Review for Learning Monocular Visual Odometry via Self-Supervised Long-Term Modeling

Can Koz

March 28, 2021

1   Summary

Monocular visual odometry aims to find position and direction camera by using the images. Existing systems use geometric methods or learning based methods or combination of both. Since fully supervised methods require data annotation self supervised methods have been proposed. Existing self supervised methods use short snippets so frame to frame estimation of pose suffers from error accumulation. In this paper, authors propose a self supervised visual odometry(VO) method to aggregate long term temporal information.

2   Strengths

- The proposed method combines geometric and learning based approaches and utilizes the the advantages of both techniques.

- The algorithm models long term temporal dependency and prevents error accumulation.

- The method achieves state of the art performance on standard VO datasets in a self supervised manner.

3   Weaknesses

- As authors mentioned in limitations section, the main weakness is the lack of diverse dataset and this can be handled with synthetic data.

- The technique results in large rotational error and authors address this issue with bias in training data.

- Two stage training mechanism makes optimization difficult.

- System fails under the over-exposure scenarios.

4   Evaluation

They use KITTI and TUM RGB-D dataset for evaluation. Based on absolute trajectory RMSE they compare their algorithm with the existing methods. They evaluated different variants of long term modeling and two stage training in ablation study. In KITTI dataset the method outperforms the state of the art methods. IN TUM RGB-D dataset they outperform the supervised baseline V2D but they can not achieve the performance of state of the art supervised methods. The authors attribute this result to lack of unlabeled data.

5 Final Comments and Future Work

Authors developed a self supervised technique which handles long term dependency for pose estimation. They achieved a state of the art performance among self supervised methods and also comparable performance with supervised methods. Instead of two stage training, end to end approach can be developed in future works.