Online, Real-Time Tracking Using a Category-to-Individual Detector: Supplementary Material

David Hall and Pietro Perona

California Institute of Technology

In the supplementary material we report the results of some additional experiments. The most important findings are reported in the manuscript. In Figure 1 we report the average performance and frame rate of each tracking algorithm on the Caltech Pedestrians Dataset. In Figure 2 we report the tracking performance of each tracking algorithm for a category detector operating at different recall rates. The rankings of the algorithms across the different datasets and across the different category detector recall rates gives us confidence that our findings are general.

In Figure 3 we report results by evaluating the tracking algorithms on every 30th frame of the datasets rather than on every frame as in the previous analysis. This experiment demonstrates how well trackers cope with large changes in appearance and in position. CIT is not the best method in this case for some settings, however, it is still 10–90x faster than the methods that do have a better average performance.

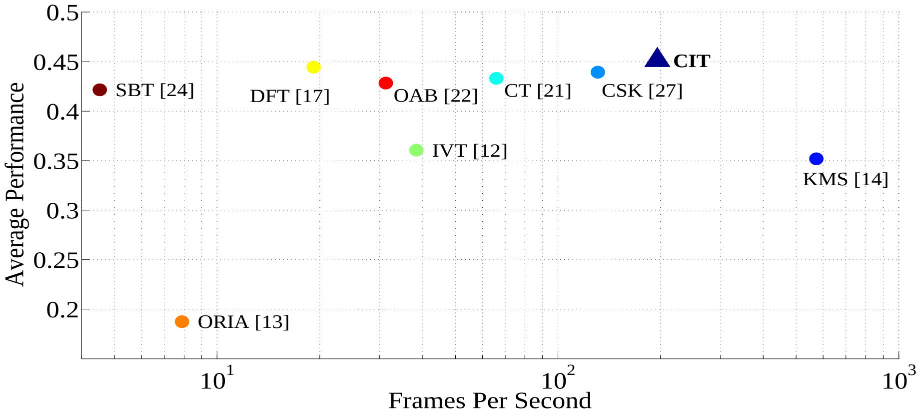

Fig. 1. The average performance and frame rate for each tracking algorithm using the Caltech Pedestrians dataset. The average performance is the same as the value in the bottom plot of Figure 5 in the manuscript. The results indicate that our method, CIT is the most accurate and slightly faster than the next most accurate. The relative performance and frame rate of the tracking methods are similar to those as reported on the Buffy dataset which can be found in Figure 6 of the manuscript. However, the absolute frame rate has increased significantly between datasets due to Caltech Pedestrians having a lower resolution than Buffy.

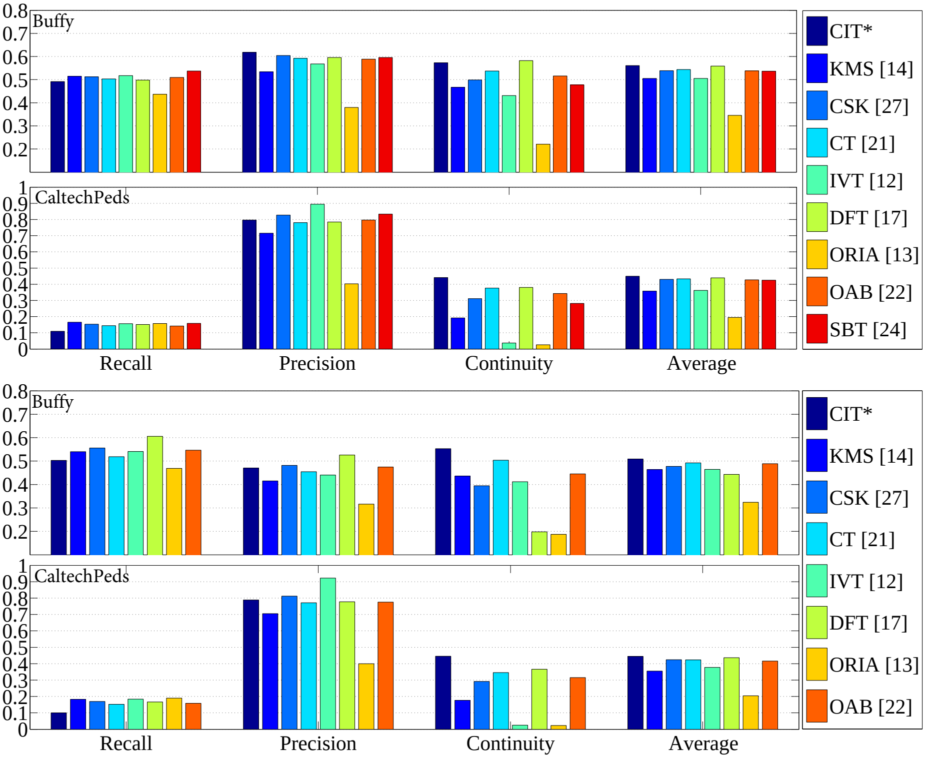

Fig. 2. Tracking Performance on Buffy and Caltech Pedestrians for a category detector operating at a recall of (top) 0.6 [Buffy], 0.7 [CaltechPeds] and (bottom) 0.7 [Buffy], 0.9 [CaltechPeds]. The results indicate that our method CIT is one of the best performers, however, most of the methods have similar performance results. The relative performance of the trackers is roughly equivalent across the two datasets as well as across the different category detector recall rates. (BSBT and SBT were omitted because they operate too slowly - at higher recall rates more objects are identified to track and so the computational requirements increase.)

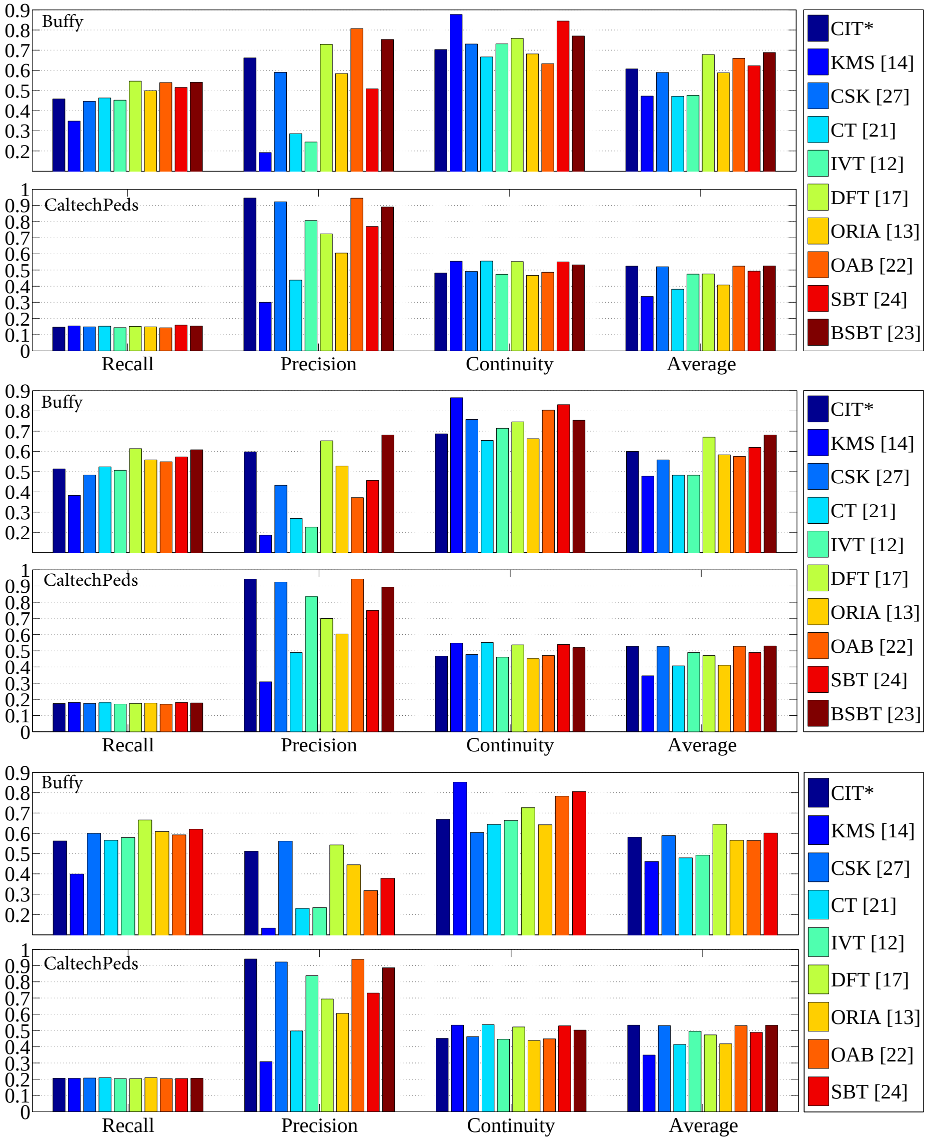

Fig. 3. Tracking Performance with a 30 frame jump on Buffy and Caltech Pedestrians for a category detector operating at a recall of (top) 0.5 [Buffy], 0.5 [CaltechPeds] (middle) 0.6 [Buffy], 0.7 [CaltechPeds] and (bottom) 0.7 [Buffy], 0.9 [CaltechPeds]. Instead of evaluating the trackers on every single frame in the datasets, they are evaluated on every 30th frame. This gives an indication of how well trackers can handle large appearance changes and large changes in position from one frame to the next. The performance of the KMS tracker suffers because the target can have large changes in position from frame to frame. CIT, CSK, DFT, OAB, SBT and BSBT are all able to handle large changes reasonably well. CIT is outperformed by DFT, OAB, SBT and BSBT on Buffy for a category detector recall of 0.5, however, these are the slowest tracking methods.