

A Novel Representation of Parts for Accurate 3D Object Detection and Tracking in Monocular Images Supplementary Material

We provide here some additional details about parameter values we could not put in the paper for spatial reasons. We also provide the graphs comparing our method against LSD-SLAM, LINE-2D and PWP3D on all the test sequences of our datasets. The datasets, including both learning and testing data, are available for download at http://cvlab.epfl.ch/data/3d_object_tracking.

Parameters Settings for the Experiments

For all our experiments, we use a MoG made by 9 components around a central pose, such as those depicted in Figure 6 of the paper. As mentioned in the paper, if the previous frame has been tracked, we add a 10^{th} component with the pose computed at the previous frame. The parameters used in all our tests are given in Table 1. For numerical comparisons, no additional temporal consistency or pose smoothing is enforced in our method. For real applications, a trajectory filter (e.g. Kalman) is recommended for avoiding visual jitter.

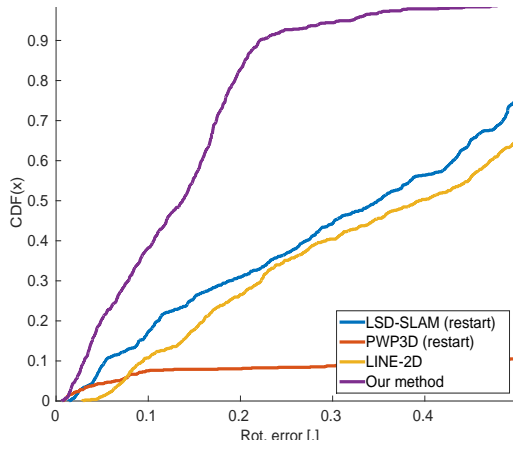
CDF of Rotation and Translation Error Graphs

In Figures 1, 2, 3, we report curves for the Cumulative Distribution functions (CDF) of the Absolute Rotation and Translation Error for all the datasets. For some frames, the groundtruth pose could not

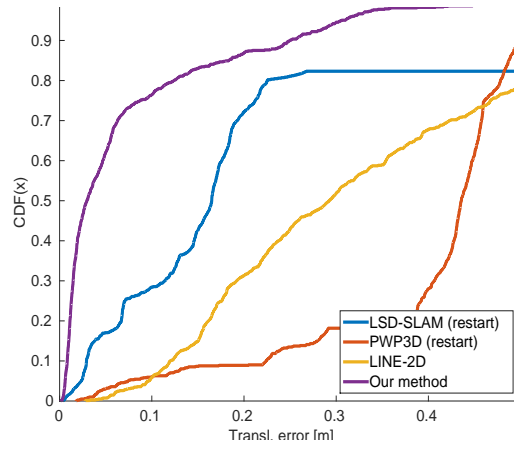
Parameter	Description	Value	units
M	num. of components in the MoG prior	9 (10)	-
T	threshold on repro.error between prior and detections	40	pixels
image size	image size for all videos and sequences	640×360	pixels
$CNN^{part-det}$ input patch size	patch size for $CNN^{part-det}$	32×32	pixels
$CNN^{part-det}$ activation	activation function for $CNN^{part-det}$	tanh	
$CNN^{part-det}$ pooling	pooling function for $CNN^{part-det}$	max over 2×2 patches	
$CNN^{cp-pred}$ input patch size	patch size for $CNN^{cp-pred}$	64×64	pixels
$CNN^{cp-pred}$ activation	activation function for $CNN^{cp-pred}$	tanh	
$CNN^{cp-pred}$ pooling	pooling function for $CNN^{cp-pred}$	max over 2×2 patches	
N_P	the number of the parts for an object	[1, 4]	-
N_V	the number of control points of the part	7	-

Table 1: Value of parameters for the experiments described in the paper.

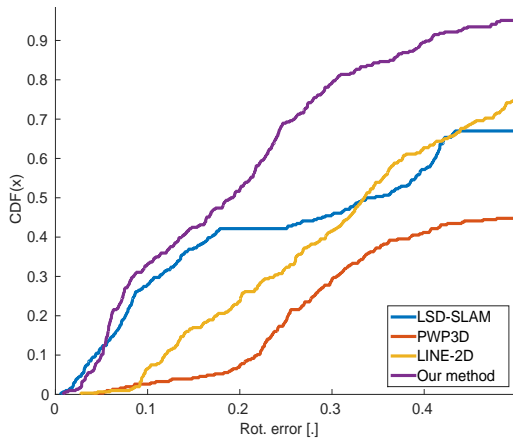
be retrieved, and those frames were not taken into account for the AUC scores computation.



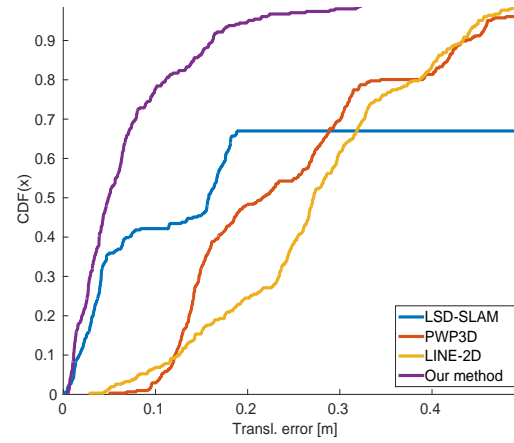
(a) Rotation error CDF: BOX dataset-video #1



(b) Translation error CDF: BOX dataset-video #1

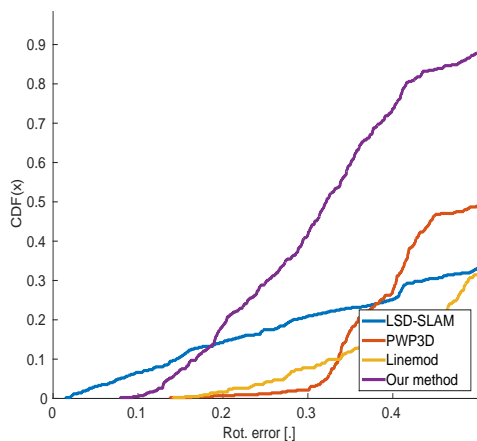


(c) Rotation error CDF: BOX dataset-video #2

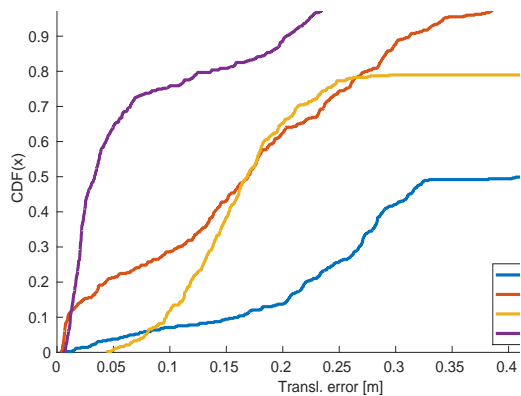


(d) Translation error CDF: BOX dataset-video #2

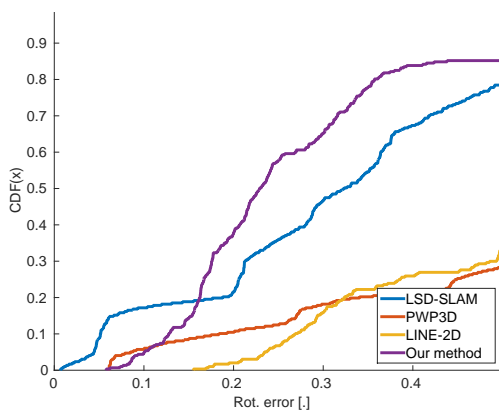
Figure 1: The Cumulative Distribution functions (CDF) of the Rotation and Translation errors on the BOX dataset videos.



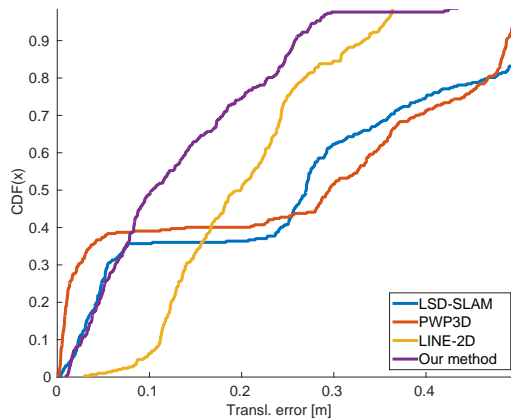
(a) Rotation error CDF: CAN dataset-video #1



(b) Translation error CDF: CAN dataset-video #1

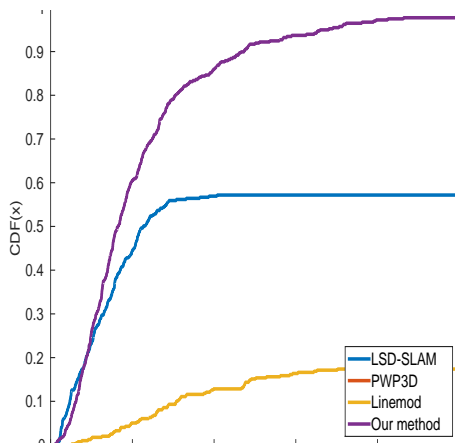


(c) Rotation error CDF: CAN dataset-video #2

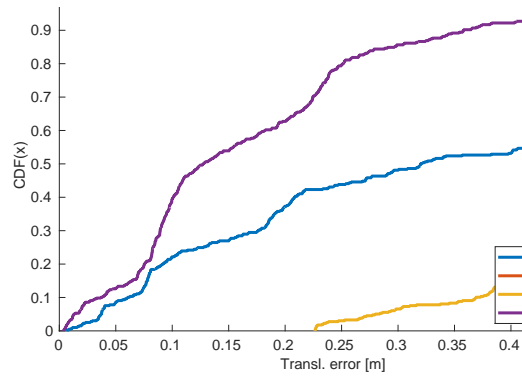


(d) Translation error CDF: CAN dataset-video #2

Figure 2: The Cumulative Distribution functions (CDF) of the Rotation and Translation errors on the CAN dataset videos. All methods perform pretty badly when estimating the rotation, due to the symmetrical shape of the target object.



(a) Rotation error CDF: DOOR dataset-video #1



(b) Translation error CDF: DOOR dataset-video #1

Figure 3: The Cumulative Distribution functions (CDF) of the Rotation and Translation errors on the DOOR dataset videos. PWP3D diverges immediately after initialization and is not visible on the curves.