Detailed Project Proposal (DPP)

Leveraging Deep Features for ORB-SLAM3
(DXSLAM-ORB3)

Student Name: Adwaith Kallungal Vrundavanan
Student ID: 22061390
MSc in AI and Robotics
Supervisor: Dr Zoe Jeffrey
16-06-24

1 Aim of the project

The aim of this project is to enhance ORB-SLAM3 by integrating HF-Net as the feature extractor, thereby improving its robustness and accuracy in complex environments. This will enable more reliable and efficient SLAM performance in diverse and challenging scenarios.

Short Description of the idea

Simultaneous Localisation and Mapping (SLAM) is a problem that has made great improvements over the last decade. In the world of robotic, feature-based visual SLAM algorithms reign supreme. They're efficient, allowing robots to navigate smoothly, and adaptable, making them perfect for long-term missions. But the existing visual SLAM algorithms use handcrafted visual features like SIFT (Lowe, 2004), Shi-Tomasi (J. Shi and Tomasi, 1994) and ORB (Ethan, 2011) which fails to extract features in complex environments. Several studies (Mur-Artal and Tardós, 2017; X. Shi et al., 2020) have identified limitations in ORB-SLAM2's ability to re-localize in environments with significant scene or viewpoint changes.

Recent developments in deep learning has seen great results with pixel-wise feature extractors (DeTone, Malisiewicz, and Rabinovich, 2018; Dusmanu et al., 2019; Tang et al., 2019) which are more robust in extracting features even in complex conditions. While ORB-SLAM3 (Campos et al., 2021) represents a state-of-the-art visual SLAM algorithm, it utilizes the aforementioned ORB feature extraction, leading to limitations in complex scenarios.

This project proposes an improvement to ORB-SLAM3 by integrating HF-Net (Sarlin et al., 2019), a deep learning-based feature extractor. Li et al. (2020) demonstrated improved performance over ORB-SLAM2 by utilizing HF-Net. This project aims to replicate and potentially surpass those results by integrating HF-Net into ORB-SLAM3.

2 Research Questions

- Can replacing the handcrafted feature extraction in ORB-SLAM3 with the deep learning-based HF-Net lead to improved performance in terms of accuracy, robustness, and efficiency?
- How does the performance of ORB-SLAM3 integrated with HF-Net compare to the original ORB-SLAM3?

Project Objective

- Integrate HF-Net as the primary feature extractor in ORB-SLAM3 to improve robustness in complex environments.
- Validate the improvement through a series of benchmark tests comparing the enhanced ORB-SLAM3 with the original version.

Project Plan

Tasks:

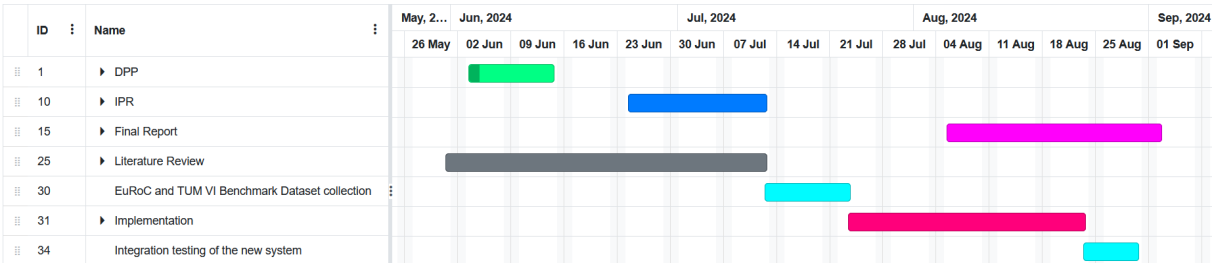

Figure 1: Gantt chat of project plan

1. Literature Review (31-05-2024:12-07-2024)
   (a) Understanding SLAM and Monocular SLAM (31-05-2024:04-06-2024)
   (b) Understanding ORB SLAM 1,2,3 (04-06-2024:06-06-2024)
   (c) Understanding SP SLAM, DX SLAM (06-06-2024:14-06-2024)
   (d) Understand ORB SLAM 3 code base (14-06-2024:02-07-2024)
   (e) Understand HF-Net (03-07-2024:12-07-2024)
2. EuRoC and TUM VI Benchmark Dataset collection (12-07-2024:23-07-2024)
3. Implementation (23-07-2024:23-08-2024)
   (a) Serialising the HF-Net (23-07-2024:31-07-2024)
   (b) writing the wrapper for serialised model in C++ (31-07-2024:08-08-2024)
   (c) Unit Testing HF-Net (09-08-2024:12-08-2024)
   (d) Intergrate the Model with ORB slam 3 (12-08-2024:23-08-2024)
4. Integration testing of the new system (23-08-2024:26-08-2024)
5. Validation of the new system with benchmark datasets (26-08-2024:30-08-2024)

Milestones:

- Serialising the HF-Net
- Integrating HF-Net to ORB-SLAM3
- Validation of the new system with benchmark datasets

References

Campos, Carlos et al. (2021). "Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam". In: *IEEE Transactions on Robotics* 37.6, pp. 1874–1890.

DeTone, Daniel, Tomasz Malisiewicz, and Andrew Rabinovich (2018). "Superpoint: Self-supervised interest point detection and description". In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 224–236.

Dusmanu, Mihai et al. (2019). "D2-net: A trainable cnn for joint detection and description of local features". In: *arXiv preprint arXiv:1905.03561*.

Ethan, Rublee (2011). "ORB: An efficient alternative to SIFT or SURF". In: *ICCV, 2011*.

Li, Dongjiang et al. (2020). "DXSLAM: A robust and efficient visual SLAM system with deep features". In: *2020 IEEE/RSJ International conference on intelligent robots and systems (IROS)*. IEEE, pp. 4958–4965.

Lowe, David G (2004). "Distinctive image features from scale-invariant keypoints". In: *International journal of computer vision* 60, pp. 91–110.

Mur-Artal, Raul and Juan D Tardós (2017). "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras". In: *IEEE transactions on robotics* 33.5, pp. 1255–1262.

Sarlin, Paul-Edouard et al. (2019). "From coarse to fine: Robust hierarchical localization at large scale". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 12716–12725.

Shi, Jianbo and Tomasi (1994). "Good features to track". In: *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 593–600. DOI: 10.1109/CVPR.1994.323794.

Shi, Xuesong et al. (2020). "Are we ready for service robots? the openloris-scene datasets for lifelong slam". In: *2020 IEEE international conference on robotics and automation (ICRA)*. IEEE, pp. 3139–3145.

Tang, Jiexiong et al. (2019). "GCNv2: Efficient correspondence prediction for real-time SLAM". In: *IEEE Robotics and Automation Letters* 4.4, pp. 3505–3512.