COMPARISON OF DSO AND ORBSLAM3 IN INDOOR ENVIRONMENTS USING MONOCULAR VISUAL SLAM

Qiyao Yang

Belarusian State University of Informatics and Radioelectronics
Minsk, Republic of Belarus

Jun Ma – Assistant

Abstract. This thesis presents a comparative study of ORB-SLAM3 and DSO in indoor environments. We first validate the local implementation on public datasets, then analyze self-recorded sequences from four perspectives: trajectory behavior, summary statistics, average per-frame runtime, and mapping visualization. The results show that ORB-SLAM3 is stronger in pose estimation and real-time efficiency, while DSO provides more informative semi-dense map appearance. Both methods degrade under low light and strong highlights, but with different failure modes, indicating clear complementarity. The study also notes key limitations, including the absence of ground truth for custom data, limited run repetitions, and acquisition choices that increase loop-closure difficulty. Future work will expand scene coverage and sample size, improve the evaluation protocol, and explore a hybrid framework that combines the strengths of both approaches.*Keywords:* semantic communication, vehicle detection, YOLOv12, CNN encoder, supernet, AWGN channel, compression ratio.

Keywords. DSO, ORB-SLAM3, monocular SLAM, indoor localization, photometric calibration, TUM RGB-D, EuRoC MAV, self-recorded data.

Introduction

Visual Simultaneous Localization and Mapping (SLAM) is a fundamental technology in robotics and computer vision that enables devices to estimate their ego-motion and simultaneously reconstruct the structure of unknown environments from image sequences [1]. It has found widespread applications in areas such as autonomous driving, indoor mobile robotics, augmented reality (AR), and virtual reality (VR). Among these, indoor environments are particularly important due to the growing demand for indoor robot navigation, warehouse automation, and AR-based indoor guidance systems. However, indoor scenes pose significant challenges for visual SLAM due to their diverse textures, varying illumination conditions, and potential dynamic objects [2].

Two prominent open-source visual SLAM systems are Direct Sparse Odometry (DSO) [3] and ORB-SLAM3 [4], representing direct and feature-based methods, respectively. DSO minimizes photometric errors without feature extraction, making it advantageous in low-texture or motion-blurred environments [3]. ORB-SLAM3 supports monocular, stereo, RGB-D, and visual-inertial modes, using ORB features for tracking, mapping, and loop closure to achieve high accuracy and robustness [4]. Their monocular versions are particularly attractive for low-cost and easy-to-deploy applications.

Experiments

To further characterize the computational cost of our self-collected sequences, we report the average per-frame tracking time of ORB-SLAM3 and DSO (mode = 0 / mode = 1) on the 509 and corridor scenarios under day and night illumination. The statistics show that ORB-SLAM3 remains near 8.3–9.2 ms per frame in all four conditions and is substantially faster than DSO; both DSO configurations span 19.2–28.9 ms per frame, i.e., roughly 2.5–3× the ORB-SLAM3 timings. This gap is consistent with the differing computational structures: DSO forms photometric residuals over many high-gradient samples inside a sliding window and performs iterative optimization, yielding a larger per-frame problem; ORB-SLAM3 relies mainly on feature extraction, matching, and comparatively light backend updates, and therefore exhibits lower time cost on our hardware and image-resolution settings.

Within DSO, mode = 1 is never faster than mode = 0 across all scenes; the largest gap appears on the corridor day sequence (≈25.0 ms/frame vs. 28.9 ms/frame), suggesting that disabling fixed photometric calibration and relying on online affine brightness compensation may increase the number of optimization variables or iterations and thus raise the average latency. Day–night changes barely affect ORB-SLAM3, whose curve is almost flat; DSO decreases to some extent at night on both 509 and the corridor, with a more pronounced drop on corridor night. This may reflect fewer reliable gradients under low light, changes in keyframe triggering, or a smaller active set of residuals in the window, shrinking each optimization step and lowering the per-frame average. Lower latency must not be equated with higher accuracy; interpretation should be combined with trajectory continuity, mapping density, and failure rates reported elsewhere.

Overall, the figure quantifies–in terms of wall-clock cost–the price of denser geometric and photometric modeling: ORB-SLAM3 is more lightweight per frame, whereas DSO trades higher per-frame cost for richer direct-map information, with mode = 0 slightly favoring average efficiency over mode = 1.

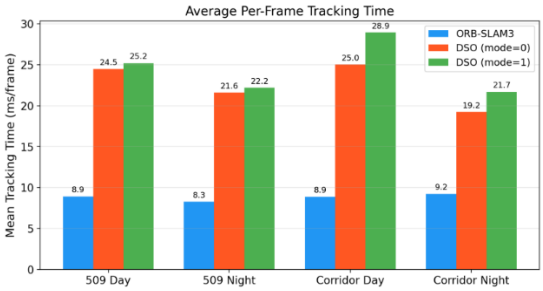

Figure 1 – Average Per-Frame Tracking Time of ORB-SLAM3 and DSO Under Different Scenarios

Table 3 – Five-time Average Measured Performance on Corridor Sequence

Metric	ORB-SLAM3 (Day/Night)	DSO mode=0 (Day/Night)	DSO mode=1 (Day/Night)
Input Frames	1805 / 1861	1804 / 1860	1804 / 1860
Output Poses	248 / 206	279 / 208	289 / 212
Average Points per Frame	212.4 / 214.3	2639.8 / 2631.0	2707.2 / 2667.9

Table 2 – Five-time Average Measured Performance on Classroom 509 Sequence

Metric	ORB-SLAM3 (Day/Night)	DSO mode=0 (Day/Night)	DSO mode=1 (Day/Night)
Input Frames	2302 / 1551	2301 / 1550	2301 / 1550
Output Poses	322 / 201	352 / 268	362 / 272
Average Points per Frame	252.1 / 218.7	2404.9 / 2065.1	2436.1 / 2035.0

Conclusion

The experiments first complete a baseline validation of the local implementation on public datasets, then turn to self-recorded sequences. This order gives the subsequent claims a clear traceability: the public benchmarks check that the pipeline, parameters, and runtime environment are sound; the custom sequences expose how the algorithms behave in realistic indoor settings. Following this protocol, the paper compares methods along four dimensions–trajectory shape, summary statistics, average per-frame time, and mapping visualization–to form a relatively complete evaluation framework.

On the self-recorded data, ORB-SLAM3 and DSO show a fairly clear division of roles. ORB-SLAM3 is overall stronger in pose estimation and trajectory stability: when illumination and texture are favorable, the estimated path aligns better with the actual motion and global consistency is easier to maintain. DSO tends to produce visually richer maps with more structural detail; its semi-dense output makes scene contours and local texture easier to perceive. Neither is free of failure modes: ORB-SLAM3 is more prone to tracking loss and map resets under night conditions and strong specular highlights; without loop closure, DSO can drift more noticeably and may also fail when facing very bright regions. The paper therefore does not frame the comparison as a simple winner–loser ranking, but as complementary strengths in localization robustness versus map interpretability.

In terms of computational cost, ORB-SLAM3's average per-frame tracking time is lower than DSO's throughout, consistent with the different front-end workload and optimization scales of feature-based and direct formulations. That suggests ORB-SLAM3 is more likely to meet tight latency budgets, whereas DSO trades higher per-frame cost for denser scene expression. Taken together with the trajectory and mapping results, system design should trade off localization stability, runtime, and map detail in a task-driven way rather than relying on a single score.

References:

1. Cadena C., Carlone L., Carrillo H. et al. // IEEE Transactions on Robotics. 2016. Vol. 32. P. 1309–1332.
2. Mur-Artal R., Tardós J. D. // IEEE Transactions on Robotics. 2017. Vol. 33. P. 1067–1080.
3. Engel J., Koltun V., Cremers D. // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2018. Vol. 40. P. 184–197.
4. Mur-Artal R., Campos C. // IEEE Transactions on Robotics. 2021. Vol. 37. P. 1874–1890.