Qeios

Peer Review

Review of: "NeRF-VIO: Map-Based Visual-Inertial Odometry with Initialization Leveraging Neural Radiance Fields"

Youqi Pan[1]

1. Peking University, China

The paper proposes a map-based visual-inertial odometry method that leverages a neural radiance field (NeRF) as a prior map. It uses a multilayer perceptron (MLP) to predict the pose of the first input image within the map for relocalization. During the tracking phase, it builds on the MSCKF framework and employs both captured and rendered images as constraints to achieve accurate pose tracking. Experiments demonstrate that the proposed method addresses the limitation of approaches like iNeRF, which rely on a good initial guess for successful relocalization. Moreover, it outperforms MSCKF in terms of tracking accuracy.

The paper introduces a certain level of innovation by using an MLP for fast relocalization within the NeRF map and by simultaneously employing two types of maps as constraints for VIO. However, the method is not clearly presented, the experiments are insufficient, and there are issues with unfair comparisons.

Method:

1. In Section III.A, Equation (3), the term *SLAM feature state* is not explained. In fact, MSCKF-based VIO systems do not maintain explicit map point states.

2. In Section III.C (page 7), the concept of *left-invariant geodesic distance* is not clearly explained, nor is the rationale for using it as a loss function. Additionally, many symbols in the derivation on pages 8–10 are left undefined (e.g., $\lfloor \omega_{i,s} \times \rfloor$). There is also no ablation study in the experiments to evaluate the effect of this loss function.

3. Sections III.D, E, and F are described too briefly, with no clear explanation of the role each component plays in the overall system.

doi.org/10.32388/UEA1C5

Experimental:

1. The relocalization MLP is trained on one sequence from a specific NeRF map and tested on other sequences from the same map, which does not reflect the model's generalization ability.

2. In Section IV.A, what is the basis for adding initialization noise to iNeRF? Meanwhile, NeRF-VIO is tested without any initialization noise, making the comparison unfair.

3. In Section IV.B, MSCKF is a map-free VIO system and does not rely on prior maps, whereas NeRF-VIO is map-based and estimates poses using a complete map. This makes the comparison between the two methods inherently unfair.

The authors should place greater emphasis on the overall architecture of the system and the relationships between its components, rather than focusing solely on algorithmic details. Additionally, it should be clearly stated that a map-based VIO system is fundamentally different from classical VIO systems, which do not require a prior map. Fair comparisons should be made with other similar map-based approaches.

Declarations

Potential competing interests: No potential competing interests to declare.