

PAPER • OPEN ACCESS

## Target Tracking Based on Correlation Filter for Scale Offset

To cite this article: Jianhao Tan *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **790** 012038

View the [article online](#) for updates and enhancements.

You may also like

- [Target tracking algorithm of intermittently updated template based on reliability evaluation](#)

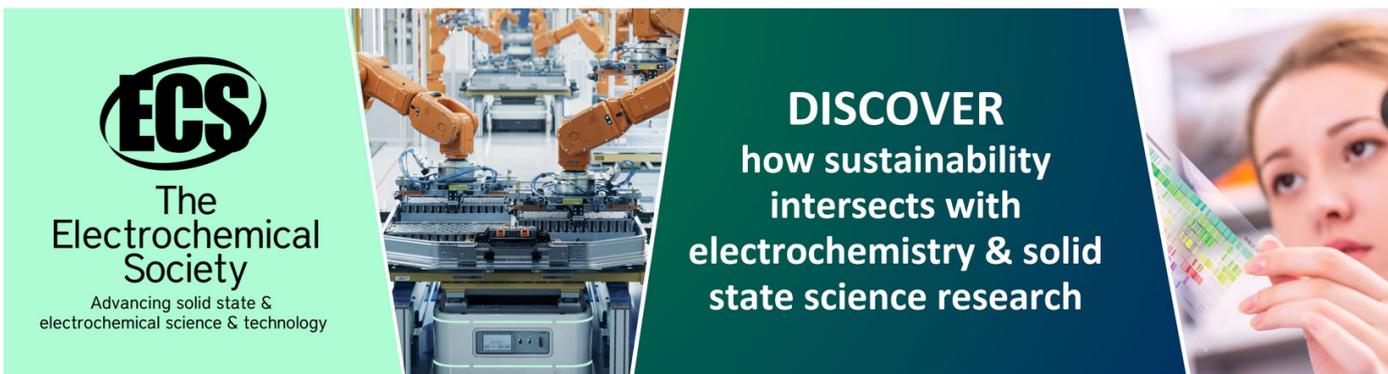
Qingsong Xie, Jingwei Shen and Zhiyong An

- [Scale Adaptive Target Tracking Algorithm Based on Correlation Filtering](#)

Fajun Lin, Haiping Wei, Yuanchen Qi et al.

- [Real-time Visual Tracking Based on Convolutional Neural Networks](#)

Rui Li and Jirong Lian



**ECS**  
The  
Electrochemical  
Society  
Advancing solid state &  
electrochemical science & technology

**DISCOVER**  
how sustainability  
intersects with  
electrochemistry & solid  
state science research

# Target Tracking Based on Correlation Filter for Scale Offset

Jianhao Tan<sup>1,2</sup>, Xiaoping Ma<sup>1,2,\*</sup> and Yaonan Wang<sup>1,2</sup>

<sup>1</sup>College of Electrical and Information Engineering, Hunan University, Changsha 410082, China

<sup>2</sup>National Engineering Laboratory for Robot Visual Perception and Control Technology, Changsha 410082, China

\*Corresponding author

**Abstract.** Existing scale estimation methods mainly use fixed-size tracking box for target tracking. However, when the moving direction of the target changes from near to far or from far to near, the size of the current tracking box cannot adapt to the scale change. In order to overcome the disadvantage of invariant size of tracking box in the scale change, this paper proposes a scale offset estimation method based on correlation filter by combining the state of previous frame and current frame of the target. Select the scale proposal box in the scale layer, and adjust the position and size of the actual tracking box in real time according to the size of the proposal box, target position and scale offset. In this paper, OTB-100 is used as the dataset. The obtained tracking results show that our algorithm Ours has better tracking performance than the existing others tracking algorithm in scale variation, and it has better tracking effect and stronger robustness in drift and occlusion events.

## 1. Introduction

With the improvement of automation technology, the application scope of computer vision is expanding. Target tracking<sup>[1]</sup> is one of the main applications of computer vision and one of the basic problems to be solved in robust control of computer vision. Target tracking is based on the state of the previous and the current moment, to find the exact location and size of the target in the video sequence at the next time<sup>[2]</sup>, which is convenient for the higher level processing and analysis of the target. However, due to the restriction of tracking environment and the limitation of tracking algorithm, the research of target tracking method needs to be further improved and developed.

In recent years, the method of target tracking based on the idea of correlation filter<sup>[3]</sup> has been widely adopted. However, during the scale change process the fixed tracking box<sup>[4]</sup> can't well solve the tracking difficulty problem caused by the scale change. Until now, target tracking method based on correlation filter is still a research hotspot in the field of computer vision<sup>[5]</sup>, especially in the tracking of target when scale changes. In literature [6], Danelljan et al.<sup>[1][2]</sup> adopted the idea of scale pyramid weight decreasing to solve the scale change problem. However, this method uses the initial size of the target as the center layer. When the scale changes, the response value of the actual tracking box will decrease, and there will be deviations between the layers of the pyramid, resulting in the accumulation of scale estimation errors. Aiming at the occlusion problem, Ma et al.<sup>[7]</sup> Used the response value of the target tracking model to determine the occlusion, but this process is greatly affected by the changes in the appearance of the tracking target<sup>[8]</sup>, which easily leads to target tracking failure.

Inspired by the word "concern"<sup>[9-10]</sup>. This paper is based on the basic principles of correlation filters, the scale proposal box is used in the scale change, which solves the problem that the size of the tracking box is invariable when the target changes from near to far or far to near scale. At the same time,



combined with the direction of scale change, the tracking box is adapted to the changes of the target on different scale layers in real time according to the operation of the proposed box size and scale offset<sup>[11]</sup>, so as to achieve accurate target tracking.

## 2. Scale Analysis

In order to make better use of the thought of the kernel correlation filter, based on previous studies, we apply the correlation filter method in the process of scale change analysis to effectively improve the robustness of target tracking.

### 2.1. Ridge Regression Analysis

Combining the previous and current states of the tracking target and the scale change, the correlation filter is regarded as a classifier according to the target tracking scale estimation principle. After the cyclic shift of the scale layer  $x_i^s$ , all the training samples on the scale layer are obtained, and the regression target of each sample is represented by  $y_i^s$ . Its purpose is to identify functions<sup>[12]</sup>  $f(x_i^s) = \alpha^T x_i^s y_i^s$  to find  $\alpha$ . The training and learning process combined with the least square error method is as follows:

$$\min_{\alpha} \sum (f(x_i^s) - y_i^s)^2 + \mu \|\alpha\|^2 \tag{1}$$

The purpose of the solution is to find  $\alpha$  that minimizes the least square error equation, according to the operation principle of diagonal matrix elements, the following formula can be obtained:

$$\alpha = x^{s*} \odot y^s / (x^{s*} \odot x^s + \mu) \tag{2}$$

### 2.2. Scale Layer Circular Matrix Analysis

In order to achieve accurate target tracking, the cyclic displacement theory is adopted to sample the objects and the samples around the objects in the scale layer.

The reference object of pixels is represented by  $x_i^s$ , using the vector of  $n \times 1$  to represent, where the reference vector is  $x_i^s = [x_0^s, x_1^s, \dots, x_{n-1}^s]$ . The scale training sample  $X^s$  was obtained by cyclic shift<sup>[12]</sup> of the reference object sample. The cyclic matrix  $X^s = \{p^w x_i^s | w=0,1,\dots,n-1\}$  is obtained by using the permutation matrix  $P$ .

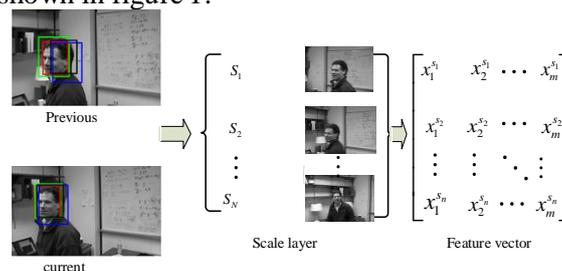
Where the permutation matrix  $P$  can be expressed as:

$$P = \begin{bmatrix} 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Where the scale cyclic matrix  $X^s$  can be expressed as:

$$X^s = \begin{pmatrix} x_0^s & x_1^s & \dots & x_{n-1}^s \\ x_{n-1}^s & x_0^s & \dots & x_{n-2}^s \\ \vdots & \vdots & \ddots & \vdots \\ x_1^s & x_2^s & \dots & x_0^s \end{pmatrix}$$

Combining the target state of the previous frame and the current frame, the cyclic sample matrix of the scale layer is obtained, as shown in figure 1.



**Figure 1.** Scale layer cyclic sample matrix.

### 2.3. Kernel Tricks Analysis

The kernel tricks<sup>[13]</sup> is used to classify nonlinear features in high-dimensional space. The initial target input is mapped to the nonlinear target tracking space, Through the nonlinear mapping function, the linear combination of the input samples can be expressed as:

$$\alpha = \sum_j \delta^s \varphi(x_j^s) \quad (3)$$

Where  $\varphi(x_j^s)$  represents the distribution of multi-scale samples in higher-dimensional space. therefore, according to the formula transformation, the least square formula on the scale layer is obtained as follows:

$$\min_{\delta^s} \sum_i (\sum_j \delta^s k(x_i^s, x_j^s) - y_i^s)^2 + \mu \left\| \sum_j \delta^s \varphi(x_j^s) \right\|^2 \quad (4)$$

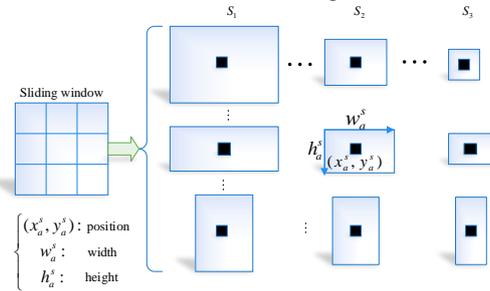
Finally, the ridge regression analysis is used to calculate the  $\delta^s$  in the frequency domain. The model is as follows:

$$\delta^s = y^s / k^{ij} + \mu \quad (5)$$

According to the above analysis, the scale response based on the correlation filter is the matrix representing the target confidence.

### 3. Select the Scale Proposal Box

According to the sample cyclic matrix to determine the position of the maximum response value. Based on the center of the target position, keep the center unchanged and slide the rectangular box to predict multiple regions with different coordinates, as shown in figure 2.



**Figure 2.** Scale proposal box selection.

Select the box most likely to be the target as the proposal box according to the size of the correlation. The selection method is as follows:

#### 3.1. Discard Strategy

Generally, only the nearest bounding box around the target is kept on the classification feature map, and the bounding box that is too far away from the center of the target is discarded. For a moving target, the size of the bounding box around the tracking target will not change much in scale relative to the bounding box far from the center of the target. The specific discarding scheme is selected step by step based on classification and regression.

Assuming that there are t scale tracking boxes, it will output 2t classification channels and 4t regression channels. The size and position of the scale proposal box are expressed by the coordinate set, such as formula (6) and formula (7).

$$A_{cls}^{w \times h \times 2t} = \{(x_i^s)^{cls}, (y_j^s)^{cls}\} \quad (6)$$

$$\begin{cases} A_{reg}^{x_i^s \times y_j^s \times 2t} = \{(x_i^s)^{reg}, (y_j^s)^{reg}\} \\ A_{reg}^{w \times h \times 4t} = \{(dx_i^s)^{reg}, (dy_j^s)^{reg}, (dw_a^s)^{reg}, (dh_a^s)^{reg}\} \end{cases} \quad (7)$$

Where  $(x_i^s, x_j^s)$  represents the central position coordinates,  $(w_a^s, h_a^s)$  represents the width and height,  $dx_i^s, dy_i^s$  and  $dw^s, dh^s$  represent the target center and position scale offsets, respectively.

### 3.2. Coincidence Ratio Comparison

In order to select the best scale proposal box, a threshold value is set to determine the maximum coincidence rate between the scale proposal box and the actual target bounding box in this paper. the upper limit value of the coincidence rate is set to 0.7, the lower limit value of the coincidence rate is set to 0.3.

## 4. Scale Offset Adjustment Tracking Box

The position corresponding to the maximum response value obtained from the scale cyclic matrix is taken as the target point. Multiple prediction boxes with different coordinate sizes are obtained with the target point as the center, and the tracking box with the maximum confidence<sup>[14]</sup> is obtained as the proposal box according to the selection method of the scale proposal box.

### 4.1. Classification and Regression

Determine the target category and location based on the size of the correlation. Where the scale proposed sample is denotable as  $g^s$ , and the target sample is denotable as  $x^s$ . Firstly, find the target position through correlation analysis, and then according to the direction of scale change to adjust the size of boundary box. classification and regression as follows.

$$N_{cls} = \varphi(g^s)_{cls} * \varphi(x^s)_{cls} \quad (8)$$

$$N_{reg} = \varphi(g^s)_{reg} * \varphi(x^s)_{reg} \quad (9)$$

Where \* indicates convolution operation,  $\varphi(g^s)$  represents the kernel space,  $\varphi(g^s)_{cls}$  and  $\varphi(g^s)_{reg}$  represent the category and location, respectively.  $\varphi(x^s)$  represents the target sample kernel space on the scale layer,  $\varphi(x^s)_{cls}$  and  $\varphi(x^s)_{reg}$  represent the category and location, respectively.

### 4.2. Calculate the Offset and Determine the Proposal Box

Based on the above analysis, the target tracking loss function is defined as:

$$loss(a_i^s, b_i^s) = \frac{1}{N_{cls}} \sum_i L_{cls}(a_i^s, (a_i^s)^s) + \mu \frac{1}{N_{reg}} \sum_i (a_i^s)^s L_{reg}(b_i^s, (b_i^s)^s)$$

Where  $i$  represents the number of proposal boxes,  $a_i^s$  represents the  $i$  proposal box with the highest similarity to the tracking target,  $(a_i^s)^s$  means that when the proposal box is a positive sample, its value is 1, otherwise it is 0.  $b_i^s$  represents a vector consisting of the center position of the actual box as well as its width and height,  $(b_i^s)^s$  represents the actual tracking box that is related to the positive sample proposal box.  $(a_i^s)^s L_{reg}$  means that the regression loss is considered only when  $(a_i^s)^s = 1$ .

In order to adapt to scale changes, the position and size of the tracking box are normalized as follows:

$$\begin{cases} dx_i^s = \frac{(x^s - x_i^s)}{w_a^s}, dy_j^s = \frac{(y^s - y_j^s)}{h_a^s} \\ dw^s = \log(w^s/w_a^s), dh^s = \log(h^s/h_a^s) \end{cases} \quad (10)$$

Where  $(x^s, y^s)$ ,  $w^s, h^s$  represent the center, width and height of the actual tracking box.

When the scale of the target changes, the proportion of positive and negative samples in the actual tracking box will change accordingly. Therefore, the offset between the selected scale proposal box and the actual box with the greatest correlation needs to be calculated. According to formula (10), the position and size of the offset can be deduced, which can be expressed as:

$$\begin{cases} (dx_i^s)^* = \frac{((x^s)^* - x_i^s)}{w_a^s}, (dy_j^s)^* = \frac{((y^s)^* - y_j^s)}{h_a^s} \\ (dw^s)^* = \log((w^s)^*/w_a^s), (dh^s)^* = \log((h^s)^*/h_a^s) \end{cases} \quad (11)$$

Combining the above formula (11), the position and size of the actual tracking box in the scale change can be deduced as shown in formula (12).

$$\begin{cases} (x^s)^* = x_i^s + (dx_i^s)^* * w_a^s, (y^s)^* = y_j^s + (dy_j^s)^* * h_a^s \\ (w^s)^* = w_a^s * e^{(dw^s)^*}, (h^s)^* = h_a^s * e^{(dh^s)^*} \end{cases} \quad (12)$$

In summary, according to the scale proposal frame and position offset, the actual frame size in the scale change is adjusted, and the coordinates corresponding to the position and size of the actual tracking box of the target as follows:  $\{(x^s)^*, (y^s)^*\}, \{(w^s)^*, (h^s)^*\}$ .

## 5. Analysis of Results

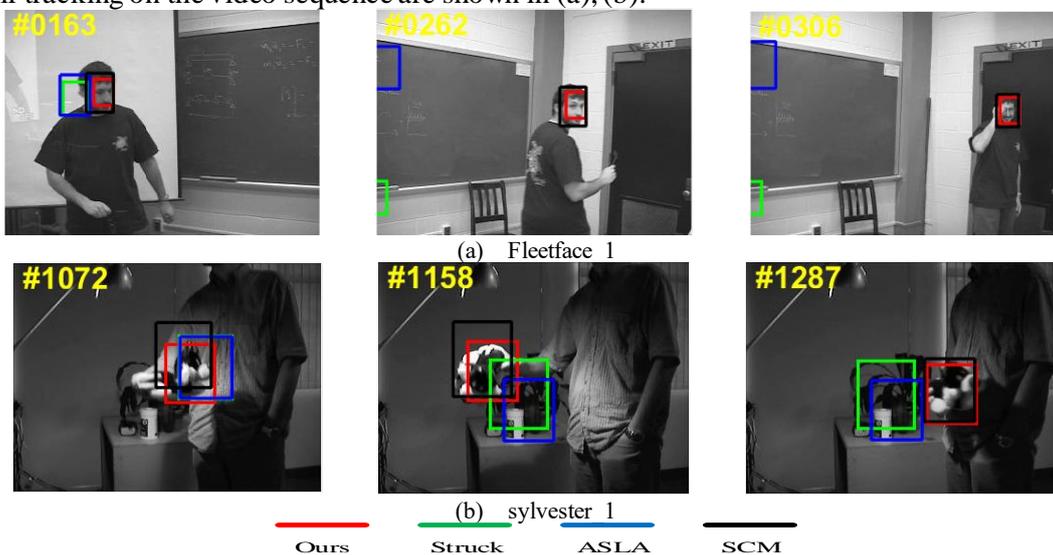
### 5.1. Experimental Configuration

The trackers involved in this paper are all completed on the MATLAB2017a programming platform, The experimental hardware environment was configured to run on an Intel(R) Core(TM) i7-4510u CPU on a 4GB PC.

In this paper, Struck, ASLA, SCM are compared and analyzed with Ours. The results of all test sequence runs were evaluated using SRE and TRE and OPE. At the same time, different tracking attributes in scale changes are considered, and the specific comparison results are shown in table 1.

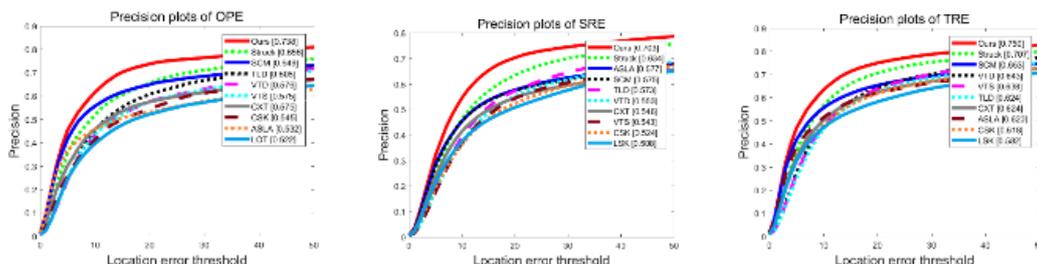
### 5.2. Overall Results Analysis

The otb-100 dataset was used as the test set for validation, the first four trackers (Ours, Struck, ASLA, SCM) with better overall tracking performance were selected through testing. The comparison results of their tracking on the video sequence are shown in (a), (b).

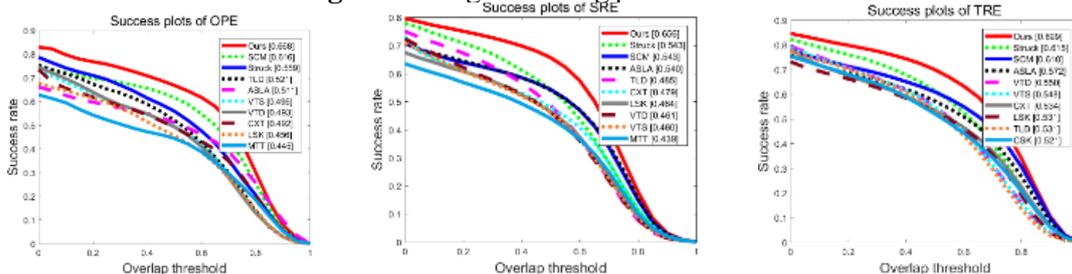


**Figure 3.** Comparison chart of tracking effect in the process of scale change.

According to the results of Ours algorithm, this paper analyzes OPE, SRE and TRE respectively. The error range for all sequences on the precision graph is [0,50] and the ordinate range is [0,0.9]. The overlap range of all sequences in the success rate graph is [0,1], and the ordinate range is [0,0.9]. As shown in Figure 4, Figure 5.



**Figure 4.** Target tracking precision chart.



**Figure 5.** Target tracking success rate chart.

In order to visualize the results, only the top 10 trackers with better accuracy are selected in Figure 4 and Figure 5. The result shows that Ours algorithm is obviously better than other tracking algorithms. The precision of Ours algorithm is 70.3%, 75.0% and 73.8% respectively on SRE, TRE and OPE, and the success rate is 65.6%, 69.9% and 66.8% respectively on SRE, TRE and OPE. In terms of time robustness (TRE), Ours tracker improves tracking accuracy by 6.1% and improves success rate by 13.7% compared to Struck tracker.

*5.3. Analysis of Different Trace Attributes*

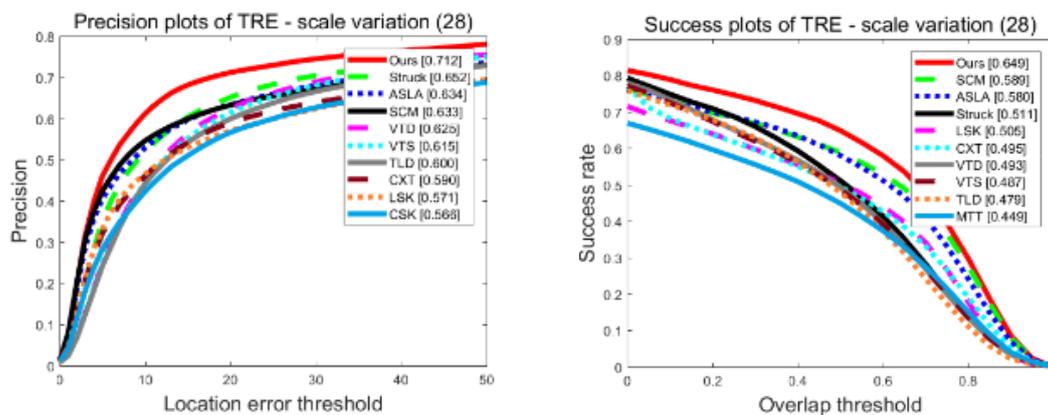
In order to demonstrate the robustness of the algorithm, the five tracking attributes of deformation (SD), background clutter (SBC), occlusion (SO), removed field of view (SOV) and scale change (SV) were analyzed. As shown in table 1.

**Table 1.** Parameter comparison of different trackers under different tracking properties.

		SD	SBC	SO	SOV	SV
Ours	precision	0.720	0.649	0.726	0.577	0.712
	success	0.709	0.651	0.679	0.609	0.649
Struck	precision	0.655	0.622	0.631	0.484	0.652
	success	0.605	0.580	0.551	0.489	0.511
ASLA	precision	0.571	0.575	0.560	0.0	0.634
	success	0.557	0.524	0.519	0.0	0.580
SCM	precision	0.635	0.600	0.633	0.0	0.633
	success	0.616	0.548	0.597	0.0	0.589

This paper compares and analyzes the tracking effect of Ours tracker and Struck, ASLA, SCM, which are 4 advanced trackers under the 5 tracking attributes of deformation(SD), background clutter (SBC), occlusion (SO), removed field of view (SOV) and scale change (SV). Table 1 shows that: Compared with the existing advanced methods, our algorithm Ours has improved tracking accuracy and tracking success rate on these five tracking attributes, and its tracking accuracy has reached 71.2% and the tracking success rate has reached 64.9% in the scale change, compared with the advanced algorithm Struck, its accuracy is improved by 9.2%, and the success rate is increased by 27%.

Combining the tracking results of the Ours algorithm in scale changes in Table 1, this paper makes an overall robust assessment of the scale changes with respect to 28 sequences representing the properties of scale changes, using time robustness (TRE) as an evaluation index, as shown in Figure 6.



**Figure 6.** Time Robustness Evaluation.

It can be seen from figure 6, Ours algorithm has improved its precision tracking performance and success rate tracking performance compared with other advanced tracking algorithms in the overall tracking test of scale change.

## 6. Summary

In this paper, a scale proposal box is proposed in the scale layer based on the correlation filter, and the offset between the proposed box size and the actual tracking box is used to overcome the problem that the position and size of the target tracking box are immutable in the scale change. In the tracking process, the correlation analysis is carried out on the scale layer based on the correlation filter, and the real-time adjustment of the position and size of the current scale layer tracking box is realized by combining the position and size of the proposed box and the offset size.

## Acknowledgements

Fund Project: National Natural Science Foundation of China (61433016).

## References

- [1] Danelljan M, Häger G, Khan F, et al. Accurate scale estimation for robust visual tracking[C]//British Machine Vision Conference, Nottingham, September 1-5, 2014. BMVA Press, 2014.
- [2] Wu Y, Lim J, Yang M H. Online object tracking: A benchmark[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2013: 2411-2418.
- [3] Dai K, Wang D, Lu H, et al. Visual Tracking via Adaptive Spatially-Regularized Correlation Filters[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 4670-4679.
- [4] Sun L, Fan B, Peng Y. Adaptive scale correlation filter tracker with feature fusion[C]//2018 Chinese Control And Decision Conference (CCDC). IEEE, 2018: 5949-5953.
- [5] Zhang F, Wang Z. Visual Tracking Algorithm Based on Improved Kernelized Correlation Filters[C]//2018 International Conference on Networking and Network Applications (NaNA). IEEE, 2018: 337-343.
- [6] Wang Yanchuan, Huang Hai, Li Shaomei, et al. Correlation Filter Tracking Based on Onilne Detection and Scale-Adaption[J]. Acta Optica Sinica, 2018, 38(02): 227-239.
- [7] Ma C, Yang X, Zhang C, et al. Long-term correlation tracking[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 5388-5396.
- [8] Lan X, Ye M, Zhang S, et al. Robust collaborative discriminative learning for rgb-infrared tracking[C]//Thirty-Second AAAI Conference on Artificial Intelligence. 2018.
- [9] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[C]//Advances in neural information processing systems. 2015: 91-99.
- [10] Chorowski J K, Bahdanau D, Serdyuk D, et al. Attention-based models for speech recognition[C]//Advances in neural information processing systems. 2015: 577-585.

- [11] Li B, Yan J, Wu W, et al. High performance visual tracking with siamese region proposal network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 8971-8980.
- [12] Xu F, Wang H, Song Y, et al. A multi-scale kernel correlation filter tracker with feature integration and robust model updater[C]//2017 29th Chinese Control And Decision Conference (CCDC). IEEE, 2017: 1934-1939.
- [13] Scholkopf B, Smola A J. Learning with kernels: support vector machines, regularization, optimization, and beyond[M]. MIT press, 2001.
- [14] Wang N, Zhou W, Tian Q, et al. Multi-cue correlation filters for robust visual tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 4844-4853.