Qeios

Peer Review

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

Cesar Debeunne[1]

1. ISAE-SUPAERO, Université de Toulouse, France

Sum-up:

This paper presents a NeRF-based VIO with a novel initialization procedure using a pose estimation MLP. The system works with a pre-trained NeRF of a scene and is initialized by fusing IMU information and the estimation of the pose of the first frame with an MLP trained similarly to the NeRF. The authors propose a geodesic loss on SE(3) with a left-invariant metric to train the pose estimation network, which is stated as a novelty. Once the VIO is initialized, it refines multiple poses in a MSCKF scheme using IMU integration and feature observation from both captured and rendered images. The rendered images that are used for filtering are rendered at the same pose as the closest camera frame. The proposed system is compared to INeRF for pose estimation and to a regular MSCKF for state estimation on an indoor dataset. NeRF-VIO's pose estimator for initialization appears clearly superior to iNeRF, both in terms of accuracy and latency. NeRF-VIO's performance for state estimation outperforms the standard MSCKF and is close to NeRF-VIO with ground truth initialization.

Major Remarks:

- IV. A. One of the contributions is your geodesic loss; you should add an experimental validation of this loss function by comparing pose estimation MLP training with a baseline loss on SE(3).
- IV. B. It is surprising that you don't mention the performance of NeRF-VINS. You use their datasets, and their results are reported in their paper. The system is pretty similar; I believe that you can have interesting interpretations even if it is superior to your system. To me, it is not a no-go to admit that a similar system has better performance than you, as long as you provide a solid analysis of the results.

doi.org/10.32388/D7Q6S9

The main difference between the two is that you render images at the same pose using SSIM to filter dynamic parts of the images, while NeRF VINS renders images with a small baseline to increase the information. Moreover, the contribution is not about the VIO but about the initialization, so you can bring nuance here.

- IV. C. A visual illustration is interesting here, but the best would be to do an ablative study on the proposed dataset: evaluate the performance of NeRF VIO with and without grid-based SSIM.

Minor Remarks:

- Figure 1: This figure is not clear at all. Do you refine the NeRF with the VIO? Do you refine the feature positions? Why is the 1st IMU state in the backend; is it refined continuously? Do you even extract keypoints in the frontend?
- III. E. In Eq (20), you don't take into account the focal length and the center point of the camera? I suppose that you use the bearing vectors of the features then. If so, specify it in the text; otherwise, correct eq (20).
- "are employed in ORB-SLAM[14] and ORB-SLAM2[15]" – now you can even add ORB-SLAM3; you should cite the last update only
- In III. The explanation of the terms in (3) is disturbing as the time step k doesn't appear in (3); adding the time step k in (3) would help.
- III. C. It is not clear if the network returns a pose in SE(3) or in \mathfrak{se}(3).
- III. C. You don't discuss your choice of \mathbf{a} for the geodesic distance.
- III. C. "Since the original data naturally lies in \mathfrak{se}(3)" – to me, there is a mistake here; the original data naturally lies in SE(3).
- IV. In NeRF-VINS, they specify that the NeRF was trained on Table 1 for sequences 1-4 and Table 5 for sequences 5-8. Here, you say that you just train on Table 1; can you explain?

Conclusion:

The contributions of the paper are minor but pretty interesting, especially for the initialization of the VIO. The presentation of the method is clear overall, with a few edits needed to avoid confusion. However, the experimental validation is lacking ablative studies and comparison to other systems. I warmly recommend the author to read carefully the NeRF-VINS paper that presents a similar system but with

way more experimental insights. Major revisions, especially on section IV, are needed for this paper to be suitable for journal publication.

Declarations

Potential competing interests: No potential competing interests to declare.