Qeios

Peer Review

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

Varun Ravi Kumar[1]

1. Qualcomm, San Diego, United States

Strengths

- Novel Methodology: The coarse-to-fine matching pipeline, combined with GAL and Inter-Intra Contrastive Learning, addresses key challenges in image-to-point cloud registration, such as viewpoint variations and modality gaps.

- New Dataset: The Carla Intersection Dataset fills a gap in existing datasets by focusing on traffic camera localization in diverse urban and rural environments. This provides a valuable resource for future research.

- Strong Generalization: TrafficLoc demonstrates robust performance across unseen environments (e.g., TestT10 in Town10) and varying camera setups, showcasing its adaptability.

2. Comprehensive Evaluation

- The authors conduct thorough experiments on multiple datasets (Carla Intersection, KITTI, NuScenes, USTC), ensuring the method's applicability to both traffic and in-vehicle cameras.

- Ablation studies effectively isolate the contributions of each component (e.g., GAL, DTA), providing clear evidence of their impact on performance.

3. Technical Rigor

- The methodology is well-designed, leveraging advanced techniques like Transformers, PointNet, and Vision Transformers (ViT).

- The inclusion of dense training alignment and geometry-guided attention loss highlights the authors' attention to detail in addressing common pitfalls in cross-modal registration.

4. Clarity of Presentation

- The paper is well-structured, with clear sections explaining the problem, methodology, experiments, and results.

doi.org/10.32388/AAJMMQ

- Figures (e.g., Figure 2, Figure 6) and tables (e.g., Table 1, Table 2) effectively support the narrative and make complex concepts more accessible.

Weaknesses

1. Ambiguities in Problem Formulation

- The problem statement lacks clarity regarding certain symbols. For instance:

- It is unclear whether $ p_i $ in Equation (1) needs to be transformed into homogeneous coordinates.

- The role of $ K $ (camera intrinsic matrix) in the projection function $ F(\cdot) $ is not explicitly explained, which could confuse readers unfamiliar with camera geometry.

2. Insufficient Explanation of FPS

- While Farthest Point Sampling (FPS) is mentioned for generating super-points, the description of how clustering and mapping are performed after generating $ M $ super-points is vague. This limits reproducibility and understanding of the feature extraction process.

3. Limited Analysis of Super-Point Localization

- The authors use $ L_{\text{det}} $ to locate super-points but do not analyze the correspondence between super-points and image patches. This omission leaves room for questions about the reliability of coarse matching.

4. Efficiency Concerns

- FPS is known to be computationally expensive. The authors do not explore alternative sampling methods or justify why FPS was chosen over more efficient techniques.

5. Baseline Comparisons

- The baselines primarily use ResNet for image feature extraction, while the proposed method leverages DUSt3R. This discrepancy makes it difficult to attribute performance gains solely to TrafficLoc's innovations versus the choice of feature extractor.

- Similarly, the description of PiMAE for point cloud feature extraction is brief and lacks depth.

6. Threshold Selection

- The thresholds $ \tau_r $ and $ \tau_t $ for registration recall (RR) are not justified. It is unclear whether these values are based on practical requirements or arbitrary experimental settings.

7. Precision vs. Recall Trade-off

- While the method achieves high precision, the relatively low recall suggests a significant number of

false negatives. This issue warrants further investigation, particularly in real-world applications where missed detections could have critical consequences.

Suggestions for Improvement

1. Clarify Problem Formulation

- Explicitly define all symbols in the problem statement section. For example:

- State whether $ p_i $ is transformed into homogeneous coordinates.

- Explain the role of $ K $ in the projection process.

2. Expand on FPS

- Provide a more detailed explanation of how FPS is integrated into the pipeline, including clustering and mapping steps. Consider visual aids to clarify this process.

3. Analyze Super-Point Correspondence

- Conduct experiments to evaluate the quality of super-point localization and its impact on coarse matching accuracy.

4. Explore Alternative Sampling Methods

- Compare FPS with other efficient sampling techniques (e.g., random sampling, grid-based sampling) to assess their impact on performance and computational efficiency.

5. Improve Baseline Comparisons

- Ensure a fair comparison by using consistent feature extractors (e.g., DUSt3R for all methods) or clearly stating the rationale for differences.

- Expand the description of PiMAE and its role in point cloud feature extraction.

6. Justify Thresholds

- Explain the rationale behind selecting $ \tau_r $ and $ \tau_t $ for RR. If these thresholds are arbitrary, consider sensitivity analysis to demonstrate robustness.

7. Address Precision-Recall Trade-off

- Investigate the causes of low recall and propose strategies to improve it without compromising precision.

8. Enlarge Dataset

- Increase the size of the Carla Intersection Dataset by simulating additional scenarios. This will enhance the diversity and realism of the dataset, further validating the method's generalization ability.

9. Highlight Novelty

- Summarize the main contributions more succinctly in the introduction to emphasize the novelty and significance of the work.

Overall Assessment

Clarity: 4.5/5

The paper is generally well-written, but some ambiguities in the problem formulation and technical details need clarification.

Novelty: 4.5/5

The proposed method introduces several innovative components (e.g., GAL, Inter-Intra Contrastive Learning) that address key challenges in image-to-point cloud registration.

Impact: 4/5

TrafficLoc demonstrates strong performance and generalization, making it a valuable contribution to the field. However, its reliance on pre-existing 3D maps may limit real-world applicability.

Final Rating: 4.3/5

Conclusion

This paper presents a significant advancement in traffic camera localization, supported by a novel dataset and extensive experimentation. While there are areas for improvement, particularly in clarity and efficiency, the overall quality of the work is high. Addressing the suggested improvements will further enhance the paper's rigor and impact.

Declarations

Potential competing interests: No potential competing interests to declare.