Long-Term Object Tracking Method Based on Dimensionality Reduction

Dongfang Zhao, Wenjing Kang(✉), and Gongliang Liu

School of Information Science and Engineering,
Harbin Institute of Technology, Weihai, China
catmouseand@foxmail.com, {kwjqq,liugl}@hit.edu.cn

Abstract. Long-term object tracking is challenging as target objects often undergo drastic appearance changes over time. Recently, the FDSST algorithm has performed very well which reduces number of the FFT by dimensionality reduction and operates at real-time. But in case of long-term tracking the performance of FDSST degrades, and the existing long-term tracking methods cannot guarantee the accuracy and real-time performance simultaneously. To solve the above problems, we input a set of sample patches of the target appearance to a multi-channel correlation filter to locate the position of the target in a new frame. At the same time, the number of FFTs is reduced by dimensionality reduction, and an online SVM is trained as the detector to ensure the accuracy of target tracking. Finally, we get a method to track long-term object accurately and in real time. To evaluate the method, we did extensive experiments on a benchmark with 100 sequences. The results show that the proposed method performs well both in accuracy and real-time performance and outperforms than the state-of-the-art methods.

Keywords: Long-term tracking · Dimensionality-reduction · Online SVM · Ream-time tracking

1 Introduction

Robust tracking is a fundamental problem in computer vision, with many applications in areas such as robotics, surveillance and human- machine interfaces [1]. Meanwhile, object tracking is challenging due to the large appearance variation caused by deformation, illumination change, heavy occlusion, sudden motion, and target disappearance. Most of the existing tracking algorithms usually could solve one of the problems, they will fail when other problems arise in the same time. Danelljan et al. proposed the FDSST algorithm in 2017 aimed to improve the accuracy and the robustness of the tracker [2]. This novel scale adaptive tracking approach learns separate discriminative correlation and has an outstanding performance when illumination change or the sudden motion occurs. But the FDSST will usually fail even when the object is under slight occlusion.

In the process of long-term object tracking, occlusion often occurs. Therefore, the performance of FDSST is not ideal for long-term tracking. But FDSST is very fast, usually up to 80FPS during the experiment in my computer. In long-term tracking,

M. Jia et al. (Eds.): WiSATS 2019, LNICST 280, pp. 529–536, 2019.
https://doi.org/10.1007/978-3-030-19153-5_54

people generally solve the problem of target tracking failure by adding a detector. In [3], Ma et al. activated an online trained detector to recover the target object. This online detector can effectively solve the problem of target tracking failure caused by occlusion.

In order to solve the problem of low speed in long-term object tracking, we refer to the methods of dimensionality reduction and the online detector, then apply those methods in the long-term object tracking. The experiments show that our method could solve the problem of occlusion very well, and the method can achieve real time results while maintaining accuracy.

2 Related Work

Object tracking is always the most active research topic in computer vision. In this section, we discuss the most closely related approaches. In recent years, the correlation filter has been widely used, because the operator is readily transferred into the Fourier domain, then the algorithm using the correlation filter would own the fastest speed. In 2010, Bolme et al. firstly proposed a Minimum Output Sum of Squared Error filter (MOSSE) [4] that is computationally efficient. Many effort has been made to improve tracking performance since then using correlation filters. In 2012, Henriques, et al. proposed an extension include kernelized correlation filter [5]. In 2013, Galoogahi, et al. proposed a novel framework for learning a multi-channel detector/filter efficiently in the frequency domain. This multi-channel filter is better than the single-channel correlation filters [6]. In 2015, Henriques et al. derive a new Kernelized Correlation Filter (KCF), and based on the KCF, Henriques J F proposed a fast multi-channel extension of linear correlation filters, which called Dual Correlation Filter (DCF) [7]. In the same year, Danelljan et al. proposed Spatially Regularized Discriminative Correlation Filters (SRDCF) for tracking. This filter can be learned on a large number of negative training samples without damaging the positive ones [8].

Recently, long-term visual tracking has received attention. Kalal et al. divided the tracking task into tracking, learning and detection (TLD) [9] where tracking and detection facilitates each other, the results from the tracker provide training data to update the detector. And the detector re- initializes the tracker when it fails. In [10] Supancic et al. proposed a simple and effective system that trains online SVM by learning a large number of negative training examples. The good effect of the system proves the practicability of the above theory in long – term target tracking. In [11], Kang et al. proposed a continuous correlation filter to achieve subpixel object locations in continuous domain. The author learn an online random fern classifier to redetect the target in case of tracking failure. Due to its excellent classification strategy, the CCFT tracker can locate the target with high accuracy.

In this paper, in order to speed up the tracking of long-term target, we adopted the method of dimensionality reduction, which reduced the number of FFT needed to be calculated. By feature dimension reduction, we obtained a higher speed, and the tracking precision is also very high.

3 Proposed Algorithm

3.1 Multi-channel Discriminative Correlation Filters

Our tracking approach is based on FDSST. The FDSST based tracking approaches learn an optimal correlation filter for locating the target in a new frame.

At each location n in a rectangular domain, let target sample f consists of a dimensional feature vector $f(n) \in \mathbb{R}^d$. Then denote feature channel $l \in \{1, \ldots \mathrm{d}\}$ of f by f^l. This is achieved by minimizing the L^2 error of the correlation response compared to the desired correlation output g,

$$\varepsilon = ||\mathrm{g} - \sum_{l=1}^{d} h^l \odot f^l||^2 + \lambda \sum_{l=1}^{d} ||h^l||^2 \tag{1}$$

The symbol $\odot$ denotes circular correlation. The symbol λ is a weight parameter. Note that the domains of f^l, h^l and g all have the same dimension and size.

Transforming (1) to the Fourier domain using Parseval's formula would solve (1) efficiently. The filter that minimizes (1) is given by,

$$H^l = \frac{\overline{G}F^l}{\sum_{k=1}^{d} F^k \overline{F^k} + \lambda}, l = 1, \ldots, d. \tag{2}$$

Here, the capital letters denote the discrete Fourier transform (DFT) of the corresponding quantities. The bar $\overline{G}$ denotes complex conjugation.

To apply the filter in a new frame t, we extracted a sample z_t from a considered region. We will get the DFT of the correlation scores y_t by computed in the Fourier domain

$$Y_t = \frac{\sum_{l=1}^{d} \overline{A_{t-1}^l} Z_t^l}{B_{t-1} + \lambda} \tag{3}$$

Here, we update the numerator and the denominator of the filter by

$$A_t^l = (1 - \eta) A_{t-1}^l + \eta \overline{G} F_t^l, l = 1, \ldots, d \tag{4}$$

$$B_t = (1 - \eta) B_{t-1} + \eta \sum_{k=1}^{d} \overline{F_t^k} F_t^k \tag{5}$$

Finally, finding the maximum correlation score, we get the estimate of the current target.

3.2 Dimensionality Reduction

To reduce the number of FFT, we employ a dimensionality reduction strategy. We obtain P_t by minimizing the reconstruction error of the target template u_t

$$\varepsilon = \sum_{n} ||u_t(n) - P_t^T P_t u_t(n)||^2 \tag{6}$$

We update the u_t by $u_t = (1-\eta)u_{t-1} + \eta f_t$. When the orthonormality constraint $P_t P_t^T = I$, then the Eq. (7) is minimized. Next we got a solution by performing an eigenvalue decomposition of the auto-correlation matrix

$$C_t = \sum_{n} u_t(n) u_t(n)^T \tag{7}$$

We set the rows of P_t to the $\overline{d}$ eigenvectors of C_t corresponding to the largest eigenvalues.

The filter is updated using the compressed training sample $\widetilde{F}_t = \mathcal{F}\{P_t f_t\}$ and compressed target template $\widetilde{U_t} = \mathcal{F}\{P_t u_t\}$ as

$$\widetilde{A}_t^l = \overline{G}\widetilde{U_t^l}, l = 1, \ldots, \widetilde{d} \tag{8}$$

$$\widetilde{B}_t = (1-\eta)\widetilde{B}_{t-1} + \eta \sum_{k=1}^{\widetilde{d}} \overline{\widetilde{F}_t^k} \widetilde{F}_t^k \tag{9}$$

Now, we can do less number of FFT to reduce the computational cost.

The correlation scores at the test sample z_t at the test sample z_t are obtained similarly to Eq. (3) by applying the filter on the compressed sample $\widetilde{Z}_t = \mathcal{F}\{P_{t-1} z_t\}$,

$$Y_t = \frac{\sum_{l=1}^{\bar{d}} \overline{\widetilde{A}_{t-1}^l} \widetilde{Z}_t^l}{\widetilde{B}_{t-1} + \lambda} \tag{10}$$

3.3 Online Detector

A robust tracking algorithm requires a detection module to recover the target from potential tracking failures caused by occlusion. For each tracked target z, we use $y_t = ifft\{Y_t\}$ to be the confidence score. When the confidence score is below a pre-defined threshold, then the algorithm start up the online detector and finds the target back. In our approach, we use an online SVM classifier as the detector, and then we train the SVM classifier by drawing dense training samples around the estimated position. Given a training set $\{(v_i, c_i) | i = 1, 2, \ldots, N\}$ with N samples in a frame, the V_i denotes the feature vector generated by the $i-th$ sample and $c_i \in \{+1, -1\}$ is the class label. In the end, we get the objective function,

$$\min_h \frac{\lambda}{2} ||h||^2 + \frac{1}{N} \sum_i \ell(h; (v_i, c_i)) \tag{11}$$

Where $\ell(h; (v, c)) = \max\{0, 1 - c\langle h, v\rangle\}$. The notation $\langle h, v\rangle$ indicates the inner product between h and v. And we apply the passive-aggressive algorithm [12] to update the hyper plane parameters efficiently.

$$h \leftarrow h - \frac{\ell(h; (v, c))}{||\nabla_h \ell(h; (v, c))||^2 + \frac{1}{2\tau}} \nabla_h \ell(h; (v, c)) \tag{12}$$

Where $\nabla_h \ell(h; (v, c))$ is the gradient of the loss function in terms of h and $\tau \in (0, +\infty)$ is a hyper-parameter that controls the update rate of h.

Algorithm: Proposed tracking algorithm: iterate at frame t

Input:

Image It

Previous target position p_{t-1} and scale s_{t-1}

Translation model $A_{t-1,trans}$ and $B_{t-1,trans}$

Scale model $A_{t-1,scale}$ and $B_{t-1,scale}$

For t= 2:T

Max response=max(response);

If Max response< threshold

Activate detection module.

End

End

Until End of the video sequence

Output:

Estimated target position p_t and scale s_t

Updated translation mode $A_{t,trans}$, $B_{t,trans}$

Updated scale model $A_{t,scale}$, $B_{t,trans}$

4 Experiment

4.1 Experimental Settings

Datasets. We evaluate the proposed algorithm on a large benchmark dataset [13] that contains 100 videos.

Desktop Configuration. We implement our tracker in MATLAB on an ADM Ryzen-5 3.20 GHz CPU with 16 GB RAM, and use the Vlfeat toolbox.

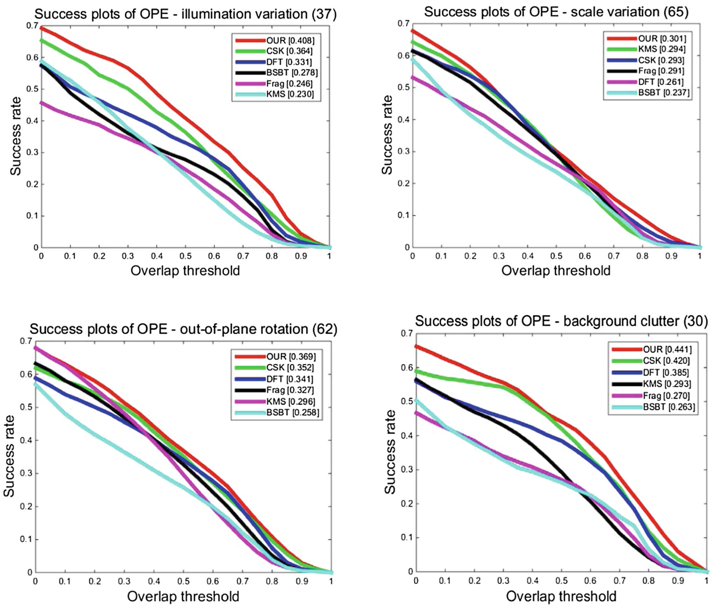

Fig. 1. Overlap success plots over four tracking challenges

Fig. 2. An experiment to track long-term object. The first and second columns prove that our method can effectively recover the target lost due to occlusion and deformation, and the third and fourth columns prove the robustness of our method when tracking the long-term target.

4.2 Comparisons

We compared the proposed algorithm with conventional and state-of-the-art algorithms which are available as source code. They are CSK [14], KMS [15], DFT [16], Frag [17] and BSBT [18]. The parameters for those algorithms are set default.

Figure 1 shows that the tracking results on OTB-100. The results show that our algorithm could track the object accurately. And our algorithm performs better than those Start-of-the-art trackers.

In addition, we tested the robustness of our method in tracking long-term targets. Figure 2 proved that the method we proposed could be a good way to track and reposition the lost object.

In the experiment, our algorithm can reach the speed of 60FPSD. This proves that our method of dimensionality reduction is effective. From Table 1, we could see that our algorithm has a high score than others in DP, OS and CLE. The results show that our algorithm performs better than others.

Table 1. The score of 6 trackers in 6 sequences. (The best results are in italics)

Sequence name		Bird2	Car2	CarDark	Coupon	CarScale	David2
Our tracker	OS (%)	*0.980*	*1.000*	*1.000*	*1.000*	*0.813*	*1.000*
	DP (%)	*0.980*	*1.000*	0.802	*1.000*	*0.516*	0.994
	CLE(%)	*9.250*	*1.370*	*4.090*	*2.470*	*13.54*	3.26
CSK	OS (%)	0.485	*1.000*	*1.000*	*1.000*	0.651	*1.000*
	DP (%)	0.485	0.797	*0.992*	*1.000*	0.448	*1.000*
	CLE (%)	31.12	10.36	3.23	6.14	83.01	*2.33*
KMS	OS (%)	0.253	0.074	0.275	0.107	0.536	0.337
	DP (%)	0.293	0.074	0.595	0.086	0.298	0.095
	CLE (%)	31.51	121.44	34.12	109.29	39.11	45.53
DFT	OS (%)	0.717	0.278	0.649	1.000	0.647	0.767
	DP (%)	0.707	0.278	0.514	1.000	0.448	0.767
	CLE (%)	27.03	100.64	29.39	5.77	79.95	15.02
BSBT	OS (%)	0.061	0.377	0.700	0.462	0.131	0.156
	DP (%)	0.061	0.377	0.700	0.462	0.131	0.156
	CLE (%)	324.54	129.11	61.69	59.23	220.95	159.44
Frag	OS (%)	0.152	0.380	0.081	0.141	0.631	0.758
	DP (%)	0.141	0.380	0.115	0.141	0.317	0.825
	CLE (%)	66.90	88.10	79.64	72.26	31.07	15.74

5 Conclusion

In this paper, in order to accelerate the speed of the long term object tracking, we use the dimensionality reduction method to reduce the number of FFTs and increase the target tracking speed to 60 FPS. At the same time, we use multi-channel discriminative correlation filters and give this filter a set of sample patches of the target appearance to ensure tracking robustness. In addition, we use an online training SVM to re-track the lost targets. This method successfully solves the problem that the accuracy and speed cannot be maintained simultaneously in long-term object tracking through dimensionally reduction and online SVM detector. Extensive experiment results on a large-scale benchmark demonstrate that our method could track the long-term object at real time and accurately.

References

1. Yilmaz, A., Javed, O., Shah, M.: Object tracking: a survey. ACM Comput. Surv. **38**(4), 13 (2006)
2. Danelljan, M., Häger, G., Khan, F.S., et al.: Discriminative scale space tracking. IEEE Trans. Pattern Anal. Mach. Intell. **39**(8), 1561–1575 (2017)
3. Ma, C., Huang, J.B., Yang, X., et al.: Adaptive correlation filters with long-term and short-term memory for object tracking. Int. J. Comput. Vis. **126**, 771–796 (2018)
4. Bolme, D.S., Beveridge, J.R., Draper, B.A., et al.: Visual object tracking using adaptive correlation filters. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2544–2550. IEEE (2010)
5. Henriques, J.F., Caseiro, R., Martins, P., Batista, J.: Exploiting the circulant structure of tracking-by-detection with kernels. In: Proceedings of the European Conference on Computer Vision (2012)
6. Galoogahi, H.K., Sim, T., Lucey, S.: Multi-channel correlation filters. In: Proceedings of IEEE International Conference on Computer Vision (2013)
7. Henriques, J.F., Caseiro, R., Martins, P., et al.: High-speed tracking with kernelized correlation filters. IEEE Trans. Pattern Anal. Mach. Intell. **37**(3), 583–596 (2015)
8. Danelljan, M., Häger, G., Khan, F.S., Felsberg, M.: Learning spatially regularized correlation filters for visual tracking. In: Proceedings of IEEE International Conference on Computer Vision (2015)
9. Kalal, Z., Mikolajczyk, K., Matas, J.: Tracking-learning-detection. IEEE Trans. Pattern Anal. Mach. Intell. **34**(7), 1409 (2012)
10. Supancic, J.S., Ramanan, D.: Self-paced learning for long-term tracking. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2013)
11. Kang, W., Li, X., Li, S., Liu, G.: Corrected continuous correlation filter for long-term tracking. IEEE Access **6**, 11959–11969 (2018)
12. Crammer, K., Dekel, O., Keshet, J., Shalev-Shwartz, S., Singer, Y.: Online passive-aggressive algorithms. J. Mach. Learn. Res. **7**, 551–585 (2006)
13. Wu, Y., Lim, J., Yang, M.H.: Object tracking benchmark. TPAMI (2015)
14. Henriques, J.F., Caseiro, R., Martins, P., Batista, J.: Exploiting the circulant structure of tracking-by-detection with kernels. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7575, pp. 702–715. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33765-9_50
15. Comaniciu, D., Ramesh, V., Meer, P.: Kernel-based object tracking. PAMI **25**(5), 564–577 (2003)
16. Fragkiadaki, K., Shi, J.: Detection free tracking: exploiting motion and topology for segmenting and tracking under entanglement. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2073–2080. IEEE (2011)
17. Adam, A., Rivlin, E., Shimshoni, I.: Robust fragments-based tracking using the integral histogram. In: CVPR (2006)
18. Stalder, S., Grabner, H., Van Gool, L.: Beyond semi-supervised tracking: tracking should be as simple as detection, but not simpler than recognition. In: ICCV Workshops, vol. 3 (2009)