LONG-TERM OBJECT TRACKING BASED ON SIAMESE NETWORK

Kaiheng Dai, Yuehuan Wang, Xiaoyun Yan

The School of Automation, Huazhong University of Science and Technology,Wuhan, China

Introduction

◦ **Motivation**

--Current Siamese network trackers only use the initial target patch in the first frame to track the best candidate in a new frame, which may make the tracker fail to tracked target when the appearance change drastically.

--Multi-template fusion scheme should inherit the good performance of Siamese network tracker, especially target re-detection and real-time tracking.

◦ **Our work**

--We propose a multi-template fusion scheme which introduce a stable template without update, a soft template with a thin update, and an adaptive template with a large update; then fuse the results of those templates together with a simple but effective cascade architecture (fig 2).

--We propose patch template update scheme based on optical flow to make the template update more credible.

Method

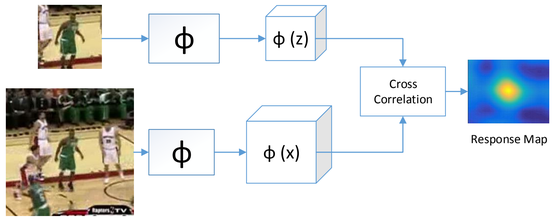


Fig 1 The Baseline Method[1]

◦ **multi-template cascade fusion scheme**

Step 1: Use the stable .template to compute the respond map, if the result within the threshold, output the result, otherwise, step 2.

Step 2:Use the stable template and soft stable template to compute the fusion respond map, if the result within the threshold, output the result, otherwise, step 3..

Step 3:Use the stable template, soft stable template and adaptive template to fusion tracking.

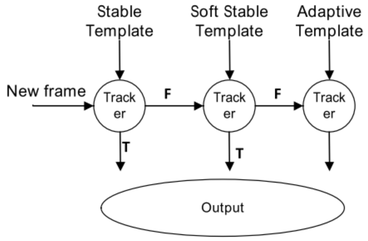


Fig 2 The cascade architecture of fusion tracking with three templates.

◦**Baseline Siamese CNNs Architecture**

Inputs	Conv Block	Max Pool	Conv Block	Max Pool	Conv Only	Conv Only	Conv Only
Z:63×63×3 X:127×127×3	5×5×3×96	3×3	5×5×48×256	3×3	3×3×128×384	3×3×192×384	3×3×192×256

patch template update mechanism

$$T_j^{i+1} = (1 - \alpha\pi_j)T_j^i + \alpha\pi_j\phi(y_j^i)$$

where α is a update weighting factor, T_j^i denotes the j-th patch template at i-th frame,$\phi(y_j^i)$ is the optimal candidate,, $\pi_j = O^i/O^t$where $O^t$is the number of given pixels covered by the target box in template frame, then estimated optical flow in optimal candidate found in the current frame, O^i is the number of those pixels in the optimal candidate.

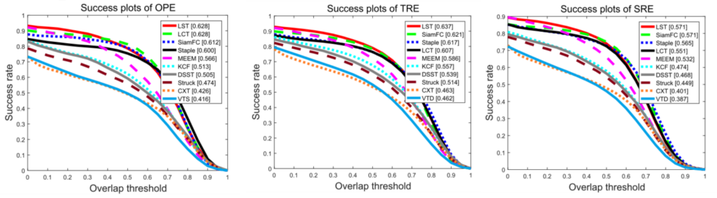
References

[1].Luca Bertinetto, Jack Valmadre, Joao F. Henriques, An- dreaVedaldi, and Philip H. S. Torr, “Fully-convolutional siamese networks for object tracking,” in Proc. ECCV. 2016, Springer International Publishing.

[2].Huchuan Lu, Shipeng Lu, Dong Wang, Shu Wang, and Henry Leung, “Pixel-wise spatial pyramid-based hybrid tracking,” IEEE Transactions on Circuits and System- s for Video Technology, vol. 22, no. 9, pp. 1365–1376, 2012.

Experiment Result

◦**Experiment results‘on OTB2013’**



Conclusion

The extensive empirical evaluations on tracking benchmark OTB2013 demonstrate that the proposed method not only can inherit the good performances of SiamFc tracker in re-detection and common appearance changes but also performs well in heavy appearance variations where SiamFc may fail. Moreover, it run fast because the appearance variations are so slight in most of tracking time that only one stable template can handle well.