

Supplemental Material for "Structure Preserving Object Tracking"

1. Videos

In the supplemental material, we present a number of selected videos (because of the 30 Mb space limitations) to illustrate the performance of our SPOT tracker. We present videos that correspond to Experiment 1 and videos that correspond to Experiment 2:

- Experiment 1 – multiple objects: We present tracks obtained using the mst-SPOT tracker (single scale) for the following videos: *Air Show*, *Car Chase*, *Hunting*, *Red Flowers*, *Shaking*, and *Sky Diving*. Different colors represent different objects.
- Experiment 2 – single object: We present tracks obtained using the mst-SPOT tracker (single scale) for the following videos: *Coke Can*, *Dollar*, *Occl. Face 2*, and *Tiger 2*.

2. Additional Results

As mentioned in the paper, we developed a single-scale and a multiple-scale version of our tracker. The results obtained with the single-scale tracker are presented in our paper. Here, we present results obtained using multi-scale trackers. The multiple-scale trackers are run at three scales for each frame (viz. relative scales 0.9, 1.0, and 1.1), and select the highest posterior probability among the three scales to determine the location and scale of the object. In all experiments, we assume that the aspect ratio of the rectangular bounding box is fixed.

Table 1 presents the performance obtained by the TLD [1], no-SPOT, and mst-SPOT trackers when tracking multiple objects in multiple scales in terms of average distance error (Err.). See the paper for details on how the performance criteria are computed. The results presented in Table 1 are in line with those presented in the main paper: our mst-SPOT tracker outperforms trackers without structural constraints on most of the videos. The differences between single-scale and multi-scale mst-SPOT trackers are rather small, which may be due to a ceiling effect.

Table 2 presents the performance obtained by the TLD, no-SPOT, and mst-SPOT trackers when tracking a single object. As in the experiments presented in the paper, the mst-SPOT tracker uses two parts. The results in Table 2 show that all three trackers have very similar performances when tracking single objects (on multiple scales). The mst-SPOT does not improve the results in this set of experiments; it is actually outperformed by the single-scale mst-SPOT tracker. At present, we are still unsure as to why this happens.

References

- [1] Z. Kalal, J. Matas, and K. Mikolajczyk. P-n learning: Bootstrapping binary classifiers by structural constraints. In *CVPR*, pages 49–56, 2010. 1, 2

Table 1. Performance of three multi-scale model-free trackers on multiple-object videos measured in terms of average distance in pixels between centers of the predicted and the ground-truth bounding box (lower is better). The results are averaged over five runs and over all target objects in each video. The best performance on each video is boldfaced.

	TLD [1]	no-SPOT	mst-SPOT
Air Show	21.3	11.6	7.6
Car Chase	22.4	69.2	4.1
Parade	8.8	19.8	15.7
Red Flowers	40.2	72.6	13.7
Hunting	133.5	125.9	17.6
Sky Diving	5.8	19.5	9.8
Shaking	14.3	30.9	9.9
Basketball	15.6	34.5	9.4
Skating	90.3	88.4	27.3
Avg. rank	2.1	2.7	1.2

Table 2. Performance of three multi-scale model-free trackers on single-object videos measured in terms of average distance in pixels between the centers of the predicted and the ground-truth bounding box (lower is better). The results are averaged over five runs. The best performance on each video is boldfaced.

	TLD [1]	no-SPOT	mst-SPOT
Sylvester	5.7	12.7	15.6
David	4.5	4.9	5.5
Cola Can	9.7	9.6	18.7
Occl. Face 1	14.8	9.0	6.5
Occl. Face 2	14.8	8.6	8.0
Surfer	8.1	10.8	12.6
Tiger 1	10.3	53.5	10.3
Tiger 2	23.2	36.9	11.6
Dollar	65.3	5.7	8.6
Cliff bar	8.6	13.8	89.2
Tea Box	51.1	12.4	27.6
Girl	28.2	12.2	10.4
Avg. rank	2.0	1.9	2.0