INF791 - CV on 3D data

Summary report of selected articles

Werikson Alves - 96708
Federal University of Viçosa
Viçosa, Brasil
e-mails: werikson.alves@ufv.br

1 Trabalhos Relacionados

This report presents a review of five relevant articles in the field of visual SLAM and 3D reconstruction with deep learning. The selected works address different strategies to improve the accuracy, robustness, and generalization of mapping systems based on monocular, stereo, and RGB-D cameras.

BA-Net Tang and Tan [2019] proposes a differentiable optimization network inspired by dense bundle adjustment (BA). The architecture performs joint depth and pose adjustment based on convolutional feature errors. The use of base depth maps reduces the dimensionality of the optimization. Evaluations on ScanNet and KITTI show that BA-Net outperforms classical and supervised approaches in depth and trajectory metrics, highlighting the effectiveness of the end-to-end approach.

DeepFactors Czarnowski et al. [2020] presents a dense and probabilistic monocular SLAM, which uses learned latent representations for optimization on a factor graph. Instead of mapping depth directly, the system operates in a continuous latent space. Tests on ScanNet, TUM, and ICL-NUIM indicate higher accuracy and robustness compared to methods such as CodeSLAM and CNN-SLAM. The modular structure and ability to operate in real time reinforce its practical potential.

D3VO Yang et al. [2020] combines deep estimates of depth, pose, and uncertainty in a direct pipeline for monocular VO. Depth and uncertainties are learned by neural networks and integrated into a direct optimization model. The system achieves competitive results on KITTI and EuRoC, outperforming monocular methods and rivaling stereo and visual-inertial approaches. Robustness in adverse environments highlights the contribution of the uncertainty model.

DROID-SLAM Teed and Deng [2021] is a fully differentiable end-to-end visual SLAM system based on iterative pose and depth updates via Dense Bundle Adjustment (DBA). Trained only with monocular data, it generalizes to stereo and RGB-D inputs. Evaluations on four benchmarks (TartanAir, EuRoC, TUM-RGBD, ETH-3D) demonstrate superior performance in trajectory error and reconstruction. It stands out for its robustness, accuracy, and scalability.

NeRF-SLAM Rosinol et al. [2023] integrates monocular SLAM estimates with neural radiance fields (NeRF). It uses depth and pose information (via DROID-SLAM) as uncertainty-weighted supervision to optimize a NeRF model in real time. Results on the Cube-Diorama and Replica datasets show significant gains in visual fidelity (PSNR) and geometric accuracy, even with noisy data. Despite high memory consumption, it is promising for AR/VR and robotic inspection.

2 Justification for choosing the article

The article chosen for presentation was *DROID-SLAM: Deep Visual SLAM for Monocular, Stereo, and RGB-D Cameras*, due to its innovative proposal for integrating deep learning and differentiable geometric optimization. In addition to achieving superior results in key benchmarks, the system generalizes to multiple input modalities and operates end-to-end. Its methodology is strongly aligned with the project under development, which also involves monocular

3D reconstruction and data fusion for dense mapping. Thus, DROID-SLAM represents a strategic choice to deepen the technical discussion and connect theory and practice in the context of the INF791 course.

Referências

Jan Czarnowski, Tristan Laidlow, Ronald Clark, and Andrew J. Davison. Deepfactors: Real-time probabilistic dense monocular slam. *IEEE Robotics and Automation Letters*, 5(2):721–728, April 2020. ISSN 2377-3766. doi: 10.1109/LRA.2020.2965415.

Antoni Rosinol, John J. Leonard, and Luca Carlone. Nerf-slam: Real-time dense monocular slam with neural radiance fields. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3437–3444, Oct 2023. doi: 10.1109/IROS55552.2023.10341922.

Chengzhou Tang and Ping Tan. Ba-net: Dense bundle adjustment network, 2019. URL https://arxiv.org/abs/1806.04807.

Zachary Teed and Jia Deng. Droid-slam: Deep visual slam for monocular, stereo, and rgb-d cameras. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 16558–16569. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/89fcd07f20b6785b92134bd6c1d0fa42-Paper.pdf.

Nan Yang, Lukas von Stumberg, Rui Wang, and Daniel Cremers. D3vo: Deep depth, deep pose and deep uncertainty for monocular visual odometry. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.