year, Danelljan et al. [15] proposed an Efficient Convolution Operators (ECO) for tracking, which

Manuscript received November 6, 2021.
Manuscript revised December 16, 2021.
Manuscript publicized January 7, 2022.

[†]The author is with the State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China.
[††]The authors are with the Academy of Space Electronic Information Technology, Xi'an 710100, China.
[†††]The authors are with the School of Aerospace Science and Technology, Xidian University, Xi'an 710071, China.
[††††]The authors are with the Shaanxi Aerospace Technology Application Research Institute Co., Ltd., Xi'an 710100, China.
a) E-mail: wanghai@mail.xidian.edu.cn (Corresponding author)

DOI: 10.1587/transinf.2022DLP0039

optimized the CCOT algorithm in terms of speed.

3. Proposed Model

In this section, the detailed process of the proposed tracking algorithm will be introduced, mainly from the four parts of feature extraction and fusion, multi-scale estimation, adaptive model update [16], [17], and loss re-detection.

Figure 1 is the overall flow chart of the algorithm proposed in this paper. Firstly, the algorithm determines the search area of the current frame according to the tracking results in the previous frame, and performs feature extraction, including HOG features [18], CN features [19], [20], and CNN features. After that, the three features are adaptively fused based on the APCE [21] value and the response map after fusion is calculated. Secondly, by evaluating the confidence of the response map after fusion, it is judged whether the target in the current frame is lost. If the target is lost, the re-detection module will be activated, that is, the EdgeBox [22] is used to extract candidate samples, screen and score, and select the final re-detection results. If the target is not lost, the scale estimation is performed on the basis of the calculated target position. The feature extraction is performed through two methods of scale pyramid and EdgeBox respectively, and the score of each candidate result is calculated. Then the optimal solution is selected as the tracking result of the current frame. Finally, after the tracking result of the current frame is calculated, whether the target is occluded or not is judged according to the previous fusion response map confidence, if so, the updating of the tracking model is stopped, otherwise, the model is updated and the next frame is calculated.

3.1 Feature Extraction and Fusion

The algorithm in this paper mainly extracts HOG features, CN features and CNN features [23], then adaptively fuses them. The adaptive fusion of multiple features is mainly divided into three steps. Firstly, the two manual features of HOG feature and CN feature are fused. Secondly, the response maps calculated by multiple convolution layers in the CNN feature are fused. Finally, the response maps calculated by the two methods are adaptively fused to obtain the final prediction result. The most important point is that after obtaining the response map of the manual feature and the convolution feature, they need to be fused. If the two response maps are simply added together, the ideal complementary learning effect cannot be achieved. Therefore, an adaptive fusion method is proposed in this paper, which can automatically determine the weight ratio between the two according to the confidence of different feature response maps. the APCE is used to measure the confidence of the response map and the calculation formula of APCE is:

$$APCE = \frac{|F_{\max} - F_{\min}|^2}{mean\left(\sum_{w,h}(F_{w,h} - F_{\min})^2\right)} \tag{1}$$

where $F_{\max}$, $F_{\min}$ and $F_{w,h}$ denote the maximum value, minimum value, and the value at the (w, h) in the response map. According to the APCE of the response map, the fusion weight ω of the two response maps can be calculated:

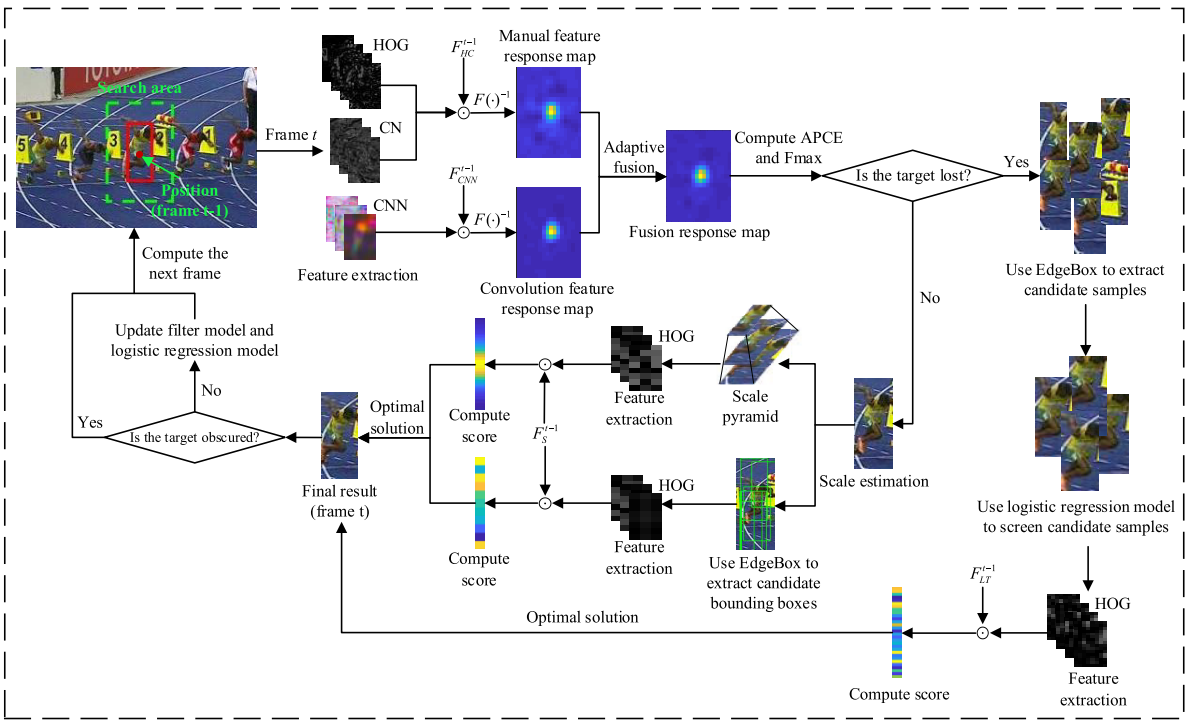


Fig. 1 The overall flow chart of the algorithm

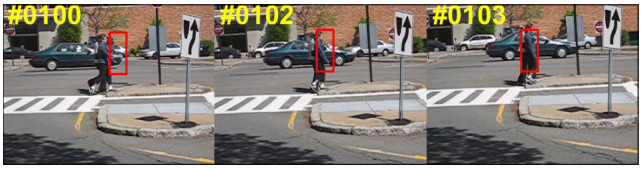


Fig. 2 Tracking results

(a) scale pyramid estimation results

(b) improved scale estimation results

Fig. 3 Scale estimation results before and after improvement

$$\alpha = \frac{APCE_{hc}}{APCE_{hc} + APCE_{cnn}} \tag{2}$$

$$\omega = \frac{\rho}{1 + e^{1-\alpha}} \tag{3}$$

where ρ is a hyperparameter, through which the approximate floating range of the fusion weight ω of the manual feature response map can be set. The final fusion response map is:

$$resp = (1-\omega) \times \frac{resp_{cnn}}{\max(resp_{cnn})} + \omega \times \frac{resp_{hc}}{\max(resp_{hc})} \tag{4}$$

Figure 2 shows the tracking results of 100^{th}, 102^{nd} and 103^{rd} frames in the video sequence. It can be seen from the figure that at the 100^{th} frame, the position of the target cannot be accurately tracked, and there is a short target loss.

3.2 Multi-Scale Estimation

In this paper, the scale pyramid is combined with the EdgeBox candidate bounding box proposal, and the scale estimation is performed after the target position in the current frame is determined. This method solves the problem that the aspect ratio of the target box remains unchanged when the scale pyramid is used alone for scale estimation, and has a better scale estimation effect.

Figure 3 (a) and Fig. 3 (b) are the tracking results of the video sequence *MotorRolling* using the scale pyramid and the improved scale estimation method, respectively. It shows that the proposed improved scale estimation method is effective.

3.3 Adaptive Model Update

By analyzing the response map in the tracking process, it can be concluded that the response map can reflect the confidence of the tracking result to a certain extent. In order to judge the status of the target more accurately, this paper uses the APCE value of the response map and its peak value as two indicators to measure the confidence of the current tracking result.

(a) without adaptive model update

(b) with adaptive model update

Fig. 4 Comparison of tracking results with and without adaptive model update

In each frame, the APCE and $F_{\max}$ of the displacement response map are calculated. If both of these values exceed the historical average value under a certain proportion, it is considered that the target in the current frame is in a relatively good state, and the tracking result of the current frame should be learned to adapt to the gradual change of the target. At this time, the displacement and scale filter models are updated in the following manner:

$$\hat{y}_t = \begin{cases} (1-\eta)\hat{y}_{t-1} + \eta\hat{y}_t, & APCE_t \geq \beta_1 \times APCE^{avg} \\ & \& \ F^t_{\max} \geq \beta_2 \times F^{avg}_{\max} \\ \hat{y}_{t-1}, & else \end{cases} \tag{5}$$

$$\hat{k}^{xx}_t = \begin{cases} (1-\eta)\hat{k}^{xx}_{t-1} + \eta\hat{k}^{xx}_t, & APCE_t \geq \beta_1 \times APCE^{avg} \\ & \& \ F^t_{\max} \geq \beta_2 \times F^{avg}_{\max} \\ \hat{k}^{xx}_{t-1}, & else \end{cases} \tag{6}$$

where η is the learning rate of the model, β_1 and β_2 are the proportional thresholds for determining APCE and $F_{\max}$ whether to be updated respectively.

Figure 4 (a) and Fig. 4 (b) show the tracking results in the video sequence *Lemming* before and after using the adaptive model update strategy. The pictures show that from the 320^{th} frame to 350^{th} frame, the target is gradually completely occluded by other surrounding objects. From the 350^{th} frame to 385^{th} frame, the target reappears in the field of view.

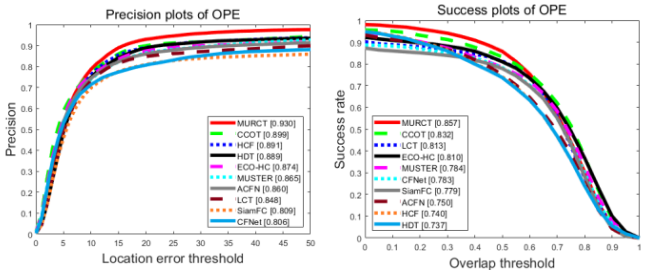
3.4 Re-Detection after Loss

In actual applications of object tracking, target loss due to various factors cannot be completely avoided. When the target is lost, it is difficult for the general tracking algorithm to find the target again, mainly because there is no better re-detection mechanism. We propose a re-detection scheme. By training a logistic regression model and a long-term filter model, when the target is judged to be lost, EdgeBox is used to extract candidate target samples in the surrounding area. The trained logistic regression model and long-term filter are used to score each sample, and the final re-detection

(a) without loss re-detection

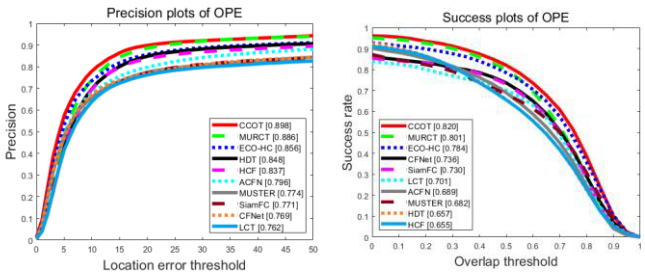
(b) with loss re-detection

Fig. 5 Tracking result of both trackers with or without loss re-detection



(a) distance precision plots (b) success overlap plots

Fig. 6 OPE results of 10 algorithms in OTB2013



(a) distance precision plots (b) success overlap plots

Fig. 7 OPE results of 10 algorithms in OTB2015

result is determined by comprehensive evaluation of the sample score and its position.

In the tracking process, a long-term filter is trained to preserve more stable and lasting target features. The long-term filter is trained by extracting the HOG features of the target area in the first frame, and in order to learn more accurate target features, it is updated only when the peak of the response map is higher than a certain threshold. After judging that the target is lost, the re-detection module is activated. The module firstly calculates the logistic regression score of the candidate samples, and then obtains a part of the candidate samples with high scores to calculate their scores through the long-term filter, and finally compares with the previous response peak to determine whether the target reappears.

Figure 5 (a) and Fig. 5 (b) show the tracking results in the video sequence *Human5* before and after using the target loss re-detection mechanism. It can be seen that the target box deviates slightly from the target position at the 175^{th} frame in Fig. 5 (a), and due to the accumulation of errors, the target is completely lost at the 177^{th} frame. After adding the loss re-detection mechanism, in the 175^{th} and 176^{th} frames of the video sequence, it is judged that the target is lost according to the response map. Then the re-detection module is activated, and the target is successfully relocated.

4. Experiment

In order to verify the effectiveness of the MURCT algorithm proposed in this paper, firstly, it was compared with 9 excellent tracking algorithms (MUSTER [24], HCF [25], HDT [26], SiamFC [9], CCOT [14], ACFN [27], CFNet [28], ECO-HC [15], and LCT [29]) on the OTB dataset. Secondly, on the VOT2016 dataset, MURCT was compared with 16 tracking algorithms (KCF [30], ASMS [31], DSST [32], HCF, deepMKCF [33], DAT [34], SODLT [35], SRDCF [36], CDTT [37], TricTRACK [38], MDNet [39], SiamFC, CCOT, SWCF [40], DPT [41], and SRBT [42]).

4.1 Results on the OTB Dataset

Figure 6 and Fig. 7 show the comparison results of MURCT and 9 algorithms in OTB2013 and OTB2015, respectively. It can be seen that MURCT ranked first in OTB2013, and compared with the second-ranked algorithm CCOT, the distance precision and overlap success rate have increased by 3.1% and 2.5%, respectively. In OTB2015, MURCT ranked second, and compared with the first-ranked algorithm CCOT, the distance precision and overlap success rate were only reduced by 1.2% and 1.9%, respectively.

Table 1 shows the distance precision, overlap success rate and center location error of 10 algorithms at specific thresholds in OTB2013 (I) and OTB2015 (II). It can be seen that in OTB2013, MURCT improved the center location error by 7 pixels compared with the second-ranked algorithm HCF. In OTB2015, MURCT ranked second in the center location error, and was only 0.8 pixels behind the first-ranked algorithm CCOT.

Figure 8 shows the tracking results of 10 tracking algorithms in video sequences *Box*. The tracking results of each algorithm are marked with different colored bounding boxes. In the video sequence *Box*, SiamFC, LCT and HDT lost the target at the 40^{th} frame due to the similarity of the surrounding environment. From the 465^{th} frame to the 500^{th} frame, all the other algorithms except CCOT,

Table 1 Results of 10 algorithms in OTB2013 (I) and OTB2015 (II)

		MURCT	MUSTER	HCF	HDT	SiamFC	CCOT	ACFN	CFNet	ECO-HC	LCT
DP	I	**93**	86.5	89.1	88.9	80.9	89.9	86	80.6	87.4	84.8
(%)	II	88.6	77.4	83.7	84.8	77.1	**89.8**	79.6	76.9	85.6	76.2
OS	I	**85.7**	78.4	74	73.7	77.9	83.2	75	78.3	81	81.3
(%)	II	80.1	68.2	65.5	65.7	73	**82**	68.9	73.6	78.4	70.1
CLE	I	**9.5**	17.9	16.5	16.7	30.4	17.4	19.6	35.9	24.1	27.7
(px)	II	15.3	32.3	23.2	20.6	33.5	**14.5**	26.1	36.4	23.1	67.6

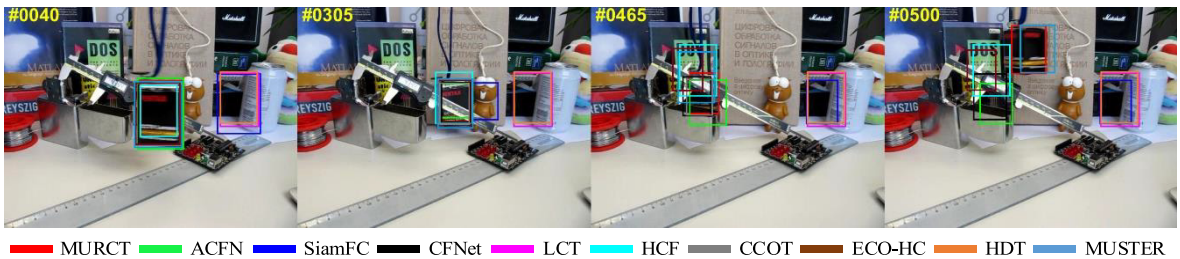


Fig. 8 Tracking results of 10 algorithms in video sequence

Table 2 Results of 17 algorithms in VOT2016

	Accuracy	Robustness	EAO		Accuracy	Robustness	EAO
MURCT	**0.5486**	22.7182	*0.2749*	SWCF	0.4998	41.3131	0.1857
MDNet	0.5443	21.2699	0.2655	ASMS	0.489	30.1444	0.2077
deepMKCF	*0.5392*	26.2313	0.2308	SRBT	0.483	*21.325*	0.2919
SODLT	0.5348	32	0.2375	DPT	0.4783	31.9389	0.2326
SiamFC	0.5245	29.8021	0.2347	TricTRACK	0.4501	35.2548	0.2077
DSST	0.5204	44.8138	0.1793	DAT	0.4474	28.3533	0.2123
SRDCF	0.5179	28.3167	0.2419	HCF	0.4361	23.8569	0.2195
CCOT	0.5155	**16.5817**	**0.3239**	CDTT	0.3984	38.3809	0.1661
KCF	0.5016	38.082	0.2005				

MUSTER and this method could not continue to track the target, because the target reappeared after being completely occluded. MUSTER has the same function of target loss re-detection as this method.

4.2 Results on the VOT2016 Dataset

Table 2 shows the evaluation results of MURCT and 16 comparison algorithms in VOT2016. The first-ranked algorithm in each item is marked in bold, the second-ranked algorithm is underlined, and the third-ranked algorithm is marked in italics. It can be seen from the table that MURCT ranks first among 17 algorithms in terms of accuracy, fourth in robustness, and third in EAO. Figure 9 and Fig. 10 are the performance rankings of tracking algorithms on the VOT dataset and in 6 different situations, respectively. The horizontal and vertical coordinates represent the robustness and accuracy of the algorithm respectively. The closer to the upper right corner of the figure, the better the performance of the tracking algorithm in terms of accuracy and robustness. It can be seen that MURCT is in a relatively prominent

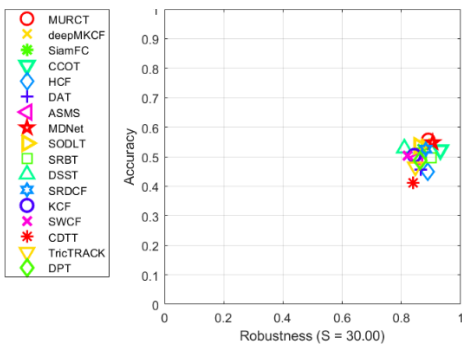
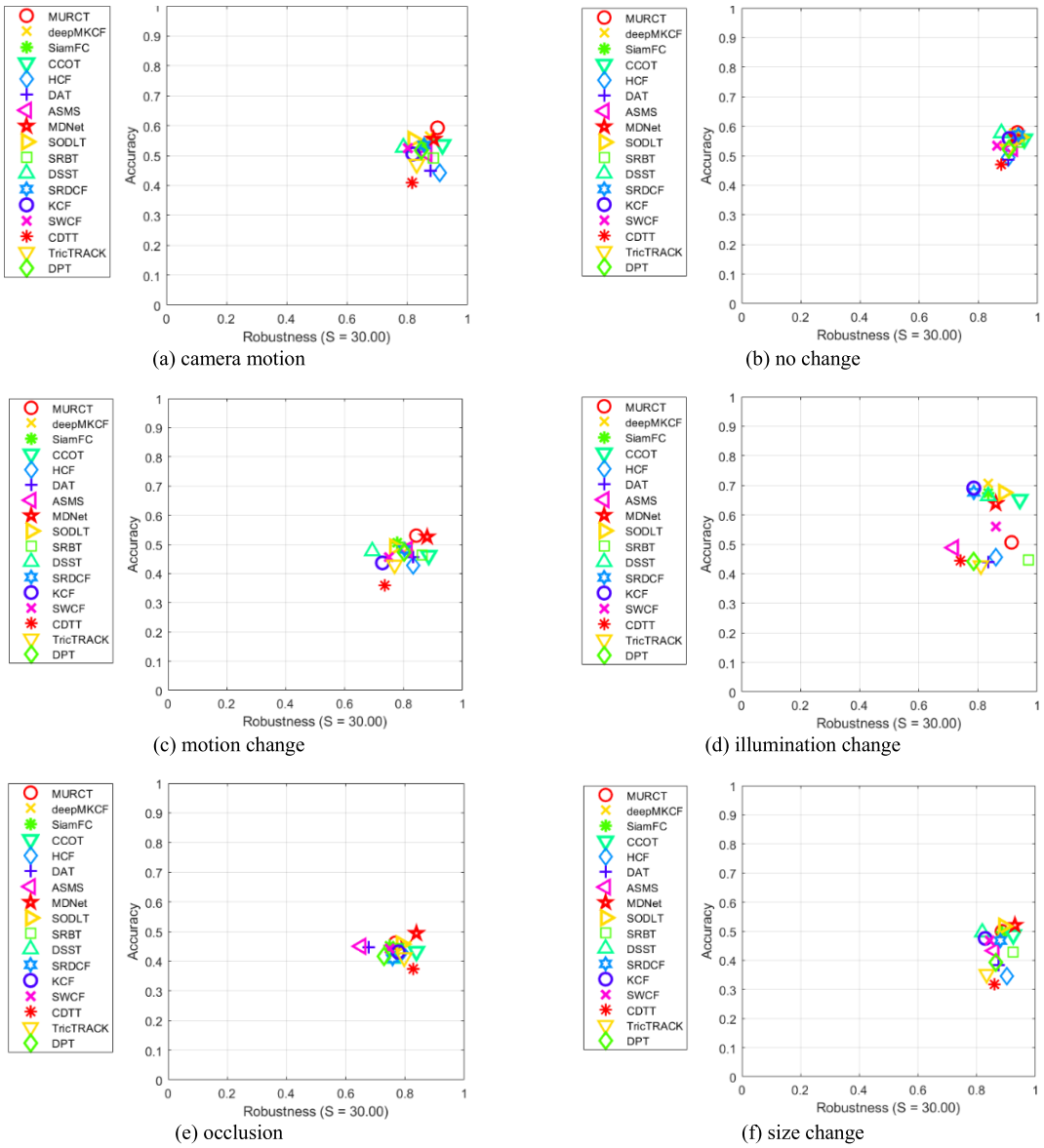


Fig. 9 VOT comparison results

position in terms of accuracy and robustness, especially in terms of camera motion and motion change, it has greater advantages than most other algorithms.

5. Conclusion

Based on the correlation filter tracking algorithm, this paper



(a) camera motion (b) no change

(c) motion change (d) illumination change

(e) occlusion (f) size change

Fig. 10 VOT comparison results under 6 different situations

proposes an improvement plan for several common problems in target tracking, such as model update and loss re-detection. We judge the current state of the target by evaluating the confidence of the tracking response map, and decides whether to update the model according to whether the target is occluded, which solves the problem of model pollution. Then, for the gradual growth of video sequences, it is inevitable that the target will be lost due to wrong estimation during tracking. We judge whether the target is lost according to the confidence of the tracking response map, and then decides whether to activate the re-detection module. Finally, this algorithm is compared with a variety of excellent algorithms on OTB and VOT datasets, which proves the effectiveness and advantages of this algorithm.

Acknowledgments

The authors would like to thank the editor and anonymous reviewers for their valuable comments on this paper. This research is supported financially by Fundamental Research Funds for the Central Universities (Grant No.JB211303).

References

[1] N. Wang, J. Shi, D.-Y. Yeung, and J. Jia, "Understanding and diagnosing visual tracking systems," In Proc. IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 Dec. 2015, pp.3101–3109, 2015.

[2] N. Zeng, H. Li, and Y. Peng, "A new deep belief network-based multi-task learning for diagnosis of Alzheimer's disease," Neural Computing and Applications, 2021. https://doi.org/10.1007/s00521-021-06149-6

[3] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," IEEE Conference on Computer Vision and Pattern Recognition, pp.2411–2418, 2013.
[4] Y. Wu, J. Lim, and M.-H. Yang, "Object Tracking Benchmark," IEEE Trans. Pattern Anal. Mach. Intell., vol.37, no.9, pp.1834–1848, 2015.
[5] M. Kristan, J. Matas, A. Leonardis, T. Vojíř, R. Pflugfelder, G. Fernández, G. Nebehay, F. Porikli, and L. Čehovin "A novel performance evaluation methodology for single-target trackers," IEEE Trans. Pattern Anal. Mach. Intell., vol.38, no.11, pp.2137–2155, 2016.
[6] N. Wax, "Signal-to-Noise Improvement and the Statistics of Track Populations," Journal of Applied physics, vol.26, no.5, pp.586–595, 1955.
[7] W. Jiang and L. Zhang, "Edge-siamnet and edge-triplenet: New deep learning models for handwritten numeral recognition," IEICE Trans. Inf. & Syst., vol.E103-D, no.3, pp.720–723, 2020.
[8] H. Kwon, Y. Kim, H. Yoon, and D. Choi, "Captcha image generation systems using generative adversarial networks," IEICE Trans. Inf. & Syst., vol.E101-D, no.2, pp.543–546, 2018.
[9] L. Bertinetto, J. Valmadre, J.F. Henriques, A. Vedaldi, and P.H.S. Torr, "Fully-convolutional siamese networks for object tracking," European Conference on Computer Vision, Springer, Cham, pp.850–865, 2016.
[10] B. Li, J. Yan, W. Wu, Z. Zhu, and X. Hu, "High performance visual tracking with siamese region proposal network," IEEE Conference on Computer Vision and Pattern Recognition, pp.8971–8980, 2018.
[11] D.S. Bolme, J.R. Beveridge, B.A. Draper, and Y.M. Lui, "Visual object tracking using adaptive correlation filters," IEEE Conference on Computer Vision and Pattern Recognition, pp.2544–2550, 2010.
[12] J.F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," IEEE Trans. Pattern Anal. Mach. Intell., vol.37, no.3, pp.583–596, 2014.
[13] C. Ma, J.-B. Huang, X. Yang, M.-H. Yang, "Adaptive correlation filters with long-term and short-term memory for object tracking," International Journal of Computer Vision, vol.126, no.8, pp.771–796, 2018.
[14] M. Danelljan, A. Robinson, F.S. Khan, and M. Felsberg, "Beyond correlation filters: Learning continuous convolution operators for visual tracking," European Conference on Computer Vision, Springer, Cham, pp.472–488, 2016.
[15] M. Danelljan, G. Bhat, F.S. Khan, and M. Felsberg, "Eco: Efficient convolution operators for tracking," IEEE Conference on Computer Vision and Pattern Recognition, pp.6638–6646, 2017.
[16] N. Zeng, Z. Wang, W. Liu, H. Zhang, K. Hone, and X. Liu, "A dynamic neighborhood-based switching particle swarm optimization algorithm," IEEE Trans. Cybern., vol.52, no.9, pp.9290–9301, 2022. doi: 10.1109/TCYB.2020.3029748.
[17] N. Zeng, D. Song, H. Li, Y. You, Y. Liu, and F. Alsaadic, "A competitive mechanism integrated multi-objective whale optimization algorithm with differential devolution," Neurocomputing, vol.432, pp.170–182, 2021.
[18] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," IEEE Conference on Computer Vision and Pattern Recognition, vol.1, pp.886–893, 2005.
[19] J. Van De Weijer, C. Schmid, J. Verbeek, and D. Larlus, "Learning color names for real-world applications," IEEE Trans. Image Process., vol.18, no.7, pp.1512–1523, 2009.
[20] M. Danelljan, F.S. Khan, M. Felsberg, and J. Van De Weijer, "Adaptive color attributes for real-time visual tracking," IEEE Conference on Computer Vision and Pattern Recognition, pp.1090–1097, 2014.
[21] M. Wang, Y. Liu, and Z. Huang, "Large margin object tracking with circulant feature maps," IEEE Conference on Computer Vision and Pattern Recognition, pp.4021–4029, 2017.
[22] C.L. Zitnick and P. Dollár, "Edge boxes: Locating object proposals from edges," European Conference on Computer Vision, Springer, Cham, pp.391–405, 2014.
[23] N. Zeng, Z. Wang, B. Zineddin, Y. Li, M. Du, L. Xiao, X. Liu, and T. Young, "Image-based quantitative analysis of gold immunochromatographic strip via cellular neural network approach," IEEE Trans. Med. Imag., vol.33, no.5, pp.1129–1136, 2014.
[24] Z. Hong, Z. Chen, C. Wang, X. Mei, D. Prokhorov, and D. Tao, "Multi-store tracker (muster): A cognitive psychology inspired approach to object tracking," IEEE Conference on Computer Vision and Pattern Recognition, pp.749–758, 2015.
[25] C. Ma, J.-B. Huang, X. Yang, and M.-H. Yang, "Hierarchical convolutional features for visual tracking," IEEE International Conference on Computer Vision, pp.3074–3082, 2015.
[26] Y. Qi, S. Zhang, L. Qin, H. Yao, Q. Huang, J. Lim, and M.-H. Yang, "Hedged deep tracking," IEEE Conference on Computer Vision and Pattern Recognition, pp.4303–4311, 2016.
[27] J. Choi, H.J. Chang, S. Yun, T. Fischer, Y. Demiris, and J.Y. Choi, "Attentional correlation filter network for adaptive visual tracking," IEEE Conference on Computer Vision and Pattern Recognition, pp.4807–4816, 2017.
[28] J. Valmadre, L. Bertinetto, J. Henriques, A. Vedaldi, and P.H.S. Torr, "End-to-end representation learning for correlation filter based tracking," IEEE Conference on Computer Vision and Pattern Recognition, pp.2805–2813, 2017.
[29] C. Ma, J.-B. Huang, X. Yang, and M.-H. Yang, "Adaptive correlation filters with long-term and short-term memory for object tracking," International Journal of Computer Vision, vol.126, no.8, pp.771–796, 2018.
[30] J.F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," IEEE Trans. Pattern Anal. Mach. Intell., vol.37, no.3, pp.583–596, 2014.
[31] T. Vojir, J. Noskova, and J. Matas, "Robust scale-adaptive mean-shift for tracking," Pattern Recognition Letters, vol.49, pp.250–258, 2014.
[32] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," IEEE Conference on Computer Vision and Pattern Recognition, vol.1, pp.886–893, 2005.
[33] M. Tang and J. Feng, "Multi-kernel correlation filter for visual tracking," IEEE International Conference on Computer Vision, pp.3038–3046, 2015.
[34] H. Possegger, T. Mauthner, and H. Bischof, "In defense of color-based model-free tracking," IEEE Conference on Computer Vision and Pattern Recognition, pp.2113–2120, 2015.
[35] N. Wang, S. Li, A. Gupta, et al., "Transferring rich feature hierarchies for robust visual tracking," arXiv preprint arXiv 2015:1501.04587, 2015.
[36] M. Danelljan, G. Häger, F.S. Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," IEEE International Conference on Computer Vision, pp.4310–4318, 2015.
[37] J. Xiao, R. Stolkin, and A. Leonardis, "Single target tracking using adaptive clustered decision trees and dynamic multi-level appearance models," IEEE Conference on Computer Vision and Pattern Recognition, pp.4978–4987, 2015.
[38] X. Wang, M. Valstar, B. Martinez, M.H. Khan, and T. Pridmore, "Tric-track: Tracking by regression with incrementally learned cascades," IEEE International Conference on Computer Vision, pp.4337–4345, 2015.
[39] H. Nam and B. Han, "Learning multi-domain convolutional neural networks for visual tracking," IEEE Conference on Computer Vision and Pattern Recognition, pp.4293–4302, 2016.
[40] E. Gundogdu and A.A. Alatan, "Spatial windowing for correlation filter based visual tracking," IEEE International Conference on Image Processing, pp.1684–1688, 2016.
[41] A. Lukežič, L.Č. Zajc, and M. Kristan, "Deformable parts correlation filters for robust visual tracking," IEEE Trans. Cybern., vol.48, no.6, pp.1849–1861, 2017.
[42] H. Lee and D. Kim, "Salient region-based online object tracking," IEEE Winter Conference on Applications of Computer Vision, pp.1170–1177, 2018.

Xianyu Wang received the B.S. degrees at the School of telecommunications engineering, Xidian University, Xi'an, China, in 2004, and the M.S. degree in telecommunications engineering from Xidian University, Xi'an, China, in 2008. He is currently pursuing the Ph.D. degree at Xidian University of telecommunications engineering, Xi'an, China. His research interests include intelligent information processing and artificial intelligence.

Cong Li received his Ph.D. degree at Xidian University, Xi'an, China, in 2018. He works as an engineer at Academy of Space Electronic Information Technology, Xi'an, where he leads the research team of "intelligent satellite system". He has published one book on machine learning and more than 10 journals and conferences. His research interests are intelligent satellite communications, software-defined multi-function satellite payloads, and so on.

Heyi Li received his B.S. and M.S. degrees, respectively, in 2018 and 2021 in School of Aerospace Science and Technology, both in Xidian University (Xi'an, China). He focuses on researching in image processing.

Rui Zhang received M.S. degree in control theory and control engineering from Xi'an University of Technology. His research field include communication and information technology.

Zhifeng Liang graduated from Fujian Provincial Space Information Engineering Research Center of Fuzhou University in 2013. Currently, he is a system engineer and mainly engaged in the design of remote sensing satellite ground application system, artificial intelligence and big data algorithm research.

Hai Wang received Ph.D. degree at Xidian University, Xi'an, China, in 2007. His research interests include circuit and systems, signal processing and artificial intelligence.