Qeios

Peer Review

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

Chenyang Li[1]

1. Lanzhou University, China

The proposed Geometry-Guided Feature Fusion (GFF) and Geometry-Guided Attention Loss (GAL) significantly improve cross-modal matching accuracy, outperforming existing methods in scenarios with large viewpoint variations. Additionally, the proposed internal contrastive learning approach effectively distinguishes pixel features and point features within the same image patch and point cloud. The introduced dense matching comparison method greatly enhances the accuracy of patch-to-point group matching. The paper evaluates the method on multiple datasets and analyzes the contributions of different modules through ablation studies, validating its effectiveness. The authors constructed the Carla Intersection Dataset, providing a new experimental benchmark for traffic surveillance camera localization tasks, and the method demonstrates a strong generalization ability across multiple datasets.

Overall, the paper is well-written and presented. I hope the following, somewhat stringent, review comments can help the authors further improve the quality of the paper.

Comment:

1. In the problem statement section, the definitions of some symbols are not clear. For example, when using equation (1), it is unclear whether pi needs to be transformed into homogeneous coordinates. This should be explicitly stated. Additionally, the symbol K is not clearly explained in the equation, and its role in the formula should be specified to avoid confusion for readers.

2. In the feature extraction section, the paper mentions using Farthest Point Sampling (FPS) to generate super-points, but the description of how FPS is combined with point cloud features,

 doi.org/10.32388/GCS1IP

especially how clustering and mapping are performed after generating M super-points, is somewhat unclear.

3. The authors use L_det to locate the super-points, which seems to be the basis for coarse matching. However, no experiments are provided to analyze the correspondence between super-points and image patches.

4. As far as I know, Farthest Point Sampling (FPS) is an inefficient operation. Have the authors considered using other, more efficient point sampling methods and conducting comparative experiments?

5. The paper mentions using DUSt3R as the image feature extractor and conducting ablation studies. However, to the best of our knowledge, the baseline methods in the paper do not use DUSt3R as the feature extractor but rather ResNet. Additionally, the description of the PiMAE point cloud feature extractor is too brief.

6. In the experimental setup section, the selection criteria for the registration recall (RR), specifically the thresholds τr and τt, are not clearly explained. It is unclear whether these thresholds are based on real-world application requirements or experimental settings. The rationale and justification for their choice are not provided.

7. In the experimental results section, the paper reports high precision but low recall, which may indicate a large number of false negatives. It is recommended to further analyze the reasons behind this.

Declarations

Potential competing interests: No potential competing interests to declare.