MUlti-Store Tracker (MUSTer): a Cognitive Psychology Inspired Approach to Object Tracking

Zhibin Hong[1], Zhe Chen[1], Chaohui Wang[2], Xue Mei[2], Danil Prokhorov[3], Dacheng Tao[1],
[1]Centre for Quantum Computation and Intelligent Systems, University of Technology, Sydney. [2]Laboratoire d'Informatique Gaspard Monge, Université Paris-Est. [3]Toyota Research Institute

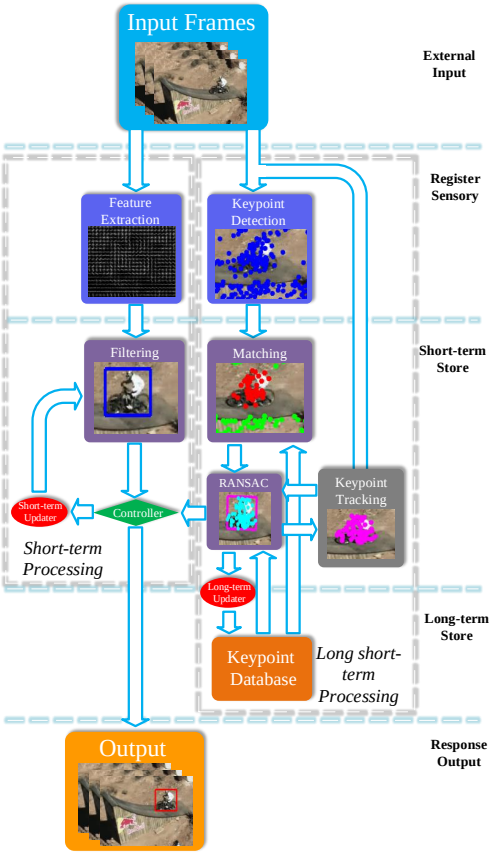

Figure 1: A system flowchart of the proposed tracker based on the Atkinson-Shiffrin Memory Model.

The online tracking research community have developed a number of trackers. Some [2] are highly sensitive and accurate in the short term, while others are relatively conservative but robust over the long term (e.g., [3]). In other words, some trackers can be regarded as short-term systems while others can be regarded as long-term systems. The power of the Atkinson-Shiffrin Memory Model (ASMM) to track objects by co-operation between the long- and short-term memory stores has motivated us to design a tracker that integrates a short- and long-term system to boost tracking performance.

In this paper, we propose the MUlti-Store Tracker (MUSTer) based on the ASMM. A system flowchart of MUSTer is shown in Figure 1. MUSTer consists of one short-term store and one long-term store that collaboratively process the image input and track the target. An Integrated Correlation Filter (ICF) is employed in the short-term store to perform short-term processing and track the target based on short-term memory and spatiotemporal consistency. This component generally works accurately and efficiently in relatively stable scenarios. In addition, another relatively conservative long-term component based on keypoint matching-tracking and RANSAC estimation is introduced to conduct the long short-term processing on the fly. This interacts with the short-term memory stored in an active set of keypoints using forward-backward tracking, and it also retrieves the long-term memory for matching and updates the long-term memory based on the RANSAC estimation results and the forgetting curve. During tracking, the outputs of both the short-term and long short-term processing are sent to a controller, which decides the final MUSTer output and the ICF update. Specifically, the short-term memory in ICF is reset when the short-term processing output is highly inconsistent with the long-term memory encoded by the output of the long short-term processing. This enables the recovery of the short-term tracking after dramatic appearance changes such as severe occlusion, the object leaving field-of-view, or rotation.

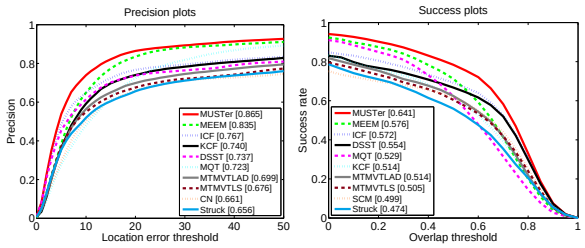

Figure 2: Quantitative comparison in CVPR2013 OOTB. The performance score for each tracker is shown in the legend. For each figure, only the top 10 trackers are presented.

The short-term component is used to provide instant responses to the image input based on short-term memory. For accurate and efficient short-term processing performance, we employ Integrated Correlation Filters (ICFs), which are based on the Kernelized Correlation Filters (KCFs) [2] and the Discriminative Scale Space Correlation Filter (DSSCF) [1]. The ICF is a two-stage filtering process that performs translation estimation and scale estimation. The long-term memory of the target appearance is modeled by a total feature database $\mathcal{M} = \mathcal{T} \cup \mathcal{B}$ that consists of a foreground (target) feature database $\mathcal{T}$ and a background feature database $\mathcal{B}$:

$$\mathcal{T} = \{(\mathbf{d}_i, \mathbf{p}_i^o)\}_{i=1}^{N_\mathcal{T}}, \quad \mathcal{B} = \{\mathbf{d}_i\}_{i=1}^{N_\mathcal{B}} . \tag{1}$$

Here, $\mathbf{d}_i \in \mathcal{R}^{128}$ is the 128-dimensional Scale-invariant Feature Transform (SIFT) descriptors of the keypoints. $N_\mathcal{T}$ and $N_\mathcal{B}$ are the respective numbers of descriptors.

The proposed tracker was implemented using Matlab & C++ with OpenCv library. The average time cost on CVPR2013 Online Object Tracking Benchmark (OOTB) [5] is 0.287s/frame on a cluster node (3.4GHz, 8 cores, 32GB RAM). We report in the paper the evalutions on OOTB [5] and ALOV++ (Amsterdam Library of Ordinary Videos) dataset [4] by comparing MUSTer with a number of state-of-the-art trackers. The results on CVPR2013 OOBT is shown in Fig. 2. The experimental results on two large datasets demonstrate that the proposed tracker is capable of taking advantage of both the short-term and long-term systems and boosting the tracking performance.

[1] Martin Danelljan, Gustav Häger, Fahad Shahbaz Khan, and Michael Felsberg. Accurate scale estimation for robust visual tracking. In *BMVC*, 2014.

[2] J.F. Henriques, R. Caseiro, P. Martins, and J. Batista. High-speed tracking with kernelized correlation filters. *TPAMI*, pages 583–596, 2015.

[3] F. Pernici and A. Del Bimbo. Object tracking by oversampling local features. *TPAMI*, 36(12):2538–2551, 2014.

[4] AW.M. Smeulders, D.M. Chu, R. Cucchiara, S. Calderara, A Dehghan, and M. Shah. Visual tracking: An experimental survey. *TPAMI*, 36(7): 1442–1468, 2014.

[5] Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang. Online object tracking: A benchmark. In *CVPR*, pages 2411–2418, 2013.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

The code to reproduce the experiments is available on https://sites.google.com/site/multistoretrackermuster/