

Relatório Resumido dos artigos selecionados

Werikson Alves
Universidade Federal de Viçosa
Viçosa, Brasil
e-mails: werikson.alves@ufv.br

1 Trabalhos Relacionados

Este relatório apresenta uma revisão de cinco artigos relevantes na área de SLAM visual e reconstrução 3D com aprendizado profundo. Os trabalhos selecionados abordam diferentes estratégias para melhorar a acurácia, robustez e generalização de sistemas de mapeamento baseados em câmeras monoculares, estéreo e RGB-D.

BA-Net Tang and Tan [2019] propõe uma rede com otimização diferenciável inspirada em bundle adjustment (BA) denso. A arquitetura realiza ajuste conjunto de profundidade e pose, com base em erros de características convolucionais. O uso de mapas de profundidade base reduz a dimensionalidade da otimização. Avaliações em ScanNet e KITTI mostram que BA-Net supera abordagens clássicas e supervisionadas em métricas de profundidade e trajetória, destacando a eficácia da abordagem end-to-end.

DeepFactors Czarnowski et al. [2020] apresenta um SLAM monocular denso e probabilístico, que utiliza representações latentes aprendidas para otimização sobre um grafo de fatores. Em vez de mapear profundidade diretamente, o sistema opera em um espaço latente contínuo. Testes em ScanNet, TUM e ICL-NUIM indicam maior precisão e robustez em relação a métodos como CodeSLAM e CNN-SLAM. A estrutura modular e a capacidade de operar em tempo real reforçam seu potencial prático.

D3VO Yang et al. [2020] combina estimativas profundas de profundidade, pose e incerteza em um pipeline direto para VO monocular. A profundidade e incertezas são aprendidas por redes neurais, e integradas a um modelo direto de otimização. O sistema atinge resultados competitivos em KITTI e EuRoC, superando métodos monoculares e rivalizando com

abordagens estéreo e visual-inerciais. A robustez em ambientes adversos evidencia a contribuição do modelo de incerteza.

DROID-SLAM Teed and Deng [2021] é um sistema de SLAM visual end-to-end totalmente diferenciável, baseado em atualizações iterativas de poses e profundidades via Dense Bundle Adjustment (DBA). Treinado apenas com dados monoculares, generaliza para entradas estéreo e RGB-D. Avaliações em quatro benchmarks (TartanAir, EuRoC, TUM-RGBD, ETH-3D) demonstram desempenho superior em erro de trajetória e reconstrução. Destaca-se pela robustez, precisão e escalabilidade.

NeRF-SLAM Rosinol et al. [2023] integra estimativas de SLAM monocular com campos de radiância neurais (NeRF). Utiliza informações de profundidade e pose (via DROID-SLAM) como supervisão ponderada por incerteza para otimizar um modelo NeRF em tempo real. Resultados nos datasets Cube-Diorama e Replica mostram ganhos expressivos em fidelidade visual (PSNR) e precisão geométrica, mesmo com dados ruidosos. Apesar do alto consumo de memória, é promissor para AR/VR e inspeção robótica.

2 Justificativa da escolha do artigo

O artigo escolhido para apresentação foi o *DROID-SLAM: Deep Visual SLAM for Monocular, Stereo, and RGB-D Cameras*, por sua proposta inovadora na integração entre aprendizado profundo e otimização geométrica diferenciável. Além de alcançar resultados superiores nos principais benchmarks, o sistema generaliza para múltiplas modalidades de entrada e opera de forma end-to-end. Sua metodologia está

fortemente alinhada com o projeto em desenvolvimento, que também envolve reconstrução 3D monocular e fusão de dados para mapeamento denso. Assim, o DROID-SLAM representa uma escolha estratégica para aprofundar a discussão técnica e conectar teoria e prática no contexto do curso INF791.

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.