Supplementary Material for "Learning to Track: Online Multi-Object Tracking by Decision Making"

Yu Xiang[1,2], Alexandre Alahi[1], and Silvio Savarese[1]
[1]Stanford University, [2]University of Michigan at Ann Arbor
yuxiang@umich.edu, {alahi, ssilvio}@stanford.edu

In this supplementary material, we present additional evaluation of our experiments in the paper "Learning to Track: Online Multi-Object Tracking by Decision Making".

1. Experimental Settings

We conduct experiments on the Multiple Object Tracking Benchmark [2] for people tracking. There are 11 sequences for training, and 11 sequences for testing in the MOT benchmark. Since the annotations of the test set are not released, we separate a validation set of 6 sequences from the 11 training sequences to conduct analysis about our framework. The training and testing splitting for validation and testing is shown in Table 1. The metrics used to evaluate the multiple object tracking performance as suggested by the MOT Benchmark is shown in Table 2. The thresholds e_0 and o_0 in tracked states are set to 10 and 0.8 respectively, and the threshold T_{lost} in lost states is set to 50 in all the experiments.

2. Analysis on Validation Set

Contribution of Different Components. In this experiment, we investigate the contribution of different components in our framework by disabling a component at one time and then examining the tracking performance on the validation set. To recap, Fig. 1 illustrates our target MDP in modeling the lifetime of a target, and Table 3 describes our feature representation used in data association for lost states. The experimental results are shown in Table 4, where we disable action a_3 in tracked states, action a_6 in lost states, FB error in optical flow ($\phi_1, \cdots, \phi_5$), Normalized Correlation Coefficient (NCC, ϕ_6 and ϕ_7), ratio between the heights of bounding box (ϕ_8 and ϕ_9), and distance between the target and the detection (ϕ_{12}) respectively. By using the full model for comparison, we can see the contribution of different components in our framework.

Cross-domain Tracking. We conduct experiments by testing the trained tracker in different scenarios to investigate the generalization capability of our method. In Table 5, we present the tracking results of trackers trained on different training sequences on the six test sequences in the validation set. As we can see from the table, trackers trained on the five training sequences achieve similar performance on the test sequences. In some cases, cross-domain testing even improves the results. The experimental results demonstrate the generalization power of our framework.

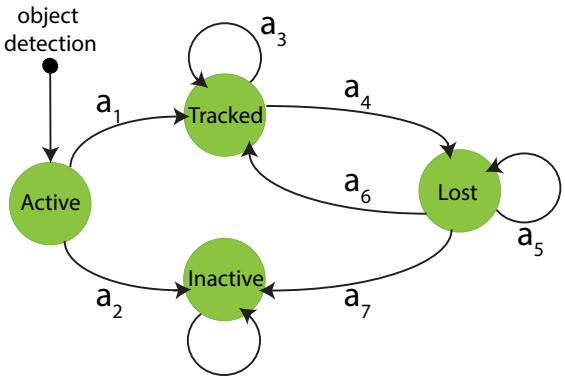

Figure 1. The target MDP in our framework.

3. Evaluation on Test Set

After the analysis on the validation set, we perform training with all the training sequences, and test the trained trackers on the test set according to Table 1. Table 6 presents detailed tracking evaluation of our framework on the 11 sequences in the test set of the MOT benchmark.

References

[1] B. Keni and S. Rainer. Evaluating multiple object tracking performance: the clear mot metrics. *EURASIP Journal on Image and Video Processing*, 2008:1:1–1:10, 2008. 2

[2] L. Leal-Taixé, A. Milan, I. Reid, S. Roth, and K. Schindler. MOTChallenge 2015: Towards a Benchmark for Multi-Target Tracking. *arXiv:1504.01942 [cs]*, 2015. 1

[3] Y. Li, C. Huang, and R. Nevatia. Learning to associate: Hybridboosted multi-target tracker for crowded scene. In *CVPR*, pages 2953–2960, 2009. 2

Training	Testing
Validation on MOT Benchmark	
TUD-Stadtmitte	TUD-Campus
ETH-Bahnhof	ETH-Sunnyday, ETH-Pedcross2
ADL-Rundle-6	ADL-Rundle-8, Venice-2
KITTI-13	KITTI-17
Testing on MOT Benchmark	
TUD-Stadtmitte, TUD-Campus	TUD-Crossing
PETS09-S2L1	PETS09-S2L2, AVG-TownCentre
ETH-Bahnhof, ETH-Sunnyday, ETH-Pedcross2	ETH-Jelmoli, ETH-Linthescher, ETH-Crossing
ADL-Rundle-6, ADL-Rundle-8	ADL-Rundle-1, ADL-Rundle-3
KITTI-13, KITTI-17	KITTI-16, KITTI-19
Venice-2	Venice-1

Table 1. Training and Testing sequences for validation and testing on the MOT Benchmark.

MOTA	Multiple Object Tracking Accuracy [1]. This measure combines three error sources: false positives, missed targets and identity switches.
MOTP	Multiple Object Tracking Precision [1]. The misalignment between the annotated and the predicted bounding boxes.
GT	The total number of ground truth trajectories.
MT	Mostly tracked targets. Percentage of ground truth trajectories that are covered by tracking output for at least 80% of their respective life span.
ML	Mostly lost targets. Percentage of ground truth trajectories that are covered by tracking output less than 20% of their respective life span.
FP	The total number of false positives.
FN	The total number of false negatives (missed targets).
IDS	The total number of identity switches [3].
Frag	The total number of times a trajectory is fragmented (i.e. interrupted during tracking).

Table 2. Evaluation metrics used for multi-object tracking.

Type	Notation	Feature Description
FB error	$\phi_1, \cdots, \phi_5$	Mean of the median forward-backward errors from the entire, left half, right half, upper half and lower half of the templates in optical flow
NCC	ϕ_6	Mean of the median Normalized Correlation Coefficients (NCC) between image patches around the matched points in optical flow
	ϕ_7	Mean of the NCC between image patches of the detection and the predicted bounding boxes from optical flow
Height ratio	ϕ_8	Mean of the ratios in bounding box height between the detection and the predicted bounding boxes from optical flow
	ϕ_9	Ratio in bounding box height between the target and the detection
Overlap	ϕ_{10}	Mean of the bounding box overlaps between the detection and the predicted bounding boxes from optical flow
Score	ϕ_{11}	Normalized detection score
Distance	ϕ_{12}	Euclidean distance between the centers of the target and the detection after motion prediction of the target with a linear velocity model

Table 3. Our feature representation for data association.

Tracker	MOTA	MOTP	GT	MT	ML	FP	FN	IDS	Frag
Full model	26.6	73.8	234	9.8%	55.1%	2,691	14,130	123	276
Disable a_3 in tracked	25.4	73.6	234	8.5%	57.7%	2,628	14,456	149	284
Disable a_6 in lost	20.9	74.0	234	3.4%	67.9%	1,895	15,951	427	269
Disable FB error	23.6	73.4	234	9.8%	50.9%	3,910	13,560	173	347
Disable NCC	23.6	73.4	234	10.7%	52.6%	3,891	13,589	148	329
Disable height ratio	24.5	73.5	234	10.7%	54.7%	3,692	13,623	119	310
Disable distance	21.4	73.5	234	9.8%	54.7%	4,235	13,704	209	336

Table 4. Analysis of our framework on the validation set by disabling different components.

Testing	Training	MOTA	MOTP	GT	MT	ML	FP	FN	IDS	Frag
TUD-Campus	TUD-Stadmitte	56.0	73.0	8	37.5%	0.0%	36	117	5	8
	ETH-Bahnhof	44.8	72.1	8	0.0%	0.0%	34	156	8	11
	ADL-Rundle-6	47.9	72.8	8	0.0%	12.5%	19	156	12	12
	KITTI-13	53.2	71.6	8	37.5%	0.0%	43	120	5	9
	PETS09-S2L1	49.0	71.8	8	25.0%	0.0%	29	138	16	10
ETH-Sunnyday	TUD-Stadmitte	46.8	76.4	30	30.0%	33.3%	266	713	9	35
	ETH-Bahnhof	43.4	77.1	30	20.0%	33.3%	217	807	28	40
	ADL-Rundle-6	48.2	76.4	30	23.3%	33.3%	148	785	29	34
	KITTI-13	47.5	76.6	30	26.7%	33.3%	250	716	9	34
	PETS09-S2L1	42.1	77.0	30	20.0%	36.7%	225	811	39	49
ETH-Pedcross2	TUD-Stadmitte	14.0	70.5	133	3.0%	75.9%	311	5053	24	79
	ETH-Bahnhof	13.3	71.0	133	3.0%	77.4%	264	5153	13	61
	ADL-Rundle-6	11.5	71.6	133	0.8%	81.2%	205	5293	44	52
	KITTI-13	13.9	70.5	133	3.0%	74.4%	316	5051	24	78
	PETS09-S2L1	11.5	71.4	133	3.8%	78.2%	272	5223	45	69
ADL-Rundle-8	TUD-Stadmitte	20.0	72.7	28	21.4%	32.1%	1715	3694	19	93
	ETH-Bahnhof	22.6	73.0	28	21.4%	32.1%	1463	3760	26	99
	ADL-Rundle-6	26.1	73.5	28	17.9%	35.7%	1048	3934	34	84
	KITTI-13	20.9	73.1	28	21.4%	32.1%	1591	3746	26	86
	PETS09-S2L1	22.1	73.3	28	17.9%	35.7%	1394	3828	63	99
Venice-2	TUD-Stadmitte	30.8	74.0	26	15.4%	23.1%	1187	3720	33	90
	ETH-Bahnhof	30.8	74.6	26	15.4%	26.9%	1109	3803	33	74
	ADL-Rundle-6	29.8	74.3	26	15.4%	23.1%	1080	3895	39	70
	KITTI-13	32.1	74.2	26	19.2%	23.1%	1182	3625	42	82
	PETS09-S2L1	29.4	74.6	24	23.1%	23.1%	1073	3880	87	88
KITTI-17	TUD-Stadmitte	60.8	72.3	9	11.1%	0.0%	36	226	6	12
	ETH-Bahnhof	60.3	72.0	9	11.1%	0.0%	30	235	6	13
	ADL-Rundle-6	57.8	73.4	9	11.1%	11.1%	20	258	10	9
	KITTI-13	59.9	71.6	9	11.1%	0.0%	46	224	4	13
	PETS09-S2L1	61.2	72.7	9	11.1%	0.0%	32	230	3	8

Table 5. Tracking performance with different pairs of training and testing sequences on the validation set.

Sequence	MOTA	MOTP	GT	MT	ML	FP	FN	IDS	Frag
TUD-Crossing	69.4	73.9	13	53.8%	7.7%	24	305	8	25
PETS09-S2L2	47.8	69.8	42	14.3%	7.1%	661	4,163	206	362
ETH-Jelmoli	32.9	73.6	45	17.8%	28.9%	639	1,041	22	71
ETH-Linthescher	27.2	74.7	197	6.1%	64.0%	191	6,262	48	107
ETH-Crossing	28.8	74.7	26	11.5%	46.2%	59	655	0	15
AVG-TownCentre	25.4	69.7	226	17.7%	33.6%	1,517	3,691	122	264
ADL-Rundle-1	16.2	71.5	32	25.0%	28.1%	3,157	4,597	49	140
ADL-Rundle-3	34.8	73.1	44	11.4%	29.5%	1,224	5,326	78	114
KITTI-16	40.4	73.0	17	0.0%	17.6%	204	775	34	66
KITTI-19	26.6	65.9	62	6.5%	22.6%	1,198	2,658	66	242
Venice-1	15.9	72.4	17	5.9%	41.2%	843	2,949	47	94
ALL	30.3	71.3	721	13.0%	38.4%	9,717	32,422	680	1,500

Table 6. Tracking performance on the test set of the MOT Benchmark.