TCTrack: Temporal Contexts for Aerial Tracking

Ziang Cao[1], Ziyuan Huang[2], Liang Pan[3], Shiwei Zhang[4], Ziwei Liu[3], Changhong Fu[1,*]
[1]Tongji University [2]National University of Singapore [3]S-Lab, Nanyang Technological University
[4]DAMO Academy, Alibaba Group
caoang233@gmail.com, ziyuan.huang@u.nus.edu, {liang.pan, ziwei.liu}@ntu.edu.sg
zhangjin.zsw@alibaba-inc.com changhongfu@tongji.edu.cn

1. Overview

This file aims to provide supplementary materials for our comprehensive framework, *i.e.*, TCTrack, including evaluation metric in Sec. 2, loss function in Sec. 3, and more results 4.

2. Evaluation criteria

In this paper, we adopt the one-pass evaluation (OPE) [5] criteria to evaluate the performance, *i.e.*, precision and success rate. The former is defined by the center location error (CLE) between ground truth and predicted results while the latter is calculated by Intersection over Union (IoU) score. They are reported in precision and success plot (SP). Additionally, the area-under-the-cure (AUC) on the SP is used to rank all trackers.

3. Loss functions

After obtaining the refined features $\mathbf{F}_t^*$, the tracking results can be calculated via a standard classification and regression network. To improve the precision of regression, we design a new branch named $\mathcal{L}_{loc1}$. For clarification, we denote the center and scale of ground truth bounding box as x, y, w, h while the prediction results as $\mathbf{X}_{pred}, \mathbf{Y}_{pred} \in \mathbb{R}^{N\times N}$. Therefore, the $\mathcal{L}_{loc1}$ can be formulated as follows:

$$\begin{aligned} \mathcal{L}_{loc1} &= \sum_{i=1}^{N}\sum_{j=1}^{N} M(i,j) * dis(i,j) \\ dis(i,j) &= \sqrt{\frac{(X_{pred}(i,j)-x)^2}{w} + \frac{(Y(i,j)_{pred}-y)^2}{h}} \end{aligned}, \tag{1}$$

where $\mathbf{M} \in \mathbb{R}^{N\times N}$ is a mask, following [1].

Therefore, the overall loss function can be summarized as follows:

$$\mathcal{L} = \mathcal{L}_{cls1} + \mathcal{L}_{cls2} + \mathcal{L}_{loc1} + \mathcal{L}_{loc2}\ , \tag{2}$$

where $\mathcal{L}_{cls1}$, $\mathcal{L}_{cls2}$, and $\mathcal{L}_{loc1}$ represent the cross-entropy, binary cross-entropy, and IoU loss, which are similar with [1].

4. More results

More detailed evaluations on four well-known aerial tracking benchmarks are presented in Table 1, Fig 2, Fig 3, and Fig 4. Furthermore, the qualitative results compared with other SOTA trackers are shown in Fig. 1. The exhaustive experiments strongly demonstrate the impressive performance of our temporal framework in aerial tracking in terms of short-term and long-term scenarios.

References

[1] Ziang Cao, Changhong Fu, Junjie Ye, Bowen Li, and Yiming Li. HiFT: Hierarchical Feature Transformer for Aerial Tracking. In *ICCV*, pages 15437–15446, 2021. 1

[2] Changhong Fu, Ziang Cao, Yiming Li, Junjie Ye, and Chen Feng. Onboard Real-Time Aerial Tracking With Efficient Siamese Anchor Proposal Network. *TGRS*, pages 1–13, 2021. 2

[3] Siyi Li and Dit-Yan Yeung. Visual Object Tracking for Unmanned Aerial Vehicles: A Benchmark and New Motion Models. In *AAAI*, pages 1–7, 2017. 3

[4] Matthias Mueller, Neil Smith, and Bernard Ghanem. A Benchmark and Simulator for UAV Tracking. In *ECCV*, pages 445–461, 2016. 4, 5

[5] Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang. Object Tracking Benchmark. *TPAMI*, 37(9):1834–1848, 2015. 1

*Corresponding author

Figure 1. Qualitative comparisons among TCTrack and other state-of-the-art trackers. Missing trackers are denoted as "x" using the corresponding color. It clearly shows that our tracker achieves superior performance in various challenging aerial scenarios, especially in fast motion, low resolution, severe occlusion, and camera motion. Best viewed on the screen with high-resolution.

Table 1. Attribute-based evaluations on UAVTrack112_L [2]. It shows the different components of our framework under various aerial-specific challenges. ARC, CM, FM, LR, OV, POC, SV, SOB, VC, and Overall represents aspect ratio change, camera motion, fast motion, low resolution, out of view, partial occlusion, scale variation, similar objects, viewpoint change, and overall performance. The best three performances are respectively highlighted with **red**, **green**, and **blue** color.

Trackers	ARC		CM		FM		LR		OV		POC		SV		SOB		VC		Overall	
	Prec.	Succ.	Prec.	Succ.	Prec.	Succ.	Prec.	Succ.	Prec.	Succ.	Prec.	Succ.	Prec.	Succ.	Prec.	Succ.	Prec.	Succ.	Prec.	Succ.
TCTrack (Ours)	**0.768**	**0.567**	**0.729**	**0.484**	**0.676**	**0.476**	**0.752**	**0.520**	**0.702**	**0.506**	**0.797**	**0.581**	**0.776**	**0.574**	0.743	**0.549**	**0.746**	**0.542**	**0.786**	**0.582**
HiFT	**0.712**	**0.528**	0.635	0.420	0.622	**0.446**	0.677	**0.478**	0.588	**0.417**	0.760	**0.557**	0.721	**0.541**	0.637	0.492	0.657	0.491	0.734	**0.551**
SiamAPN++	0.706	0.516	0.693	**0.447**	0.598	0.424	0.691	0.466	**0.697**	**0.499**	0.725	0.517	0.723	0.526	0.742	**0.543**	0.690	**0.501**	0.735	0.537
ARCF	0.609	0.359	0.625	0.361	0.510	0.300	0.618	0.348	0.402	0.214	0.633	0.374	0.629	0.386	0.668	0.416	0.622	0.389	0.640	0.399
ASRCF	0.703	0.407	0.626	0.294	0.622	0.370	**0.740**	0.403	**0.752**	**0.417**	0.762	0.432	**0.731**	0.434	**0.782**	0.442	0.675	0.411	**0.743**	0.446
AutoTrack	0.640	0.366	0.636	0.340	0.542	0.307	0.652	0.353	0.513	0.286	0.657	0.374	0.660	0.391	0.666	0.383	0.644	0.388	0.675	0.405
BACF	0.543	0.319	0.569	0.304	0.499	0.296	0.555	0.288	0.411	0.226	0.611	0.349	0.574	0.348	0.660	0.400	0.605	0.377	0.593	0.358
CCOT	0.643	0.372	0.692	0.332	0.618	0.362	0.711	0.384	0.563	0.329	0.691	0.389	0.681	0.406	0.692	0.419	**0.703**	0.439	0.691	0.422
COKCF	0.550	0.286	0.542	0.258	0.431	0.229	0.503	0.236	0.346	0.173	0.600	0.297	0.517	0.275	0.487	0.288	0.442	0.250	0.520	0.283
DaSiamRPN	0.693	0.455	0.680	0.397	**0.627**	0.390	0.732	0.454	0.552	0.349	**0.774**	0.512	0.718	0.468	**0.767**	0.505	0.677	0.438	0.729	0.479
DeepSTRCF	0.683	0.423	**0.715**	0.390	0.590	0.367	0.725	0.420	0.659	0.402	0.764	0.468	0.701	0.447	0.743	0.448	0.667	0.440	0.715	0.460
DSiam	0.459	0.269	0.520	0.267	0.432	0.253	0.512	0.288	0.366	0.220	0.510	0.293	0.498	0.306	0.539	0.320	0.497	0.311	0.512	0.321
ECO	0.641	0.386	0.660	0.331	0.605	0.374	0.694	0.399	0.569	0.354	0.705	0.404	0.672	0.420	0.689	0.407	0.668	0.437	0.684	0.436
fDSST	0.460	0.284	0.521	0.309	0.394	0.245	0.471	0.241	0.297	0.189	0.547	0.320	0.470	0.295	0.387	0.257	0.423	0.288	0.491	0.306
LUDT	0.554	0.329	0.589	0.295	0.551	0.327	0.610	0.351	0.483	0.302	0.628	0.370	0.573	0.353	0.533	0.303	0.544	0.347	0.592	0.369
LUDTplus	0.621	0.368	0.654	0.364	0.571	0.357	0.665	0.376	0.527	0.309	0.696	0.408	0.645	0.398	0.680	0.401	0.648	0.417	0.648	0.411
SiameseFC	0.647	0.405	0.683	0.372	0.568	0.357	0.690	0.416	0.616	0.369	0.712	0.450	0.676	0.437	0.697	0.443	0.644	0.416	0.690	0.452
SRDCF	0.460	0.278	0.501	0.260	0.468	0.279	0.462	0.242	0.331	0.179	0.527	0.304	0.486	0.307	0.437	0.268	0.480	0.315	0.508	0.320
STRCF	0.555	0.311	0.588	0.299	0.516	0.307	0.608	0.305	0.498	0.277	0.639	0.345	0.591	0.341	0.618	0.354	0.618	0.382	0.609	0.360
TADT	0.685	0.420	**0.755**	0.411	0.625	0.395	**0.741**	0.440	0.672	0.413	0.752	0.452	0.703	0.449	**0.763**	0.463	0.671	0.442	0.712	0.462
UDT	0.590	0.352	0.673	0.334	0.500	0.307	0.637	0.365	0.447	0.285	0.645	0.379	0.602	0.372	0.542	0.311	0.535	0.344	0.620	0.388
UDTplus	0.607	0.360	0.639	0.356	0.536	0.335	0.627	0.358	0.460	0.277	0.650	0.379	0.632	0.391	0.641	0.400	0.629	0.413	0.637	0.405
SiamRPN++	**0.741**	**0.531**	0.697	**0.444**	**0.667**	**0.451**	0.719	**0.486**	0.621	0.402	**0.773**	**0.562**	**0.763**	**0.548**	0.739	**0.539**	**0.745**	**0.520**	**0.773**	**0.559**
KCF	0.349	0.196	0.352	0.158	0.305	0.168	0.300	0.144	0.252	0.139	0.391	0.221	0.344	0.203	0.246	0.158	0.328	0.193	0.353	0.212

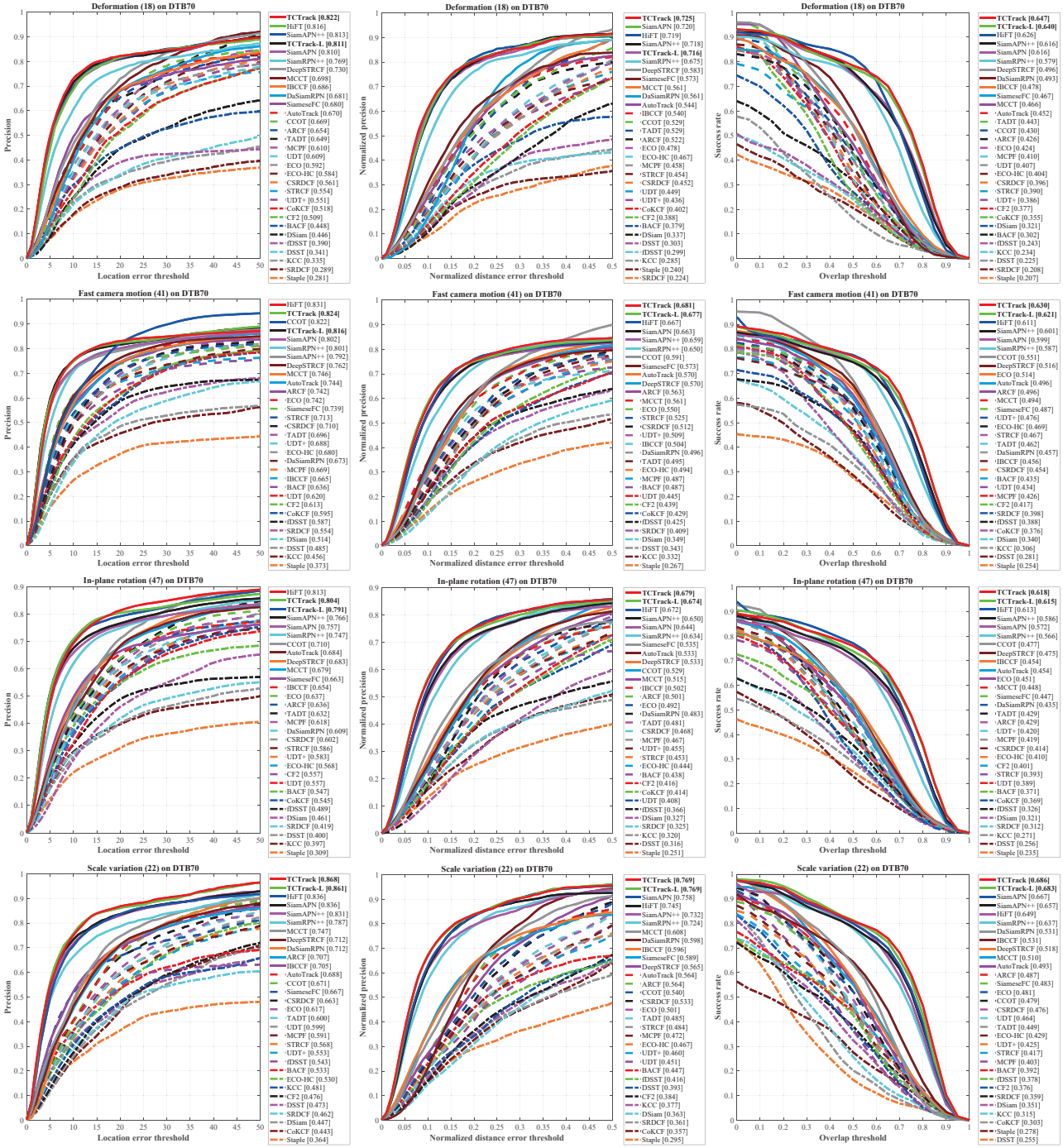

Figure 2. Attribute-based comparison among all trackers on DTB70 [3]. Our framework outperforms other trackers, especially in motion and scale variation conditions.

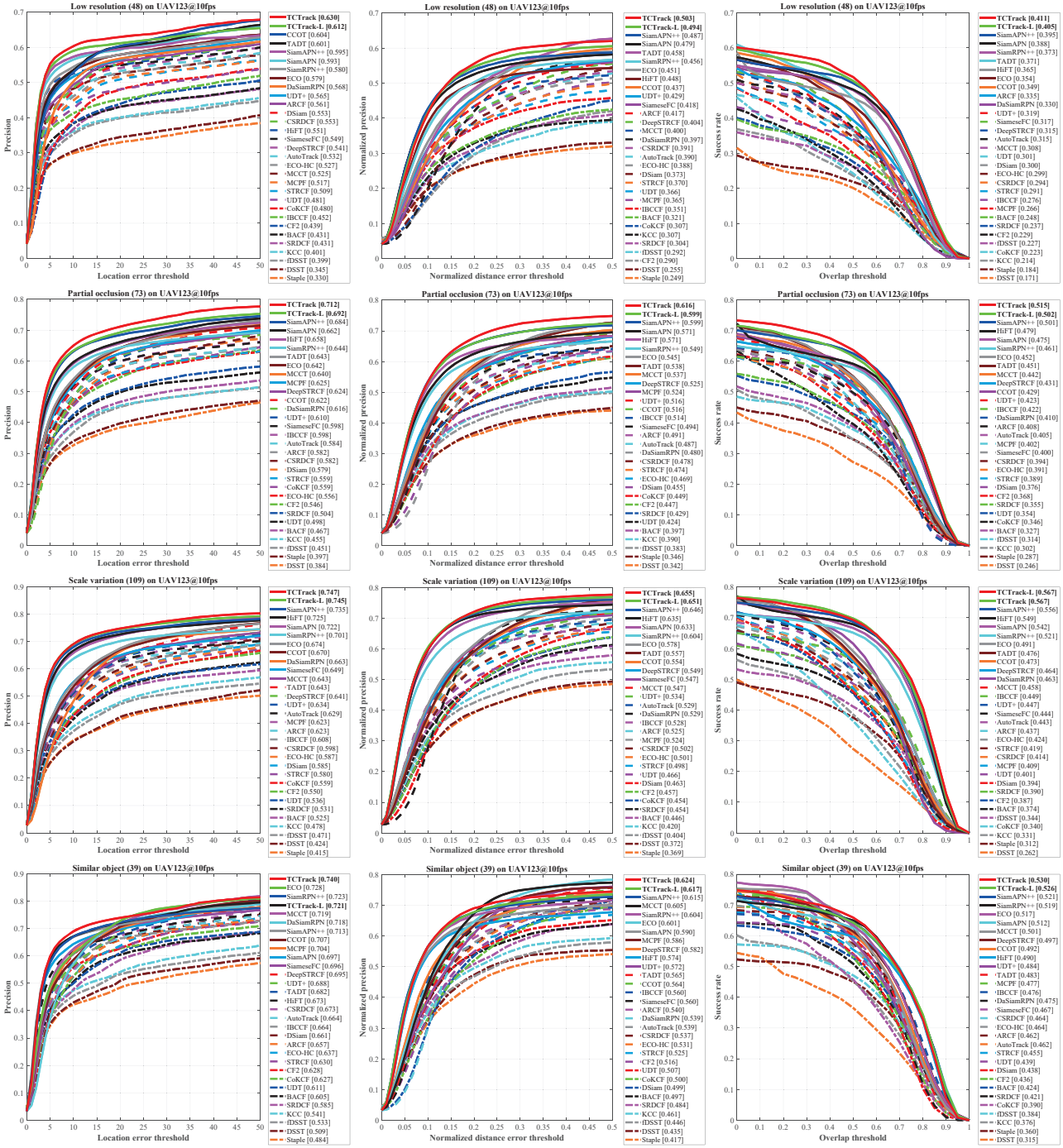

Figure 3. Attribute-based comparison among all trackers on UAV123@10fps [4]. Our framework achieves superior performance in various challenges.

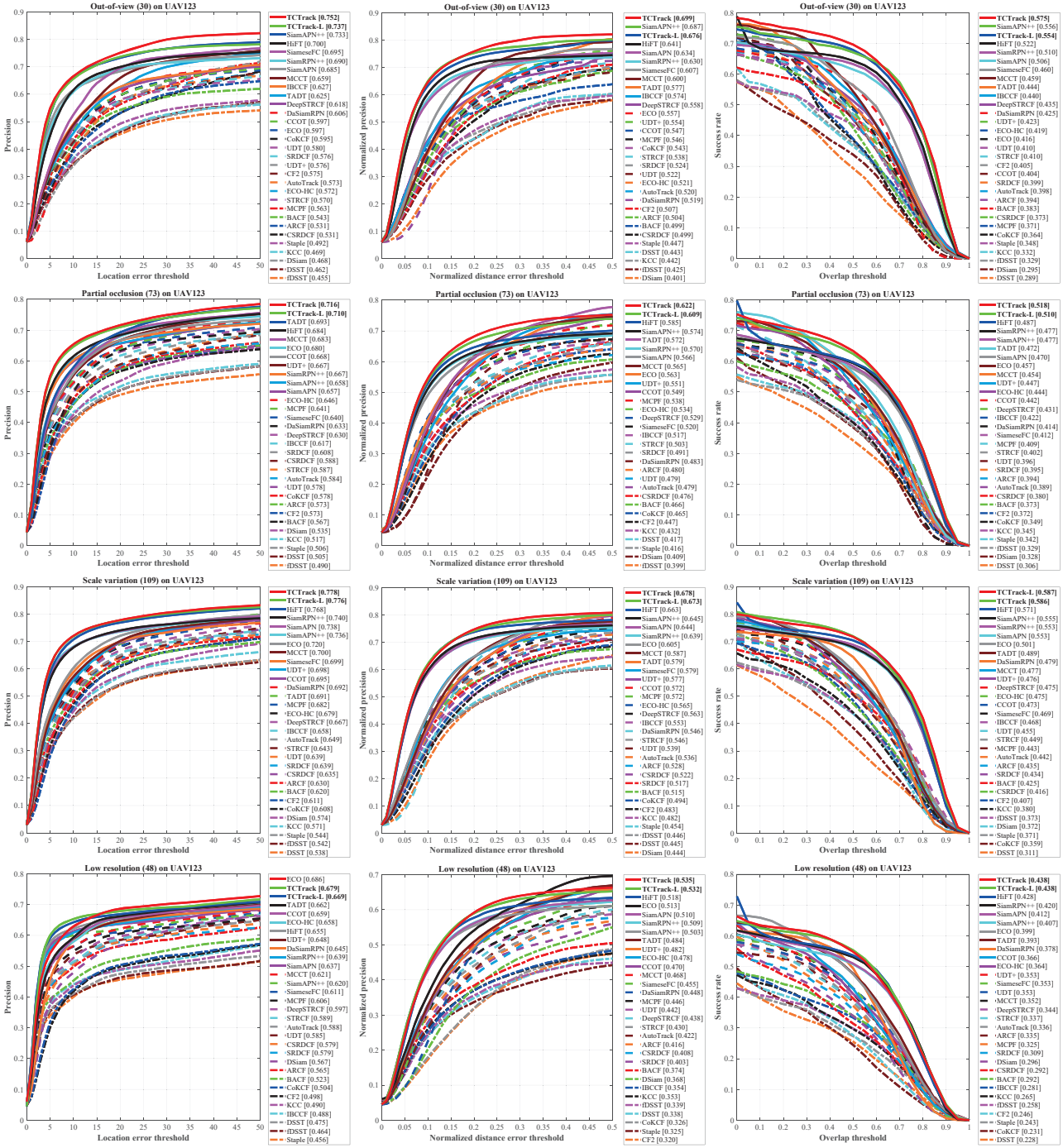

Figure 4. Attribute-based comparison among all trackers on UAV123 [4]. Attributing to our comprehensive framework, temporal contexts are effectively adopted for raising the robustness of our tracker.