Long-term target tracking combined with re-detection

Juanjuan Wang [1], Haoran Yang [1], Ning Xu [1], Chengqin Wu[1], Zengshun Zhao[1, 2, 3, *], Jixiang Zhang[1, *], Dapeng Oliver Wu[3]

1 College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao, 266590, P.R. China. wangjuanjuan6@163.com (JW); yohor.yanghaoran@gmail.com (HY); 2727021504@qq.com (NX); 736725790@qq.com (CW)
2 School of Control Science & Engineering, Shandong University, Jinan, 250061, China
3 Department of Electrical& Computer Engineering, University of Florida, Gainesville, FL 32611, USA. dpwu@ufl.edu
* Correspondence Author: zhaozs@sdust.edu.cn
* Co-Correspondence Author: zjxhii@163.com

Abstract:

The long-term visual tracking undergoes more challenges and is closer to realistic applications than short-term tracking. However, most existing methods have not been done and their performances have also been limited. In this work, we present a reliable yet simple long-term tracking method, which extends the state-of-the-art Discriminative Correlation Filters (DCF) tracking algorithm with a re-detection component based on the SVM model. The DCF tracking algorithm localizes the target in each frame and the re-detector is able to efficiently re-detect the target in the whole image when the tracking fails. We further introduce a robust confidence degree evaluation criterion that combines the maximum response criterion and the average peak-to correlation energy (APCE) to judge the confidence level of the predicted target. When the confidence degree is generally high, the SVM is updated accordingly. If the confidence drops sharply, the SVM re-detects the target. We perform extensive experiments on the OTB-2015 dataset, the experimental results demonstrate the effectiveness of our algorithm in long-term tracking.

Keywords: Discriminative correlation filtering; Long-term tracking; Re-detection

1. Introduction

While visual object tracking as a hot research topic in computer vision has been widely applied in various fields, many challenges are still not resolved especially in target disappears, partial occlusion, background clutter, studying a general and powerful tracking algorithm is a tough task

A typical scenario of visual tracking is to track an unknown object in subsequent image frames by giving the initial state of a target in the first frame of the video. In the past few decades, the visual object tracking technology has made significant progress [1] [2] [3] [4] [5] [6] [7] [8] [9] [10]. These methods are very effective for short-term tracking tasks, which is that the tracked object is almost always in the field of view. However in realistic applications, the requirement for tracking is not only to track correctly, but also to track for a longer period of time [11], during the period of time, the tracking output is wrong in the case of target objects absent, the training samples will be incorrectly annotated, leading a risk of model drift. Therefore, it is important to long-term trackers to determine whether the target is absent, and have the capability of re-detection.

Long-term tracking task also require the tracker as well as short-term tracking to maintain high-accuracy in the challenges of disappearance and occlusion, especially to stability capture the target object in long-term video [12]. Therefore, the long-term tracking presents more challenges from two aspects. Firstly, how to determine the confidence degree of the tracking results, in [13], the maximum response value of the target is used to determine the confidence of the tracking results. When the maximum peak value of the response map is lower than the threshold value, the result is determined to be unreliable. However, the response maps may fluctuate drastically when the object in occlusion or disappear condition, only using the maximum response value to judge confidence is incredibility. The APCE criterion in [14] indicates the degree of fluctuation of the response map. If the target is

undergoing fast motion, the value of APEC will be low even if the tracking is correct. However, the APCE criterion is commonly used to update trackers in [14]. Secondly, how to relocate the out-of-view targets. TLD [15] algorithm exploits an ensemble of weak classifiers for global re-detection the out-of-view. The method fails to classify the target object due to the huge number of scanning windows. LCT [13] algorithm proposes a random ferns re-detection model to detect the out-of-view target. In [16], it learns a spatial-temporal filter in a lower dimensional discriminative manifold to alleviate the influences of boundary effects, but the method still cannot solve the target disappearance problem.

This paper proposes a tracking algorithm combining the discriminative correlation filter tracker and re-detector. The proposed method aims to perform robust re-detection and relocate the target when target tracking fails. Firstly, we introduce a state-of-the-art discriminative correlation filter to track the target online. Then, the tracking confidence degree criterion, which combines the maximum response and the APCE, is used to evaluate the tracking results. Finally, we use a robust re-detection strategy based on SVM to perform re-detection. We evaluate the proposed tracking algorithm on the OTB-2015 dataset, the experimental results show that the proposed algorithm performs more stable and accurate tracking performance in the case of occlusion, background clutter, etc.

The structure of the rest of the paper is as follows: Section 2 overviews the related work. Section 3 presents the proposed method. Section 4 reports the experimental results and experimental analysis. Section 5 concludes the paper.

2. Related work

2.1. Correlation filter

Correlation filters have shown outstanding results for target tracking [17] [18]. These methods

exploit the circular correlation of the filter in frequency domain to locate the target object. Bolme et al. [4] propose the pioneering MOSSE tracker, using only gray features to train the filter. CSK tracker [19] employs illumination intensity features and applies DCFs in a kernel space. The KCF [6] further improves CSK by using multi-channel HOG features. Danelljan et al. [5] exploit color attributes of target object and learn an adaptive correlation filter.

The estimation of target scale is also an important standard for testing a well tracker. It provides improved performance while being computationally efficient. DSST tracker [20] performs translation estimation and scale estimation separately, using a scale pyramid to reply to the scale change. Li and Zhu [21] present an effective scale adaptive scheme, which defines a scale pool, turn the samples of each scale into the same size as the initial sample by the bilinear interpolation method.

The formulation of DCFs use circular correlation which implements learning efficiently by applying Fast Fourier Transform (FFT), however, it also induces circular boundary effects, which has a drastic impact on tracking performance. Danelljan et al. [22] suggest reducing these boundary effects by introducing a spatial regularization component. However, regularization will make the model optimization cost higher. Galoogahi et al. [23] propose the pre-multiplying a fixed masking matrix containing the target regions to address such deficiency of DCFs. Then using the ADMM [24] algorithm to solve the constrained optimization problem in real time. CACF [25] algorithm considers the global information, selects the background reference around the target, and adds the background penalty to the closed solution of the filter. CSRDCF [26] method distinguishes the foreground and background by segmenting the colors in the search area. LADCF [16] approach based on the DCF method, adding adaptive spatial feature selection and temporal consistency constraints alleviate the spatial boundary effects and temporal filter degradation problems that exist in the DCF method.

2.2. Long-term tracking

Kalal et al. [15] propose a tracking-learning-detection (TLD) algorithm, which decomposes the tracking task into tracking, learning and detection, among them, tracking and detection facilitates each other, the short-term tracker provides training examples for the detector, the detectors are implemented as a cascade to reduce computational complexity. Enlightened by the TLD framework, Ma et al. [13] propose a long-term correlation filter tracker using a KCF as a baseline algorithm and a random ferns classifier as a detector. The FCLT [27] trains several correlation filters on different time scales as a detector and exploits the correlation response to link the short-term tracker and long-term detector.

3. The proposed long-term tracking algorithm

In this section, we describe our tracker. In section 3.1, we introduced the main structure of our algorithm, which is shown in figure 1. In section 3.2, we introduce the tracker based on LADCF correlation filtering. In section 3.3, we introduce the confidence degree evaluation criterion and the re-detector.

3.1. The main structure of the algorithm in this paper

The proposed algorithm aims to combine both the discriminative correlation filter tracker and the re-detector for long-term tracking. First, the baseline correlation filter tracker is used to translation estimation in the tracking stage. Second, we use the maximum response value and the APCE criterion to judge the confidence level of the target. Finally, when the value of confidence is higher than the threshold, the baseline tracker achieves the tracking target alone. When the confidence level is drop sharply, it indicating the tracking failure, we do not update model and exploit the SVM model to

re-detect the target object in the current frame. The structure of the algorithm in this paper is shown in Figure 1.

Insert Figure 1

Figure 1. The structure of the algorithm in this paper

3.2. Correlation filter tracker

In this paper, we set LADCF [16] as the baseline algorithm of our tracking approach.

The LADCF algorithm proposes a new DCF-based tracking method, which utilizes the adaptive spatial feature selection and temporal consistent constraints to reduce the impact of spatial boundary effect and temporal filter degradation. The feature selection process is to select several specific elements in the filter to retain distinguishable and descriptive information, forming a low-dimensional and compact feature representation. The spatial feature selection embedded in the learning stage can be expressed as:

$$\underset{\theta,\phi}{\operatorname{argmin}} \|\theta \circledast x - y\|_2^2 + \lambda_1 \|\phi\|_0$$

$$s.t.\, \theta = \theta_\phi = diag(\phi)\theta, \tag{1}$$

where the indicator vector ϕ can potentially be expressed by θ and $\|\phi\|_0 = \|\theta\|_0$, $diag(\phi)$ is the diagonal matrix generated from the indicator vector of selected features ϕ. The ℓ_0-norm is non-convex, and the ℓ_1-norm is widely used to approximate the sparsity [24], so a temporal consistent is constructed by ℓ_1-norm relaxation spatial feature selection model:

$$\underset{\theta}{\operatorname{argmin}} \|\theta \circledast x - y\|_2^2 + \lambda_1 \|\theta\|_1 + \lambda_2 \|\theta - \theta_{model}\|_1 \tag{2}$$

where λ_1 and λ_2 are tuning parameters and $\lambda_1 << \lambda_2$.

Using the ℓ_2-norm relaxation to further simplify:

$$\underset{\theta}{\operatorname{argmin}} \|\theta \circledast x - y\|_2^2 + \lambda_1 \|\theta\|_1 + \lambda_2 \|\theta - \theta_{model}\|_2^2 \tag{3}$$

where the lasso regularization controlled by λ_1 select the spatial feature. In the above formula, the filter template model is used to increase smoothness between consecutive frames to promote time consistency. In this way, the temporal consistency of spatial feature selection can be preserved to extract and retain the diversity of static and dynamic appearance.

Since the multi-channel features share the same spatial layout, the multi-channel input is represented as $\mathrm{X} = \{\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_\mathrm{L}\}$, and the corresponding filter is represented as $\theta = \{\theta_1, \theta_2, \ldots, \theta_\mathrm{L}\}$, by minimizing, the goal can be extended to multi-channel functions with structured sparsity:

$$\underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{L} \|\theta_i \circledast x_i - y\|_2^2 + \lambda_1 \left\| \sqrt{\sum_{i=1}^{L} \theta_i \odot \theta_i} \right\|_1 + \lambda_2 \sum_{i=1}^{L} \|\theta_i - \theta_{model\,i}\|_2^2 \tag{4}$$

where θ^{j} is the j-th element of the i-th channel feature vector $\theta_\mathrm{i} \in \mathbb{R}^{\mathrm{D}^2}$. The structured spatial feature selection term calculates the ℓ_2-norm of each spatial location, and then executes the ℓ_1-norm to achieve joint sparsity.

Subsequently, utilizing ADMM [24] to optimize the above formula, we introduce relaxation variables to construct goals based on convex optimization [28], obtain the global optimal solution of the model through ADMM, and form an enhanced Lagrange operator:

$$\begin{aligned} \mathcal{L} = & \sum_{i=1}^{L} \|\theta_i \circledast x_i - y\|_2^2 + \lambda_1 \sum_{j=1}^{D^2} \left\| \sqrt{\sum_{i=1}^{L} \left(\theta_i^j\right)^2} \right\|_1 + \lambda_2 \sum_{i=1}^{L} \|\theta_i - \theta_{model\,i}\|_2^2 \\ & + \frac{\mu}{2} \sum_{i=1}^{L} \left\| \theta_i - \theta_i^{'} + \frac{\eta_i}{\mu} \right\|_2^2 \end{aligned} \tag{5}$$

where $\mathcal{H} = \{\eta_1, \eta_2, \ldots, \eta_\mathrm{L}\}$ are the Lagrange multipliers and $\mu > 0$ is the corresponding penalty parameter controlling the convergence rate [29]. As $\mathcal{L}$ is convex, ADMM is exploited iteratively to

optimize the following sub-problems with guaranteed convergence:

$$\begin{cases} \theta = arg\min_{\theta} \mathcal{L}(\theta, \theta', \mathcal{H}, \mu) \\ \theta' = arg\min_{\theta'} \mathcal{L}(\theta, \theta', \mathcal{H}, \mu) \\ \mathcal{H} = arg\min_{\mathcal{H}} \mathcal{L}(\theta, \theta', \mathcal{H}, \mu) \end{cases} \tag{6}$$

3.3. Re-detector

3.3.1. Confidence Criterion

Most existed trackers do not consider whether the detection is accurate or not. In fact, once the target is detected incorrectly in the current frame, severely occluded or completely missing, this may cause the tracking failure in subsequent frames.

We introduce a measure for determining confidence degree of target objects, which is the first step in re-detection model. The peak value and the fluctuation of the response map can reveal the confidence about the tracking results. The ideal response map should have only peak and be smooth. Otherwise, the response map will fluctuate intensely. If we continue to use the unsure samples to track target in subsequent frames, the tracking model will be destroyed. Thus, we exploit two fusion confidence degree evaluation criteria. The first one is the maximum response value $\mathrm{F_{max}}$ of the current frame.

The second one is average peak-to-correlation energy (APCE) measure which is defined as

$$\mathrm{APCE} = \frac{|\mathrm{F_{max}} - \mathrm{F_{min}}|^2}{\mathrm{mean}(\sum_{\mathrm{w,h}}(\mathrm{F_{w,h}} - \mathrm{F_{min}})^2)} \tag{7}$$

where, the $\mathrm{F_{max}}$ and $\mathrm{F_{min}}$ are the maximum response and the minimum response of the current frame, $\mathrm{F_{w,h}}$ is the element value of the w-th row and the h-th column of the response matrix.

If the target is moving slowly and is easily distinguishable, the APCE value is generally high. However, if the target is undergoing fast motion with significant deformations, the value of APCE will

be low even if the tracking is correct.

3.3.2. Target Re-detection

In this section, we describe the re-detection mechanism used in the case of tracking failure. In the re-detection module, when the confidence level is lower than the threshold, the SVM is used for re-detection. Consider a sample set $(x_1, y_1), (x_2, y_2), \ldots, (x_i, y_i), \ldots, x_i \in R^d$, including positive and negative samples, where d is the dimension of the sample, $y_i \in (+1, -1)$ is sample labels. SVM can make segmentation of positive and negative samples to obtain the best classification hyperplane. The classification plane is defined as:

$$\omega^T x + b = 0 \tag{8}$$

where w represents the weight vector and b denotes the bias term. In the case of the linearly classifiable, for a given dataset T and classification hyperplane, the following formula is used for classification judgment:

$$\begin{cases} \omega^T x + b \leq 1, y_i = -1 \\ \omega^T x + b \geq 1, y_i = +1 \end{cases} \tag{9}$$

Combining the two equations, we can abbreviate it as:

$$y\omega^T x + b \geqslant 1 \tag{10}$$

The distance from each support vector to the hyperplane can be written as:

$$d = \frac{|\omega^T x + b|}{\|\omega\|} \tag{11}$$

The problem of solving the maximum partition hyperplane of the SVM model can be expressed as the following constrained optimization problem:

$$\min \frac{1}{2} \|\omega\|^2$$

$$s.t.\, y_i(\omega^T x_i + b) \geq 1 \tag{12}$$

Next, the paper introduces the Lagrangian function to solve the above problem [30].

$$L(\omega, \lambda, c) = \frac{1}{2}\|\omega\|^2 - \sum_{j=1}^{l} c_i y_i(\omega \cdot x_i + b) + \sum_{i=1}^{l} c_i \tag{13}$$

where $c_i > 0$ is the Lagrange multiplier, the solution of the optimization problem satisfies the partial derivative of $L(\omega, \lambda, c)$ to ω and b be 0. The corresponding decision function as expressed:

$$y(x) = \mathrm{sign}(\omega^T x + b) \tag{14}$$

Import the new sample points into the decision function to get the sample classification.

For the case of linear inseparability, we use the kernel function to map it to the high-dimensional space. In this work, we use the Gaussian kernel function as follows:

$$k(x_i, x_j) = e^{\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)} \tag{15}$$

When a frame is re-detected, an exhaustive search is performed on the current frame using a sliding window, and HOG features are extracted for each image patch, and the SVM is calculated. The sample with the highest score is calculated the maximum response. When the response value is greater than the threshold, it will be used as the location of the tracking target again. For the high-confidence images, when the response value is greater than the update threshold, we utilize the following formula to update the SVM classifier weight:

$$\omega^* = \sum_{j=1}^{l} c_j^* y_j x_j \tag{16}$$

$$b^* = y_i - \sum_{j=1}^{l} y_j c_j^* (x_j \cdot x_i) \tag{17}$$

where c_j is the Lagrangian coefficient, x is the feature vector extracted from sample, and y is the label corresponding to the sample.

4. Experimental results

In this section, we evaluate the proposed algorithm on OTB-2015 benchmark 17 that contains 100

videos with comparisons to other detection-based tracking algorithms and classical correlation filtering tracking algorithms. Section 4.1 introduces the experimental platform and parameter settings of this experimental algorithm; Section 4.2 introduces the experimental datasets and evaluation criteria for this experiment; Section 4.3 describes the quantitative evaluation of the results and describes the qualitative evaluation in section 4.4

4.1. Experimental setups

The experimental software environment is MATLAB R2016a, and the hardware environment is: Intel Core i5-4200M processor, 4GB memory, Windows 8 operating system.

The regularization parameters λ_1 and λ_2 are set to 1 and 15, respectively, the initial penalty parameter $\mu = 1$, the maximum penalty parameter $\mu_{\max} = 20$, the maximum number of iterations $K = 2$, the padding parameter as $\varrho = 4$, the scale factor as $a = 1.01$, the threshold for re-detection is set to $tr = 0.13$, and the update threshold is set to $tu = 0.20$.

4.2. Experimental dataset and evaluation criteria

The OTB-2015 dataset has a total of 100 video sequences, including 11 challenges, 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) and Low Resolution (LR). The evaluation criterion adopts the tracking success rate and overlap precision in One-Pass Evaluation (OPE) as the criteria of the evaluation algorithm. The overlap success rate is defined as the percentage of frames where the bounding box overlap surpasses a threshold. The distance precision shows the percentage of frames whose estimated location is within the given threshold distance of the ground truth. Normally, the tracking success rate

threshold is 0.5, and the tracking precision threshold is 20 pixels.

4.3. Quantitative Evaluation

In this paper, we compare our algorithm with 6 state-of-the-art trackers on OTB-2015 dataset, including tracking-by-detection algorithms, including LCT [13], LMCF [14] and mainstream correlation filtering tracking algorithms, including CSK [19], KCF [6], DSST [20] and BACF [23]. Figure 2 is OPE success rate and precision plots of these algorithms, it can be seen from the figure that the proposed algorithm has been significantly improved compared with other algorithms. The precision and success rate of our method are 80.2% and 70.6%, respectively. Through experiments, we found that the short-term target tracker will learn the wrong information, resulting in the template is pollute by the wrong information and unable to track the target correctly in subsequent frames, when the target is occluded or disappears. Therefore, compared with the BACF algorithm, our method improves the precision and success rate by 10.6% and 7.3%, respectively. The LCT exploits the random fern algorithm to re-detect targets, the random fern algorithm is slow to operate, thus compared with the tracking-by-detection LCT algorithm, and the proposed algorithm improves the precision and success rate by 3.1% and 9.3% respectively. Compared with the LMCF algorithm with adding multi-peak detection, the precision and success rate of our method increased by 0.8% and 10.8% respectively.

Insert Figure 2

Figure 2. Precision and success rate plots of proposed method and state-of-art methods over OTB-2015 benchmark sequences using one-pass evaluation (OPE).

In order to further verify the superiority of our method, we analyze the tracking performance through attribute-based comparison in Table 1, which shows the complete the area-under-the-curve (AUC) score of the success plots with 11 different attributes.

Table 1. The AUC scores of success plots on OTB-2015 sequences with different attributes.

	CSK	KCF	DSST	BACF	LCT	LMCF	Ours
IV	0.343	0.449	0.538	0.512	0.497	0.517	**0.570**
SV	0.300	0.373	0.457	0.502	0.415	0.457	**0.564**
OCC	0.311	0.418	0.453	0.452	0.451	0.474	**0.573**
DEF	0.312	0.415	0.425	0.468	0.466	0.447	**0.525**
MB	0.297	0.426	0.444	0.505	0.497	0.471	**0.556**
IPR	0.348	0.433	0.483	0.479	0.510	0.448	**0.554**
FM	0.305	0.432	0.428	0.486	0.481	0.445	**0.538**
OPR	0.326	0.428	0.456	0.481	0.479	0.464	**0.563**
OV	0.234	0.364	0.358	0.464	0.413	0.436	**0.514**
BC	0.377	0.463	0.503	0.525	0.499	0.497	**0.570**
LR	0.275	0.307	0.395	0.514	0.299	0.412	**0.545**

As shown in Table 1, the proposed algorithm in this paper achieves the best performance on 11 attributes. In the case of Occlusion (OCC), our algorithm score is 9.9% higher than that of the tracking-by-detection LMCF algorithm and 12.1% higher than the short-term correlation filtering algorithm BACF. For Fast Motion (FM) images, our algorithm is 5.2% higher than the second-ranked BACF algorithm, and 5.7% higher than the LCT algorithm using random fern re-detection. In the above condition, the target model may be contaminated, which makes target tracking difficult, our

model can solve this problem by using re-detection. In the case of OPR, LCT achieves a score of 47.9%, our tracker provides a gain of 8.4%. This is because the baseline algorithm applied in this paper solves the influence of boundary effects to a certain extent, and can achieve higher accuracy when the target rotation occurs.

4.4. Qualitative Evaluation

We selected 7 representative benchmark sequences from OTB-2015 to demonstrate the effectiveness of our algorithm. The visual evaluation results are shown in Figure 3. As can be seen from the figure 3, in the "jogger" sequence, when the 70th frame target is blocked and the 84th frame target reappears in the field of view, due to the re-detection algorithm, our tracker can be track target correctly, and the short-time correlation filter tracking algorithm learns error information during occlusion, which leads to tracking errors in subsequent frames. In the "soccer" and "Matrix" sequences, due to background clutter, algorithms such as LCT and BACF lose the target, but the algorithm in this paper can correctly track the target. In the "Car4" sequence, due to the scale change problem, the scale-based DSST algorithm and the algorithm in this paper show better performance. In "shaking" sequence, the algorithm loses its target in the 17th frame due to issues such as similar lighting changes and background, but due to the addition of a re-detection algorithm, the algorithm relocates to the target at the 18th frame and keeps tracking correctly. In the "Bolt" sequence, our algorithm follows the target closely and keeps a high degree of overlap due to the rapid motion of the target and other issues. In the "Dog" sequence, when the target is deformed, our algorithm can accurately track the target, and the BACF and LMCF algorithms will have a certain offset. It can be seen from this that our algorithm achieves higher accuracy and overlap in these 7 sequences.

Insert Figure 3

Figure 3. The tracking results of each algorithm on 7 video sequences (from top to bottom are Jogging, Soccer, Matrix, Car4, Shaking, Bolt, Dog, respectively)

Furthermore, we compare our method with the baseline tracker using 7 representative benchmark sequences of OTB-2015 in Figure 4. The first three rows are short-term sequences which none of them exceed 1000 frame, the last four rows are long-term sequences, which all exceed 1000 frames.

As shown in Figure 3, in the experiments based on the short-term sequences, the LADCF tracker drifts when target objects undergo heavy occlusions (Soccer) and does not re-detect targets in the case of tracking failure. Moreover, the LADCF tracker fails to handle background clutter and deformation (Ironman, Bird1), where only tracking component is less effective to discriminate targets from the cluttered background. However, our method can track the object correctly on these challenging sequences because of the trained detector effectively re-detects target objects.

In addition, we compare our method with the baseline tracker using 4 long-term benchmark sequences of OTB-2015.

In the Sylvester and Lemming sequences, both LADCF tracker track incorrectly objects due to the rotating conditions encountered in these sequences, our method provides better robustness to these conditions. In the Liquor sequence, the LADCF tracking algorithm is similar to our algorithm before the target is occluded. But when the target is occluded, the LADCF method locates on the occluded target. In the Rubik sequence, at 854th frame, since the target object has undergone deformation and color variation, the LADCF tracker fails to track correctly. Our method is able to track successfully due

to re-detection. In our method, if the tracking fails, we perform the re-detection and initialize the tracker so that the target is re-detected. Thus, our method can correctly track the target all the time.

Overall, our method performs well in estimating positions of target objects, which can be attributed three reasons. First, the confidence criterion of our method can correctly identify the target of low confidence objects. Second, our re-detection component effectively re-detects target objects in the case of tracking failure. Third, our baseline tracker achieves adaptive discriminative filter learning on a low dimensional manifold and improves the tracking effect.

Insert Figure 4

Figure 4. The tracking results of two algorithms on 6 video sequences (from top to bottom are Soccer, Ironman, Bird1, Sylvester, Lemming, Rubik, Liquor, respectively)

5. Conclusion

This paper proposes a long-term target tracking algorithm, the two main components of proposed algorithm are a state-of-the-art DCF short-term tracker which estimate the target translation and a re-detector which re-detect the target objects in the case of tracking failure. Besides, the algorithm introduces a robust confidence criterion to evaluate the confidence value of the predicted target. When the confidence value is lower than threshold, the SVM model is used to re-detect the target objects and the template is not updated. The algorithm is suitable for long-term tracking because it can detect the accuracy of the target in real time and update the template with high reliability. Numerous experimental

results show that the proposed algorithm achieves the better performances than the other tracking algorithms.

Acknowledgments: This work was supported in part by the China Postdoctoral Science Special Foundation Funded Project (2015T80717).

Competing interests: The authors declare that they have no competing interests.

Author's contributions: ZZ and DOW proposed the original idea of the full text; JZ and JW designed the experiment; JW and NX performed the experiment; JW and HY wrote the manuscript under the guidance of ZZ. CW, JZ and JW revised the manuscript. All authors read and approved this submission.

References

1. Comaniciu, Dorin, Meer, Peter. Mean Shift: A Robust Approach Toward Feature Space Analysis [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence. 24(5): 603-619 (2002)
2. M. S. Arulampalam, S. Maskell, N. Gordon and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," in IEEE Transactions on Signal Processing, vol. 50, no. 2, pp. 174-188, (2002).
3. Kalman, R. E. A New Approach to Linear Filtering and Prediction Problems [J]. Journal of Basic Engineering, 82(1):35-45 (1960).
4. D.S. Bolme, J.R. Beveridge, B.A. Draper, Y.M. Lui, Visual object tracking using adaptive correlation filters, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, 2010, pp. 2544–2550 (2010).
5. M. Danelljan, F. Shahbaz Khan, M. Felsberg, J. Van de Weijer, Adaptive color attributes for real-time visual tracking, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1090–1097 (2014).
6. J.F. Henriques, R. Caseiro, P. Martins, J. Batista, High-speed tracking with kernelized correlation filters, IEEE Trans. Pattern Anal. Mach.Intell. 37 (3), 583–596 (2015).
7. Nam H, Han B. Learning Multi-domain Convolutional Neural Networks for Visual Tracking [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, CVPR, 4293-4302 (2016).
8. Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional siamese networks for object tracking [C] // Computer Vision-ECCV 2016 Workshops. (Springer, ECCV, 2016), 9914: 850-865.
9. Gundogdu Erhan, Alatan A A. Good Features to Correlate for Visual Tracking [J] // IEEE Transactions on Image Processing. 27(5), 2526-2540 (2018).
10. Asadi, M., Regazzoni, C.S. Tracking Using Continuous Shape Model Learning in the

Presence of Occlusion.EURASIP J. Adv. Signal Process. 2008, 250780 (2008).
11. Li T, Zhao S, Meng Q, et al. A Stable Long-Term Object Tracking Method with Re-detection Strategy [J]. Pattern Recognition Letters, 119-127 (2018).
12. Bin Yan, Haojie Zhao, Dong Wang, Huchuan Lu, Xiaoyun Yang. 'Skimming-Perusal' Tracking: A Framework for Real-Time and Robust Long-term Tracking [C]. IEEE/CVF International Conference on Computer Vision (ICCV), pp. 2385-2393 (2019).
13. C. Ma, X. Yang, C. Zhang, M.H. Yang, Long-term correlation tracking, in: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5388–5396 (2015).
14. Wang, M., Liu, Y., & Huang, Z. Large margin object tracking with circulant feature maps. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition pp. 4021-4029 (2017).
15. Z. Kalal, K. Mikolajczyk, J. Matas, Tracking-learning-detection, IEEE Trans. Pattern Anal. Mach. Intell. 34 (7) 1409-1422 (2012)
16. Xu, Tianyang et al. "Learning Adaptive Discriminative Correlation Filters via Temporal Consistency Preserving Spatial Feature Selection for Robust Visual Object Tracking." IEEE Transactions on Image Processing 28.11, 5596–5609 (2019)
17. Y. Wu, J. Lim, M.-H. Yang, Object tracking benchmark, IEEE Trans. Pattern Anal. Mach. Intell. 37 (9) 1834–1848, (2015)
18. Wu Y, Lim J, Yang M H. Online Object Tracking: A Benchmark[C]//Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, OR, USA: IEEE, 2411-2418. [DOI:10.1109/CVPR.2013.312] (2013).
19. J.F. Henriques, R. Caseiro, P. Martins, J. Batista, Exploiting the circulant structure of tracking-by-detection with kernels, in: European Conference on Computer Vision, (Springer, 2012), pp. 702–715.
20. Martin Danelljan, Gustav Häger, Fahad Shahbaz Khan, and Michael Felsberg. Accurate Scale Estimation for Robust Visual Tracking. Proceedings of the British Machine Vision Conference. BMVA Press, (2014).
21. Y. Li, J. Zhu, A scale adaptive kernel correlation filter tracker with feature integration, in: European Conference on Computer Vision, Springer, pp. 254–265 (2014).
22. M. Danelljan, G. Hager, F. Shahbaz Khan, M. Felsberg, Learning spatially regularized correlation filters for visual tracking, in: Proceedings of the IEEE International Conference on Computer Vision, pp. 4310–4318 (2015).
23. H. Kiani Galoogahi, A. Fagg, S. Lucey, Learning background-aware correlation filters for visual tracking, in: Proceedings of the IEEE International Conference on Computer Vision, pp. 1135–1143 (2017).
24. Boyd S, Parikh N, Chu E, et al. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers[M] // Boyd S, Nparikh N, Chu E, et al. Foundations and Trends in Machine Learning, Now Publishers Inc. 2011, 3(1): 1-122.
25. A. Lukezic, T. Vojir, L. C. Zajc, J. Matas, and M. Kristan, "Discriminative correlation filter with channel and spatial reliability," in IEEE Conference on Computer Vision and Pattern Recognition, pp. 4847–4856 (2017).
26. H. Kiani Galoogahi, A. Fagg, and S. Lucey, "Learning background-aware correlation filters for visual tracking," in IEEE International Conference on Computer Vision, 2017.

27. Lukežič, Alan & Čehovin Zajc, Luka & Vojíř, Tomáš & Matas, Jiri & Kristan, Matej. (2017). FCLT - A Fully-Correlational Long-Term Tracker. (2017)
28. F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Structured sparsity through convex optimization," Statistical Science, pp. 450–468 (2012).
29. D. P. Bertsekas, Constrained optimization and Lagrange multiplier methods. Academic press, (1982).
30. Joachims, T. Making Large-Scale SVM Learning Practical. In: Scholkopf, B., Burges, C. and Smola, A., Eds., Advances in Kernel Methods Support Vector Learning, Chapter 11, MIT Press, Cambridge, 169-184 (1999).