Qeios

Peer Review

Review of: "TrafficLoc: Localizing Traffic Surveillance Cameras in 3D Scenes"

Yunshuang Yuan[1]

1. Independent researcher

The paper proposes a novel method for localizing traffic cameras within a 3D reference map. Extensive experiments demonstrate that the proposed modules — namely, the Geometry-guided Attention Loss, Inter-Intra Contrastive Learning, and Dense Matching Alignment Module — are effective in improving localization performance.

From my understanding, the CARLA dataset assumes that the 3D point cloud reference is accumulated and downsampled from a series of scans, resulting in relatively homogeneous sparsity. This assumption likely supports more robust geometrical feature learning. I am curious about the role that point cloud sparsity plays in the final localization results. Specifically, do the KITTI and nuScenes datasets exhibit similar point cloud characteristics as CARLA? Besides, at the accumulation step of the point clouds, the simulated dataset may assume these point clouds could be spatially perfectly aligned. In reality, at which alignment accuracy does the proposed method require?

In this paper, the authors focus on recovering the 6-DoF pose of traffic cameras. I believe a similar method could be beneficial for calibrating on-board camera parameters (both distortion and intrinsics) relative to on-board LiDARs. Have the authors conducted experiments in this setting? If so, how accurate is the resulting automatic calibration?

Minor Issues:

The expressions for the positive and negative margins in Equations (9) and (10) are somewhat confusing. In Equation (9), they appear very likely to be multiplicative factors, whereas in Equation (10), they are presented as functions.

In Equation (12), does the softmax operation run over the dimension "c"?

Section 5.3, Line 1: "Table 1sumarizes" → wrong spacing

Declarations

Potential competing interests: No potential competing interests to declare.