LaSOT: A High-quality Benchmark for Large-scale Single Object Tracking —Supplementary Material—

Table 1. Details of 70 object categories in LaSOT and comparison with existing dense benchmark. Best viewed when zoomed-in.

NUS-PRO [4]		OTB-2015 [7]		TC-128 [5]		UAV123 [6]		VOT-17 [3]		NfS [2]		LaSOT	
class	# entries	class	# entries	class	# entries	class	# entries	class	# entries	class	# entries	class	# entries
person	193	person	36	person	45	person	48	person	19	ball	21	airplane	20
head	60	head	26	head	16	car	30	head	5	person	20	basketball	20
car	31	car	12	sphere	8	drone	10	fish	4	animal	10	bear	20
airplane	20	toy	8	2D print	5	wakeboard	10	motorcycle	4	vehicle	9	bicycle	20
boat	20	2D print	4	bicycle	5	boat	9	car	3	shuffeboard	8	bird	20
helicopter	20	cuboid	3	car	5	building	5	drone	3	face	6	boat	20
motorcycle	20	bird	2	ball	4	truck	5	ant	2	cup	4	book	20
drone	1	motorcycle	1	toy	4	bicycle	3	ball	2	dollar	4	bottle	20
-	-	deer	1	hand	3	bird	3	bird	2	aircraft	4	bus	20
-	-	bottle	1	kite	3	-	-	toy	2	airboard	2	car	20
-	-	panda	1	logo	3	-	-	bag	1	fish	2	cat	20
-	-	board	1	cuboid	3	-	-	book	1	motorcycle	2	cattle	20
-	-	can	1	boat	2	-	-	butterfly	1	drone	2	chameleon	20
-	-	dog	1	cup	2	-	-	cable	1	bicycle	2	coin	20
-	-	transformer	1	fish	2	-	-	crab	1	bird	2	crab	20
-	-	bicycle	1	guitar	2	-	-	cat	1	bag	1	crocodile	20
-	-	-	-	bird	2	-	-	flamingo	1	yoyo	1	cup	20
-	-	-	-	microphone	2	-	-	frisbee	1	-	-	deer	20
-	-	-	-	torso	2	-	-	glove	1	-	-	dog	20
-	-	-	-	motorcycle	2	-	-	hand	1	-	-	drone	20
-	-	-	-	airplane	2	-	-	helicopter	1	-	-	electricFan	20
-	-	-	-	board	1	-	-	leaf	1	-	-	elephant	20
-	-	-	-	bottle	1	-	-	rabbit	1	-	-	flag	20
-	-	-	-	can	1	-	-	sheep	1	-	-	fox	20
-	-	-	-	deer	1	-	-	-	-	-	-	frog	20
-	-	-	-	ring	1	-	-	-	-	-	-	gameTarget	20
-	-	-	-	torus	1	-	-	-	-	-	-	gecko	20
-	-	-	-	-	-	-	-	-	-	-	-	giraffe	20
-	-	-	-	-	-	-	-	-	-	-	-	goldfish	20
-	-	-	-	-	-	-	-	-	-	-	-	gorilla	20
-	-	-	-	-	-	-	-	-	-	-	-	guitar	20
-	-	-	-	-	-	-	-	-	-	-	-	hand	20
-	-	-	-	-	-	-	-	-	-	-	-	hat	20
-	-	-	-	-	-	-	-	-	-	-	-	helmet	20
-	-	-	-	-	-	-	-	-	-	-	-	hippo	20
-	-	-	-	-	-	-	-	-	-	-	-	horse	20
-	-	-	-	-	-	-	-	-	-	-	-	kangaroo	20
-	-	-	-	-	-	-	-	-	-	-	-	kite	20
-	-	-	-	-	-	-	-	-	-	-	-	leopard	20
-	-	-	-	-	-	-	-	-	-	-	-	licensePlate	20
-	-	-	-	-	-	-	-	-	-	-	-	lion	20
-	-	-	-	-	-	-	-	-	-	-	-	lizard	20
-	-	-	-	-	-	-	-	-	-	-	-	microphone	20
-	-	-	-	-	-	-	-	-	-	-	-	monkey	20
-	-	-	-	-	-	-	-	-	-	-	-	motorcycle	20
-	-	-	-	-	-	-	-	-	-	-	-	mouse	20
-	-	-	-	-	-	-	-	-	-	-	-	person	20
-	-	-	-	-	-	-	-	-	-	-	-	pig	20
-	-	-	-	-	-	-	-	-	-	-	-	pool	20
-	-	-	-	-	-	-	-	-	-	-	-	rabbit	20
-	-	-	-	-	-	-	-	-	-	-	-	racing	20
-	-	-	-	-	-	-	-	-	-	-	-	robot	20
-	-	-	-	-	-	-	-	-	-	-	-	rubicCube	20
-	-	-	-	-	-	-	-	-	-	-	-	sepia	20
-	-	-	-	-	-	-	-	-	-	-	-	shark	20
-	-	-	-	-	-	-	-	-	-	-	-	shark	20
-	-	-	-	-	-	-	-	-	-	-	-	sheep	20
-	-	-	-	-	-	-	-	-	-	-	-	skateboard	20
-	-	-	-	-	-	-	-	-	-	-	-	spider	20
-	-	-	-	-	-	-	-	-	-	-	-	squirrel	20
-	-	-	-	-	-	-	-	-	-	-	-	surfboard	20
-	-	-	-	-	-	-	-	-	-	-	-	swing	20
-	-	-	-	-	-	-	-	-	-	-	-	tank	20
-	-	-	-	-	-	-	-	-	-	-	-	tiger	20
-	-	-	-	-	-	-	-	-	-	-	-	train	20
-	-	-	-	-	-	-	-	-	-	-	-	truck	20
-	-	-	-	-	-	-	-	-	-	-	-	turtle	20
-	-	-	-	-	-	-	-	-	-	-	-	umbrella	20
-	-	-	-	-	-	-	-	-	-	-	-	volleyball	20
-	-	-	-	-	-	-	-	-	-	-	-	yoyo	20

1. Details of 70 Object Categories in LaSOT and Comparison with Existing Dense Benchmarks

LaSOT consists of 70 object categories with each containing 20 videos, as shown in Tab. 1. Most of 70 classes are chosen form the 1,000 classes in ImageNet [1], with a few exceptions such as *drone* and *gametarget*, which are carefully selected by the experts for tracking. The selection of each category must be agreed upon by all the experts to ensure its usability for visual tracking. In addition, we also compare the object categories of different dense benchmarks. As shown in Tab. 1, the number of object categories in LaSOT is two times more than that of existing benchmarks (*e.g.*, TC-128 [5] with 27 classes). Moreover, LaSOT eliminates the category bias of dataset for tracking while others do not.

2. Traing/Testing Split in Protocol II

In protocol II, we split LaSOT into *training* and *testing* sets. The *training* set contains of 1,120 videos (*i.e.*, 16 sequences for each category) with 2.83M frames in total. The rest 280 videos (*i.e.*, 4 sequences for each category) with 690K frames are used for testing.

Table 2. Comparison between *training* and *testing* sets of LaSOT.

	Video	Min frames	Mean frames	Median frames	Max frames	Total frames	Total duration
$\text{LaSOT}_{\text{training}}$	1,120	1,000	2,529	2,043	11,397	2.83M	26.2 **hours**
$\text{LaSOT}_{\text{testing}}$	280	1,000	2,448	2,102	9,999	690K	6.3 **hours**
LaSOT	1,400	1,000	2,506	2,053	11,397	3.52M	32.5 **hours**

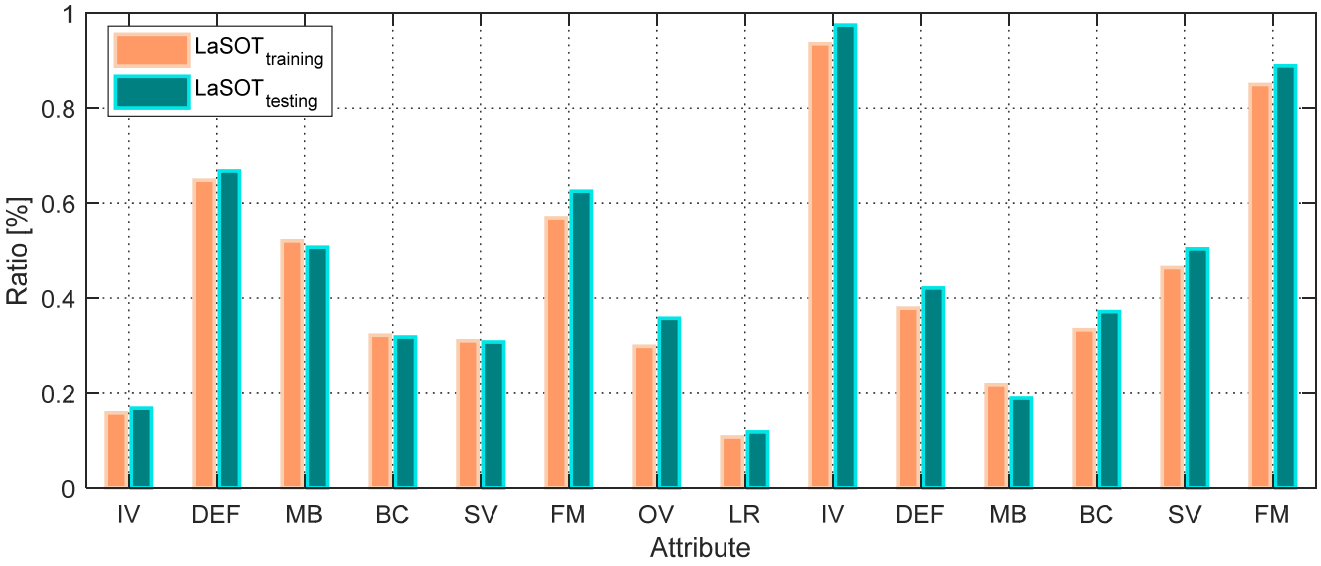

Figure 1. Comparison of sequence distribution in each attribute between *training* and *testing* sets. Best viewed in color.

Tab. 2 reports the detailed comparison between the *training* and the *testing* sets of LaSOT. We observe that the *min frames*, *mean frames*, *median frames* and *max frames* are similar between these two subsets. In addition, as shown in Fig. 1, we can see that the ratios of sequences in all 14 attributes are similar. Both Tab. 2 and Fig. 1 evidence the consistency of our training/testing split.

3. Detailed Attribute-based Performance under Protocol I

Fig. 2 shows the performance of trackers on each attribute using precision under protocol I.

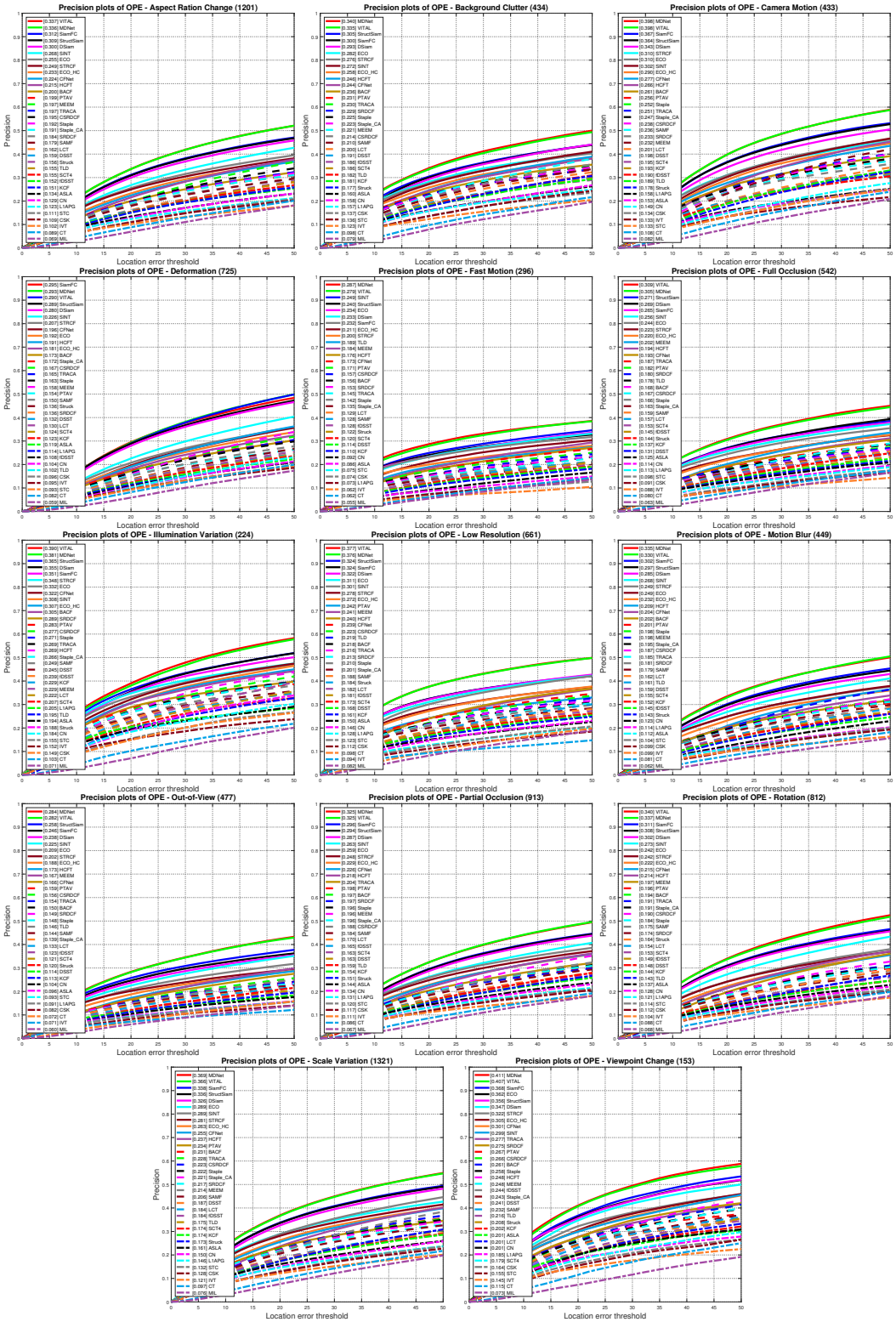

Figure 2. Performance of trackers on each attribute using precision under protocol I. Best viewed in color.

Fig. 3 shows the performance of trackers on each attribute using normalized precision under protocol I.

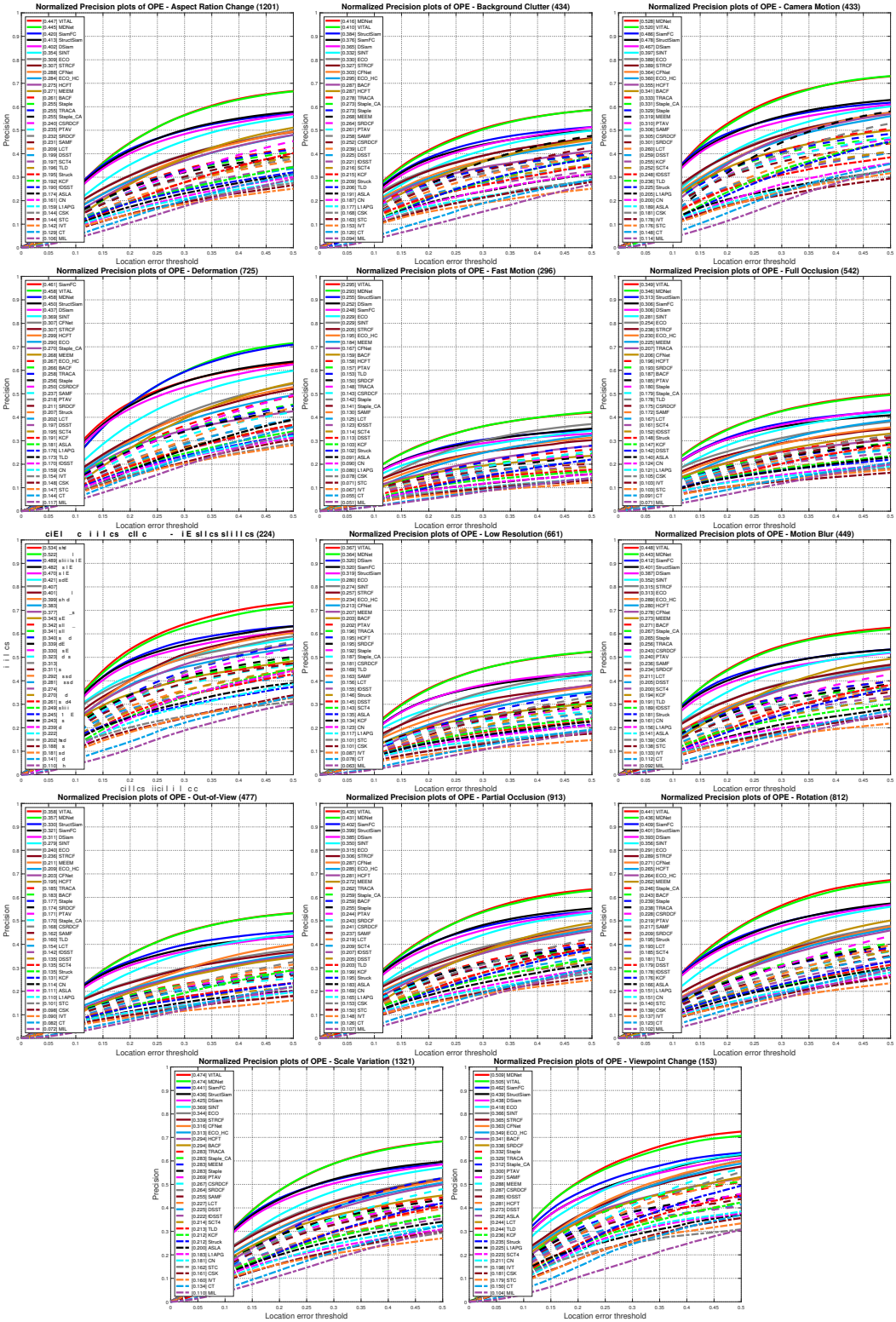

Figure 3. Performance of trackers on each attribute using precision under protocol I. Best viewed in color.

Fig. 4 shows the performance of trackers on each attribute using success under protocol I.

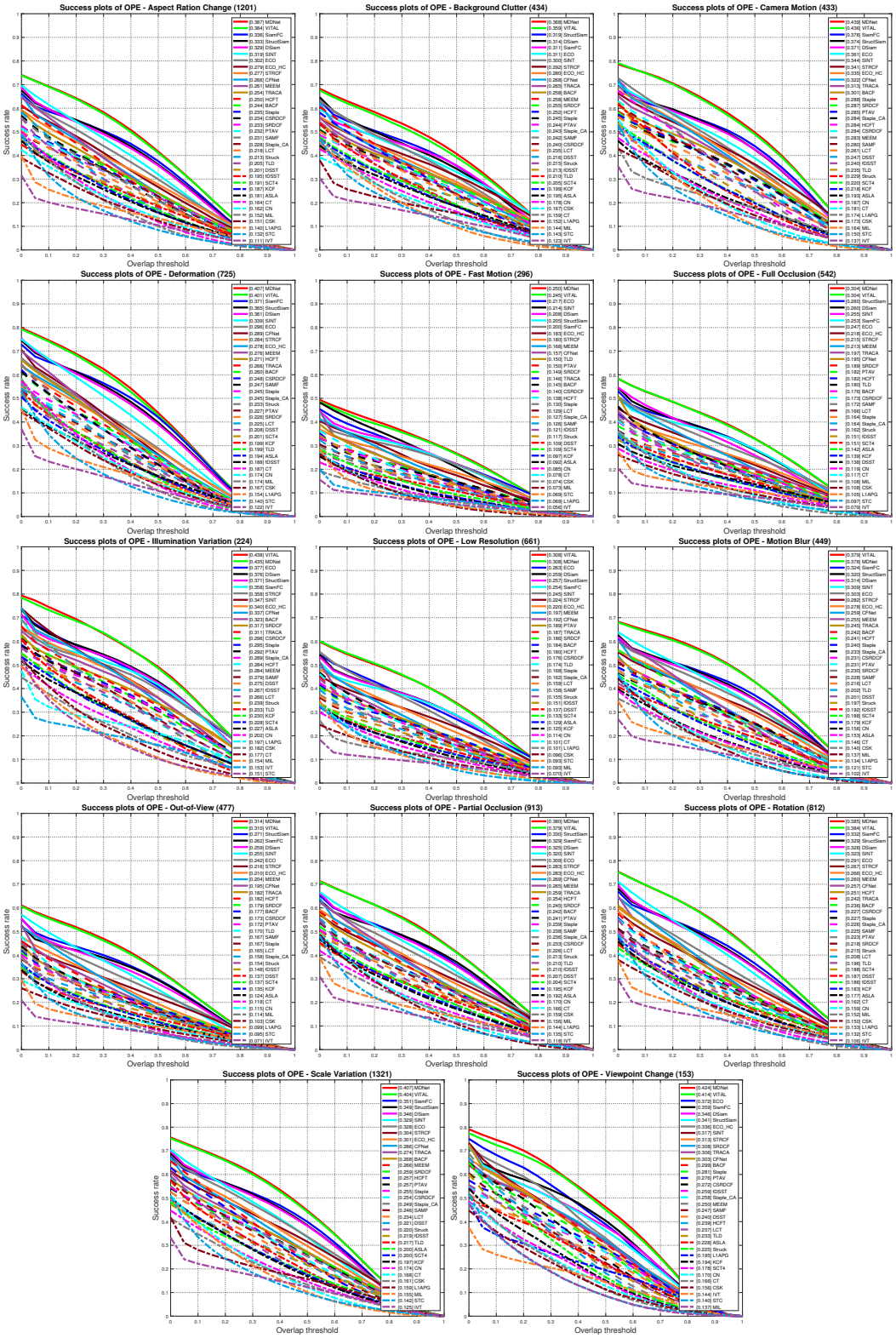

Figure 4. Performance of trackers on each attribute using success under protocol I. Best viewed in color.

4. Detailed Attribute-based Performance under Protocol II

Fig. 5 shows the performance of trackers on each attribute using precision under protocol II.

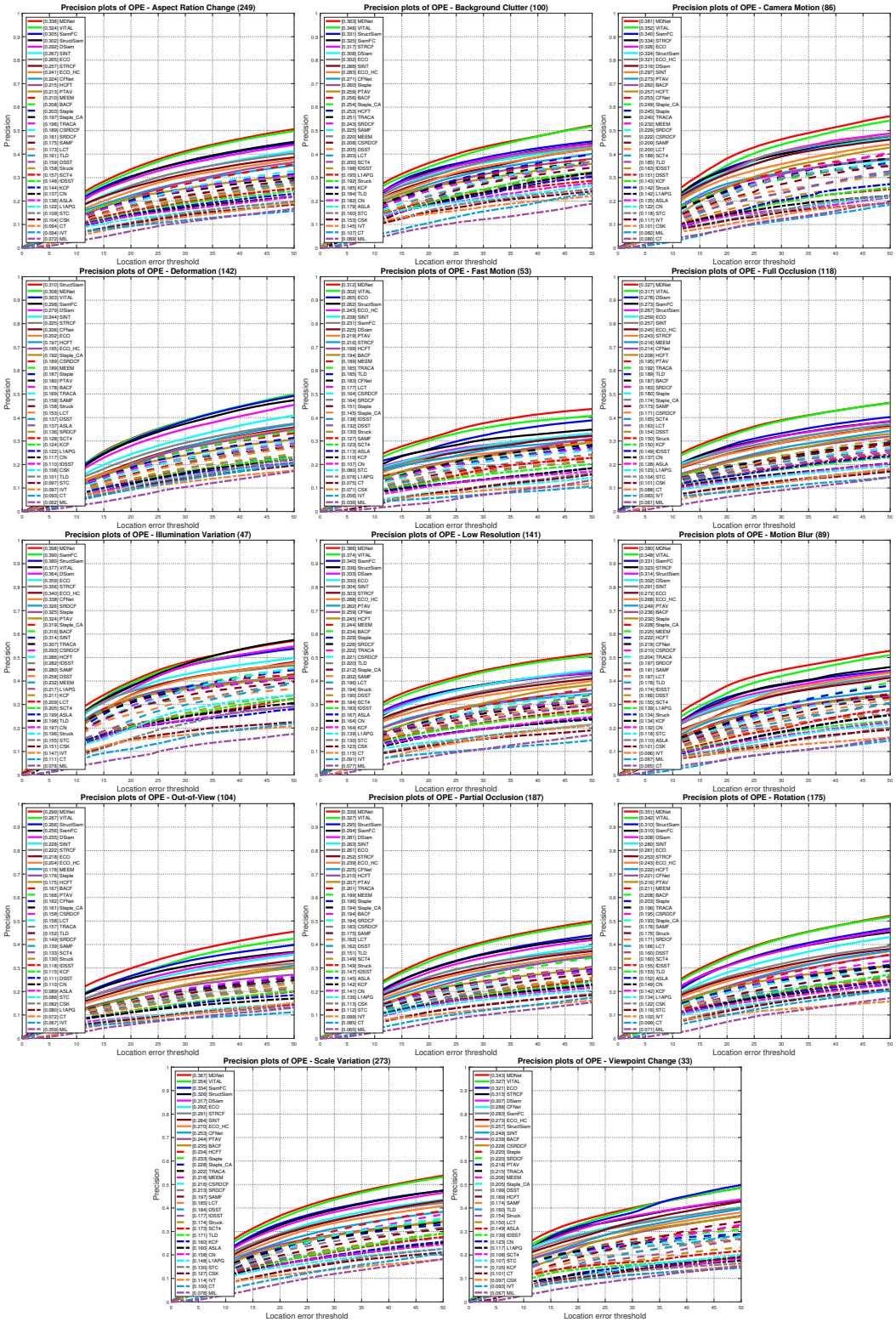

Figure 5. Performance of trackers on each attribute using precision under protocol II. Best viewed in color.

Fig. 6 shows the performance of trackers on each attribute using normalized precision under protocol II.

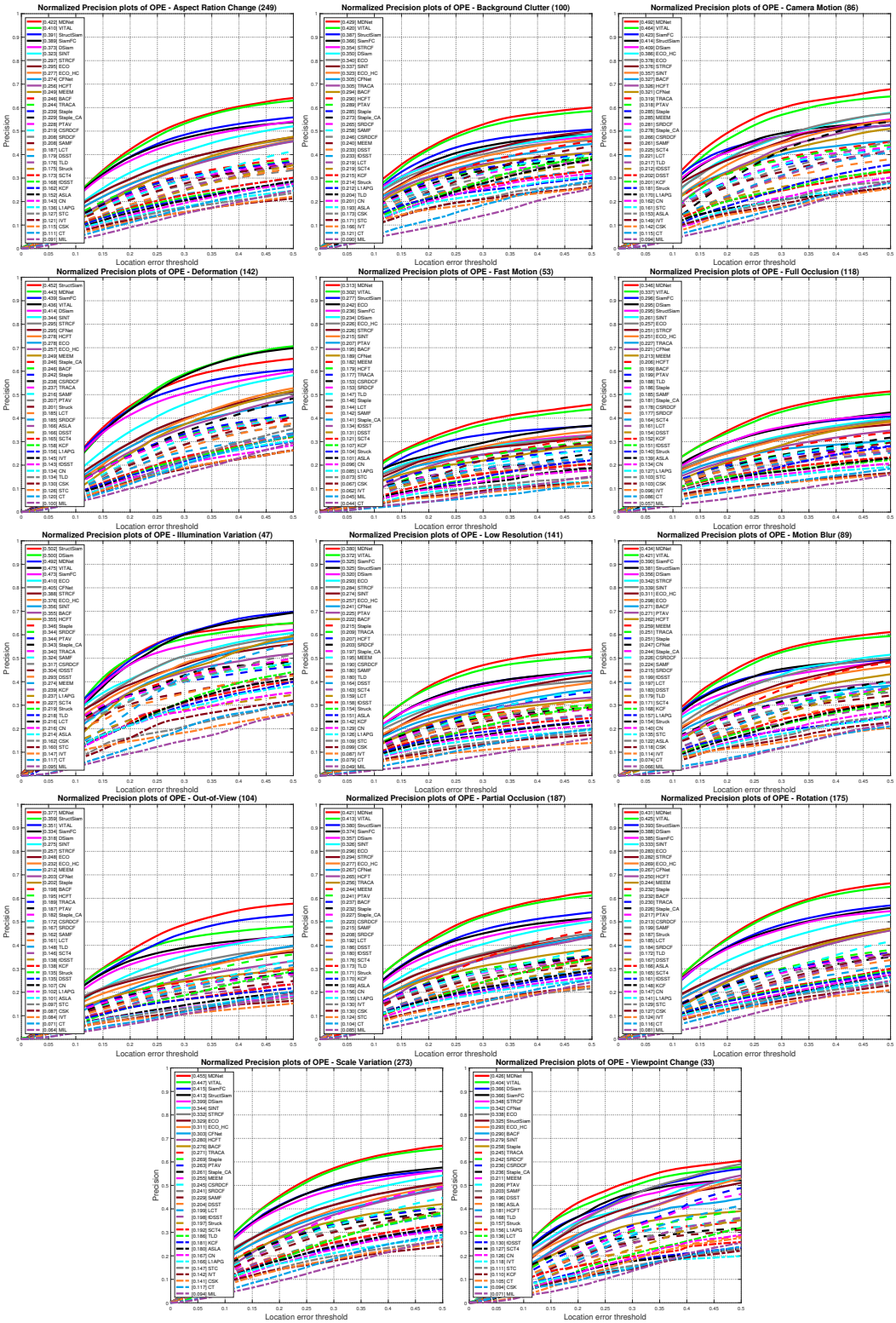

Figure 6. Performance of trackers on each attribute using precision under protocol II. Best viewed in color.

Fig. 7 shows the performance of trackers on each attribute using success under protocol II.

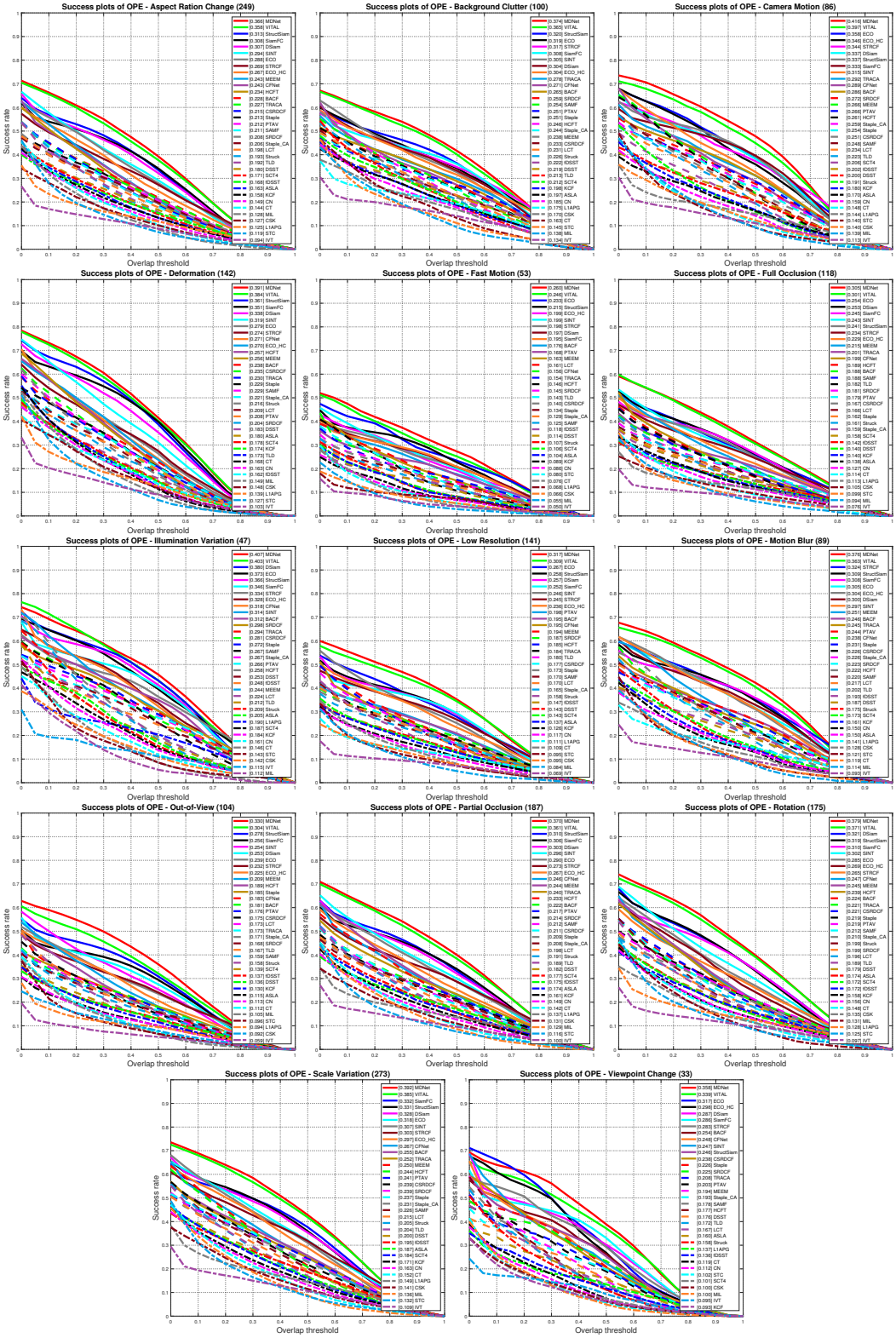

Figure 7. Performance of trackers on each attribute using success under protocol II. Best viewed in color.

References

[1] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, 2009. 2

[2] Hamed Kiani Galoogahi, Ashton Fagg, Chen Huang, Deva Ramanan, and Simon Lucey. Need for speed: A benchmark for higher frame rate object tracking. In *ICCV*, 2017. 1

[3] Matej Kristan et al. The visual object tracking vot2017 challenge results. In *ICCVW*, 2017. 1

[4] Annan Li, Min Lin, Yi Wu, Ming-Hsuan Yang, and Shuicheng Yan. Nus-pro: A new visual tracking challenge. *TPAMI*, 38(2):335–349, 2016. 1

[5] Pengpeng Liang, Erik Blasch, and Haibin Ling. Encoding color information for visual tracking: Algorithms and benchmark. *TIP*, 24(12):5630–5644, 2015. 1, 2

[6] Matthias Mueller, Neil Smith, and Bernard Ghanem. A benchmark and simulator for uav tracking. In *ECCV*, 2016. 1

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