Visual SLAM for Drones

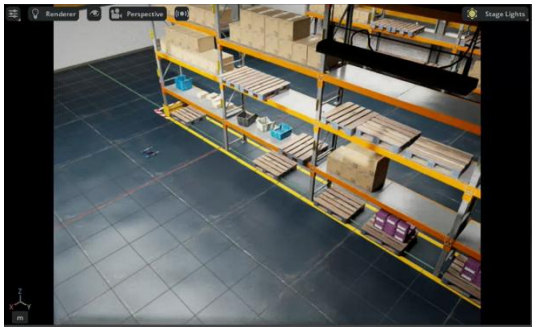
Sameh Lahouar

Professor : Alcherio Martinoli
Assistant(s) : Alexander Wallen Kiessling

The goal of this project was to benchmark and compare state-of-the-art VIO and VI-SLAM algorithms under a unified and reproducible evaluation framework. Three representative systems were selected: ORB-SLAM3, OKVIS2, and OpenVINS, covering optimization-based SLAM approaches with loop closure as well as filtering-based odometry.

A unified benchmarking pipeline was developed to evaluate all algorithms under identical conditions. On the EuRoC MAV dataset, selected Machine Hall and Vicon Room sequences of varying difficulty were processed using each algorithm's native data interface (ASL format for ORB-SLAM3 and OKVIS2, ROS 2 bags for OpenVINS).

To complement real-world benchmarking, a simulation pipeline was implemented using NVIDIA Isaac Sim and the PegasusSimulator framework. A quadrotor equipped with a monocular camera and IMU was simulated in a warehouse environment. Camera and IMU data were published as ROS 2 topics, recorded as bags, and converted to the required formats to reuse the same evaluation pipeline as for real data.



Simulation environment used in the experiments

Trajectory accuracy was evaluated using the evo toolkit, focusing on Absolute Pose Error (APE) for global accuracy and Relative Pose Error (RPE) for local motion consistency. Additional metrics were introduced to provide deeper insight into algorithm behavior, including CPU and RAM usage and feature tracking statistics over time.

The quantitative and qualitative benchmarking results on the EuRoC MAV dataset reveal clear and consistent performance differences between the evaluated algorithms. Table 1 summarizes the overall observed behavior across accuracy, robustness, and computational cost.

Criterion	ORB-SLAM3	OKVIS2	OpenVINS
Global accuracy (APE)	✓	✓	✗
Local consistency (RPE)	✓	✓✓	✓
Robustness	✓✓	✓	✗
Computational efficiency	✗	✗	✓✓
Loop closure capability	✓	✓	✗

Summary of observed algorithmic characteristics

As shown in the table, ORB-SLAM3 and OKVIS2 consistently outperform OpenVINS in terms of global accuracy and robustness, particularly on medium and difficult sequences involving aggressive motion and fast rotations. OpenVINS, while computationally efficient, exhibits significant drift accumulation and fails on the most challenging trajectories due to the absence of loop closure.

Between the two VI-SLAM systems, ORB-SLAM3 achieves the strongest global accuracy (lowest APE), benefiting from map reuse and loop closure, whereas OKVIS2 demonstrates superior local motion consistency (lowest RPE), reflecting accurate short-term estimation. These accuracy gains come at the cost of higher CPU and memory usage compared to OpenVINS.

Overall, the results highlight a fundamental trade-off: filtering-based VIO offers low computational cost but limited robustness, while optimization-based VI-SLAM achieves higher accuracy and reliability, particularly for long trajectories and aggressive motion, at increased computational expense. Consequently, no single algorithm is optimal across all criteria, and the appropriate choice depends on motion profile, robustness requirements, and available computational resources.