Learning Spatial-Temporal Regularized Correlation Filters for Visual Tracking

Feng Li[1], Cheng Tian[1], Wangmeng Zuo *[1], Lei Zhang[2], and Ming-Hsuan Yang[3]

[1]School of Computer Science and Technology, Harbin Institute of Technology, China
[2]Department of Computing, The Hong Kong Polytechnic University, China
[3]School of Engineering, University of California, Merced, USA
fengli_hit@hotmail.com, tcoperator@163.com, wmzuo@hit.edu.cn, csdzhang@comp.polyu.edu.hk, mhyang@ucmerced.edu

In the supplementary material, we first present the results of all 11 video attributes on the OTB-2015 dataset, then provide a comparison of temporal CF variation against frames between SRDCF and our STRCF on more video sequences. Finally, we show some visualization results of different trackers on several video sequences.

1. Evaluation Based on Video attributes

For fair comparisons, we compare our STRCF with 11 state-of-the-art trackers with hand-crafted features, *i.e.* SRDCF [6], BACF [8], ECO-HC [3], SRDCFDecon [4], Staple [1], Staple+CA[10], SAMF+AT [2], DSST [5], SAMF [9], MEEM [11] and KCF [7]. Fig. 1 shows that our STRCF performs favorably against the other trackers including SRDCF on all attributes and ranks the second place among all competing trackers on mean success overlap plots.

2. Analysis on Temporal CF Variations between SRDCF and STRCF

To highlight the differences of SRDCF and STRCF on model learning, here we visualize the temporal CF variation (*i.e.* $\frac{\|\mathbf{f}_t - \mathbf{f}_{t-1}\|^2}{z}$, where z is the normalization factor) against frames on more video sequences. From Fig. 2, we can draw the following conclusions: (i) Benefited from online PA algorithm, the temporal CF variation of STRCF is much smaller than SRDCF in the first few frames. This is helpful when the target suffers from significant appearance variations in the beginning, such as the deformation changes in Fig. 2a and 2b. In these two cases, our STRCF can successfully follow the target while SRDCF fails to track it. (ii) STRCF can provide a more robust CF model (*i.e.* smaller temporal CF variations) than SRDCF by *passively* updating the CFs in most frames. (iii) Compared to SRDCF, our STRCF is insensitive to various slow appearance variations (*i.e.* the out-of-plane rotation in Fig. 2c and 2d, and the cluttered background in Fig. 2e). Besides, it can also adapt to the sudden appearance variations (*i.e.* target re-appearance in Fig. 2f) by *aggressively* updating the CFs.

In this work, we show that with the introduction of the temporal regularization, STRCF can provide a more robust appearance model than SRDCF, thus leading to better performance.

*Corresponding author.

3. Qualitative Evaluation

We also perform qualitative evaluation of different trackers on several video sequences. For clearer visualization, we show the results of STRCF and 4 state-of-the-art trackers based on hand-crafted features, including ECO-HC [3], BACF [8], SRDCF [6] and SRDCFDecon [4]. The tracking results on on 6 video sequences are shown in Fig. 3.

References

[1] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, and P. Torr. Staple: Complementary learners for real-time tracking. In *CVPR*, 2016. 1

[2] A. Bibi, M. Mueller, and B. Ghanem. Target response adaptation for correlation filter tracking. In *ECCV*, 2016. 1

[3] M. Danelljan, G. Bhat, F. S. Khan, and M. Felsberg. Eco: Efficient convolution operators for tracking. In *CVPR*, 2017. 1

[4] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg. Adaptive decontamination of the training set: A unified formulation for discriminative visual tracking. In *CVPR*, 2016. 1

[5] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg. Discriminative scale space tracking. *TPAMI*, PP(99):1–1, 2016. 1

[6] M. Danelljan, G. Hager, F. Shahbaz Khan, and M. Felsberg. Learning spatially regularized correlation filters for visual tracking. In *ICCV*, 2015. 1

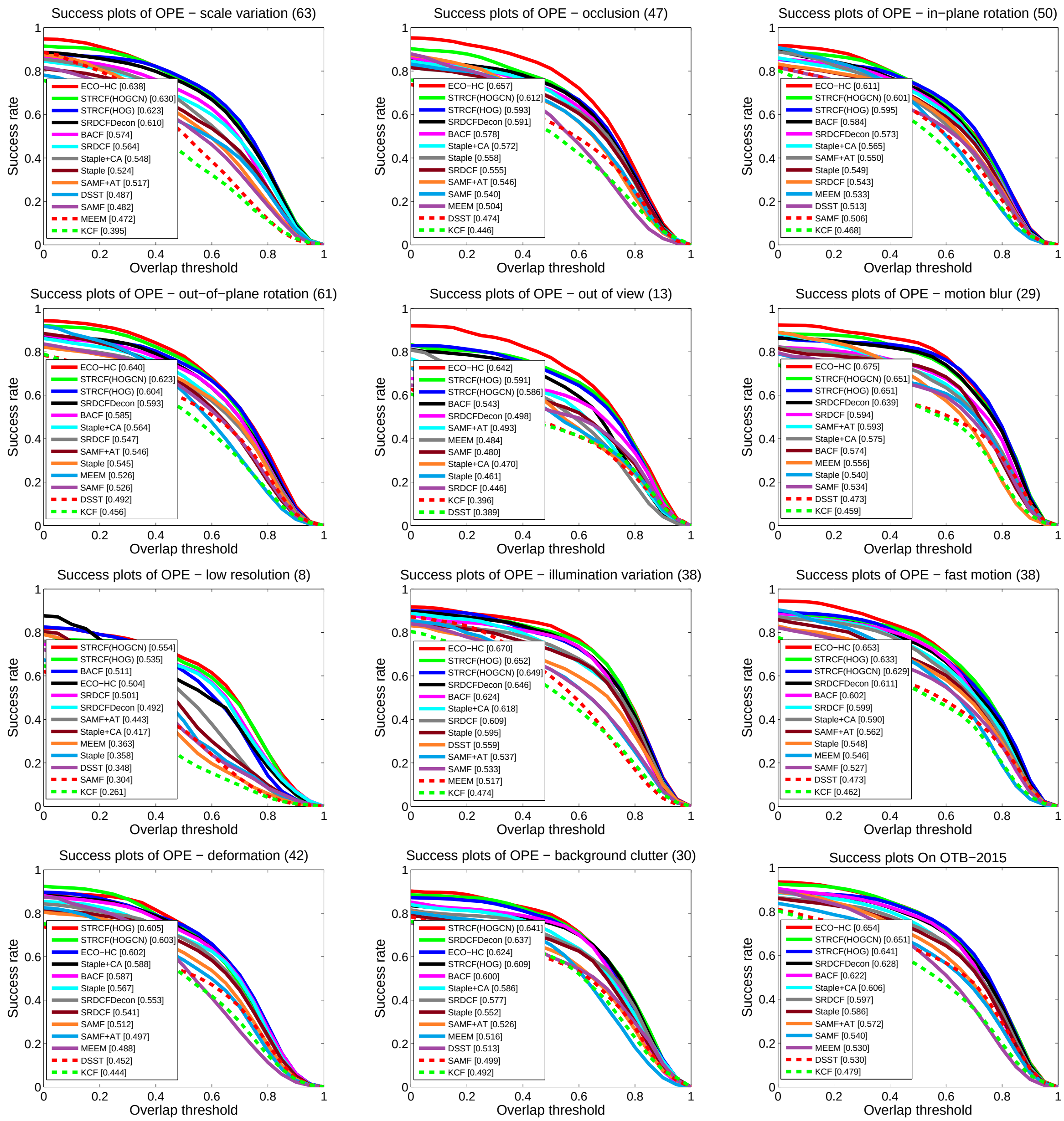


Figure 1: Overlap success plots of the competing trackers under all attributes on the OTB-2015 dataset.

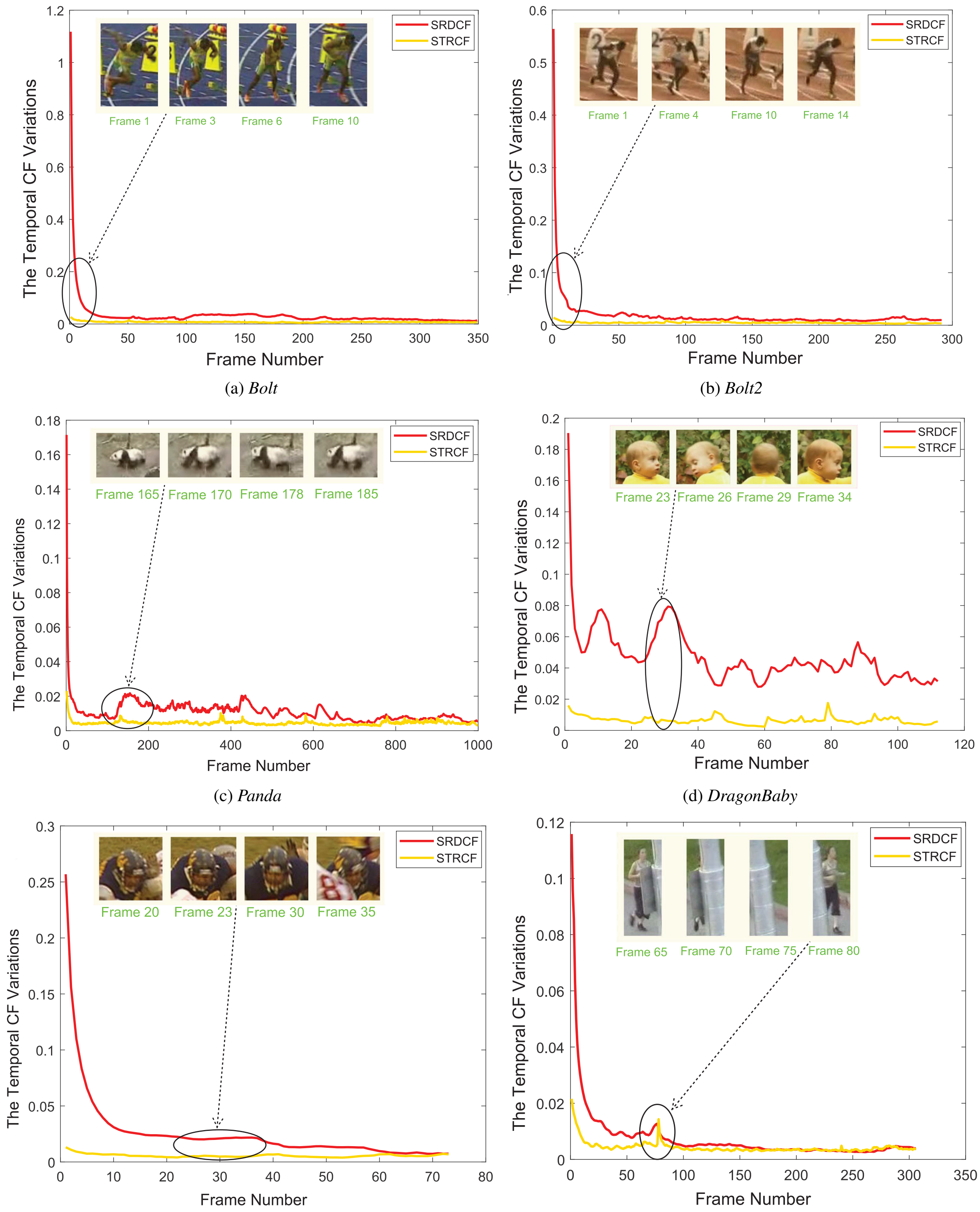
[7] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista. High-speed tracking with kernelized correlation filters. *TPAMI*, 37(3):583–596, 2015. 1

[8] H. Kiani Galoogahi, A. Fagg, and S. Lucey. Learning background-aware correlation filters for visual tracking. In *ICCV*, 2017. 1

[9] Y. Li and J. Zhu. A scale adaptive kernel correlation filter tracker with feature integration. In *ECCVW*, 2014. 1

[10] M. Mueller, N. Smith, and B. Ghanem. Context-aware correlation filter tracking. In *CVPR*, 2017. 1

[11] J. Zhang, S. Ma, and S. Sclaroff. Meem: robust tracking via multiple experts using entropy minimization. In *ECCV*, 2014. 1



(a) *Bolt* (b) *Bolt2*

(c) *Panda* (d) *DragonBaby*

(e) *Football1* (f) *Jogging*

Figure 2: Comparison of the temporal CF variation against frames between SRDCF and our STRCF on 6 video sequences (*i.e. Bolt*, *Bolt2*, *Panda*, *Football1*, *DragonBaby* and *Jogging*).

Figure 3: Qualitative evaluation on 6 video sequences (*i.e*. *CarScale*, *Dog*, *Girl2*, *Human3*, *Panda* and *Trans*). We show the results of STRCF, ECO-HC, BACF, SRDCF and SRDCFDecon with different colors, respectively.