Attention based visual odometry

Jagennath Hari(jh7454) Navoday Borkar(nyb2005) Venkata Amith Palacherla (vp2201)

January 26, 2023

1 Introduction

Odometry plays a crucial role in self-localization techniques used for the autonomous navigation of agents (Robots and Vehicles) in indoor and outdoor environments. A common technique is computing the Visual Odometry(VO) using a sequence of RGB images to determine the pose/odometry of the agent while moving. Traditionally, VO was computed using optical flow(Lucas-Kanade[5] or Gunner Farneback[1]), which commonly fail in featureless environments (e.g. white walls). The newer, learning-based methods are of two types - 1) Act as a supplement to traditional methods, and 2) End-to-End learning-based architecture to compute odometry. DeepVO[3], TartanVO[4], D3VO[6], DAVO[2], and DeepAVO[7] are a few learning-based methods to compute the odometry of the agent. However, even with the advancement of deep learning, the traditional methods outperform the learning based-methods except in a few scenarios. In this project, we propose to build a novel attention-based neural network architecture which will use a sequence of 15-25 RGB images and their Depth map to compute the odometry of the agent. We plan to test the network in a sparse featured environment where traditional methods fail to compute odometry.

2 Deliverable

This work aims to -
1) Create an Attention-based neural network model to compute odometry.
2) Create an indoor sparse feature environment dataset to test VO algorithms.
3) Evaluate and compare the performance with other methods on standard datasets.

3 Datasets

We would like to use the following dataset to train and test the proposed neural network model -
1) Kitti dataset
2) 7 Scenes
3) TUM VI dataset

References

[1] Gunnar Farnebäck. "Two-frame motion estimation based on polynomial expansion". In: *Scandinavian conference on Image analysis*. Springer. 2003, pp. 363–370.

[2] Xin-Yu Kuo et al. "Dynamic Attention-based Visual Odometry". In: *2020 CVPRW* (2020).

[3] Sen Wang et al. "DeepVO: Towards End-to-End Visual Odometry with Deep Recurrent Convolutional Neural Networks". In: *CVPR* (2017).

[4] Wenshan Wang, Yaoyu Hu, and Sebastian A. Scherer. "TartanVO: A Generalizable Learning-based VO". In: *CVPR* abs/2011.00359 (2020).

[5] Liu Xiaochen et al. "Evaluation of Lucas-Kanade based optical flow algorithm". In: *2018 IEEE CSAA Guidance, Navigation and Control Conference (CGNCC)*. 2018, pp. 1–6. DOI: 10.1109/GNCC42960.2018.9018982.

[6] Nan Yang et al. "D3VO: Deep Depth, Deep Pose and Deep Uncertainty for Monocular Visual Odometry". In: *CVPR* abs/2003.01060 (2020).

[7] Ran Zhu et al. "DeepAVO: Efficient Pose Refining with Feature Distilling for Deep Visual Odometry". In: *cvpr* abs/2105.09899 (2021).