End-to-end Flow Correlation Tracking with Spatial-temporal Attention (Supplementary Material)

Zheng Zhu[1,2], Wei Wu[3], Wei Zou[1,2,4], Junjie Yan[3]
[1]Institute of Automation, Chinese Academy of Sciences, Beijing, China
[2]University of Chinese Academy of Sciences, Beijing, China
[3]SenseTime Group Limited, Beijing, China
[4]TianJin Intelligent Tech.Institute of CASIA Co.,Ltd, Tianjin, China
{zhuzheng2014,wei.zou}@ia.ac.cn {wuwei,yanjunjie}@sensetime.com

1. Selection of aggregated frame numbers

Here, we show the performance of OTB2015 when different aggregated frame numbers T are adopted in training and tracking. The performance is evaluated by area under curve (AUC) of success plot. As shown in Table 1, the accuracy improves as more frames are aggregated. The improvement saturates at 6 frames. So we keep aggregated frame numbers $T = 6$ in training and tracking.

Table 1: AUC of OTB2015 with different aggregated frame numbers

value of T	2	3	4	5	6	7	8
AUC	0.587	0.613	0.621	0.637	**0.655**	0.651	0.654

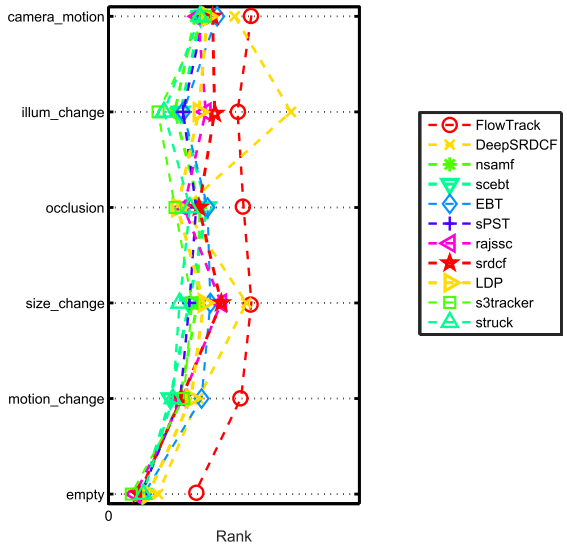

Figure 1: EAO scores for each attribute on the VOT2015 dataset. *Empty* denotes frames with no labeled attribute. Best viewed on color display.

2. Additional results with attributes on VOT2015

Here, further experimental evaluation on the VOT2015 dataset with 60 videos are presented. In the VOT2015 dataset, each frame is labeled with 5 different attributes: camera motion, illumination change, occlusion, size change and motion change. The performance is evaluated by expected average overlap (EAO) measure. As shown in Figure 1, our approach (FlowTrack) achieves the best results on 4 attributes.

3. Additional results with attributes on VOT2016

In this section, we provide further experimental evaluation on the VOT2016 dataset with 60 videos. In the VOT2016 dataset, each frame is labeled with five different attributes: camera motion, illumination change, occlusion, size change and motion change. Figure 2 visualizes the EAO of each attribute individually. Our approach (FlowTrack) ranks *1st* in 4 attributes and *2nd* in 1 attributes.

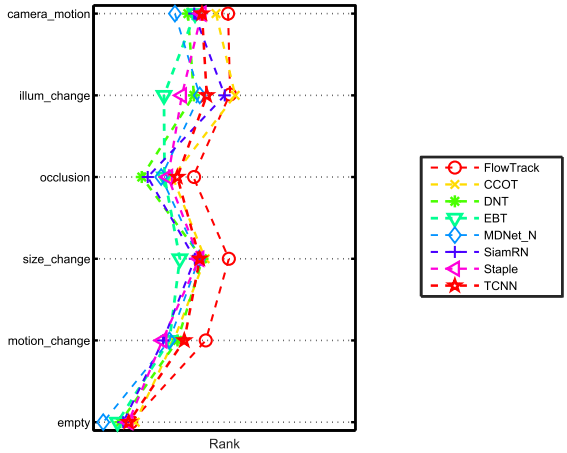

Figure 2: EAO scores for each attribute on the VOT2016 dataset. *Empty* denotes frames with no labeled attribute. Best viewed on color display.