Multi-sensor Fusion via Smoothing and Mapping for Harsh Environments

*

Abhishek Mathur
MRSD
Carnegie Mellon University
Pittsburgh, US
armathur@andrew.cmu.edu

Rodrigo Lopes Catto
MRSD
Carnegie Mellon University
Pittsburgh, US
rlopesca@andrew.cmu.edu

Yogita Choudhary
MSR
Carnegie Mellon University
Pittsburgh, US
ychoudha@andrew.cmu.edu

I. Introduction

Simultaneous Localization and Mapping (SLAM) remains a fundamental researched problem in robotics. It enables autonomous mobile systems to navigate and understand previously unseen environments. The complexity of SLAM arises from the need to estimate both a robot's trajectory and the map of the surrounding environment simultaneuously, despite uncertainties in motion and perception. Although significant progress has been made in SLAM over the past two decades, deploying SLAM in unstructured and dynamically changing environments still poses considerable challenges. These challenges become more prominent when the robot is required to operate in real-world conditions characterized by limited visibility, uneven terrain, sensor noise, and sparse perceptual cues.

While existing state-of-the-art SLAM methods have demonstrated robust performance in well-defined and structured environments such as urban streets, indoor office spaces, and academic campus datasets, they often rely on the presence of clearly identifiable landmarks and predictable motion dynamics. Techniques such as ORB-SLAM, DSO, and LIO-SAM have achieved remarkable accuracy under such assumptions. However, these assumptions often break down under natural or adverse conditions, such as forest trails, caves, or disaster zones, where GPS may not be available, lighting conditions may vary drastically, and visual features may be scarce or highly repetitive. In these cases, SLAM systems can suffer from poor localization accuracy, inconsistent map quality, or even catastrophic failures due to poor data association or loop closure errors. Therefore, developing robust SLAM systems that can generalize to such challenging and dynamic settings is critical to ensuring reliable navigation in real-world environments.

Reliable navigation and mapping under adverse conditions demand SLAM systems that are not only accurate and consistent but also adaptive to rapid changes in environmental conditions, sensor quality, and motion dynamics. To address these demands, our project focuses on advancing robust multi-sensor fusion techniques tailored specifically for SLAM in such operationally harsh settings. We believe that leveraging the complementary strengths of different sensor modalities—such as cameras, LiDARs, and IMUs—can significantly enhance system robustness by compensating for the weaknesses of individual sensors under specific environmental conditions.

In this project, we aim to build upon and extend the existing work on LVIO-SAM [3], which integrates stereo camera, LiDAR, and IMU data through factor graph optimization with preintegrated IMU measurements. Similarly, LVI-SAM [13] offers a tightly-coupled framework that fuses visual, inertial, and LiDAR data to provide robust mapping capabilities in real-time. While both of these systems show promising results and achieve state-of-the-art performance, they have primarily been evaluated on datasets collected in structured and relatively benign environments such as the CMU Campus and KITTI datasets. This limited diversity in testing environments may not sufficiently stress the system's ability to handle the broad spectrum of real-world operational conditions.

Therefore, our work seeks to extend these frameworks to operate effectively in more diverse and difficult environments. In particular, we focus on the M2P2 dataset [4], which includes multimodal sensor data collected across various terrain types (paved, on-trail, and off-trail) and lighting conditions (well-lit, low-light, and complete darkness). The **major contributions** for our course project our as follows:

- A comparative performance evaluation of the LVI-SAM and LVIO-SAM algorithms was carried out using the KITTI dataset, which represents structured environmental conditions.
- The LVI-SAM was previously evaluated on the custom datasets used by the authors and was not evaluated on KITTI dataset. This work presents the first (to our knowledge) systematic assessment of LVI-SAM on KITTI dataset.
- Performance analysis was further extended by benchmarking LVI-SAM and LVIO-SAM on the M2P2 dataset, characterizing their efficacy in unstructured environments.

II. Background and Related Work

In the field of sensor odometry, various approaches have been developed to enhance accuracy and robustness. Visual-inertial odometry (VIO) systems, such as the Multi-State Constraint Kalman Filter (MSCKF) [5], and keyframe-based methods [6], have demonstrated effective fusion of visual and inertial data. Optimization-based VIO systems, including VINS-Mono [7], further improved state estimation through nonlinear optimization techniques. Lidar-based odometry methods, such as LOAM [8], and LeGO-LOAM [9], have provided low-drift and real-time solutions by leveraging geometric information from lidar scans. The integration of visual and lidar data has been explored in systems like LIMO [10], and DVL-SLAM [11], which enhanced odometry by combining sparse depth information with visual inputs. More recent advancements include LIO-SAM by [12], which tightly couples lidar and inertial measurements using smoothing and mapping techniques. Based on this, LVI-SAM [13] integrates visual, lidar, and inertial data within a factor graph framework, achieving robust and accurate state estimation. LVI-SAM integrates a visual-inertial subsystem (VIS) and a lidar-inertial subsystem (LIS), combining their outputs to enhance localization. In contrast, LVIO-SAM [3] fuses stereo camera, LiDAR, and IMU data within a single unified system using a centralized factor graph. Unlike LVI-SAM, which processes IMU data in both VIS and LIS, LVIO-SAM utilizes it only once, reducing redundancy.

III. Methods

A. LVIO SAM

In LVIO SAM, the robot's state and its trajectory are estimated using sensor measurements by posing the problem as a maximum a posteriori (MAP) problem. The factor graph contains four types of factors: lidar odometry factors, IMU preintegration factors, visual odometry factors, and a state prior factor. The factor graph is optimized with GTSAM [15] using incremental smoothing and mapping [26]. The preintegrated motion estimation from the IMU helps correct the skew in the point cloud and provides initial estimates for optimizing the LiDAR odometry. This refined LiDAR odometry is then used to estimate the IMU bias and serves as the initial guess for visual odometry triangulation. The visual odometry, in turn, acts as a between factor in the system's overall motion estimation. Fig. 1 shows the high-level block diagram for the LVIO-SAM algorithm.

B. LVI SAM

In LVIO SAM, the robot's state and trajectory are estimated using lidar, visual, and inertial measurements by formulating the problem as a maximum a posteriori (MAP) optimization over a factor graph. The system, LVI-SAM, is divided into a visual-inertial system (VIS) and a lidar-inertial system (LIS), each capable of independent operation while leveraging estimates from the other. Visual odometry is computed by minimizing residuals from visual reprojection and IMU preintegration, with depth information for visual features optionally extracted from lidar frames. Lidar odometry is obtained by scan-matching lidar features to a sliding-window feature map, aided by de-skewing with IMU data. The VIS provides initial guesses for the LIS scan matching, and the LIS provides initialization support for the VIS. Loop closures are first detected visually and then refined through lidar matching. All constraints from visual odometry, lidar odometry, IMU preintegration, and loop closures are jointly optimized using iSAM2. Fig. 2 shows the high-level block diagram for the LVIO-SAM algorithm.

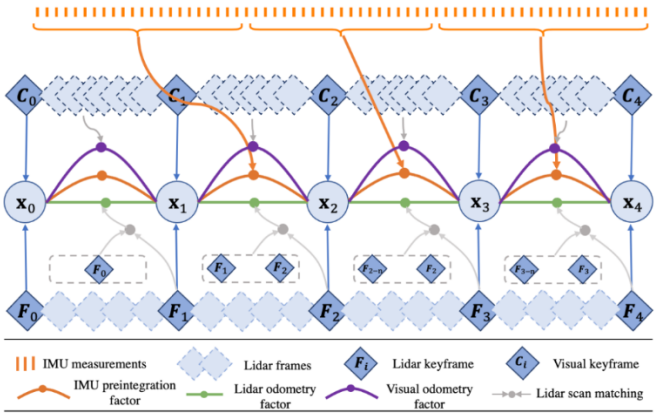

Fig. 1. Method for LVIO-SAM

IV. Code Used

For this project, we leveraged the open-source repository code LVIO-SAM [16], a tightly-coupled lidar-visual-inertial odometry system based on factor graph optimization, and LVI-SAM [13], which extends similar principles for robust mapping in diverse environments. These repositories provide a strong baseline for state estimation by tightly integrating LiDAR, camera, and IMU data to optimize pose trajectories and map consistency over time. The systems are implemented primarily in C++ within the Robot Operating System (ROS) framework, specifically targeted at ROS Noetic. They make use of key third-party libraries, including GTSAM for implementing the underlying factor graph optimization and IMU preintegration algorithms, as well as OpenCV for processing and managing image data for feature tracking and visual odometry.

To adapt the code for our experimental needs, we began by setting up a ROS workspace and resolving all package dependencies to ensure compatibility with the ROS Noetic environment. We made minor but essential modifications to configuration and launch files to support the specific formatting and calibration parameters required by the M2P2 dataset. These adjustments included tuning IMU noise parameters, camera intrinsics, LiDAR range settings, and feature extraction thresholds to improve performance under varying lighting and terrain conditions. Furthermore, we developed additional

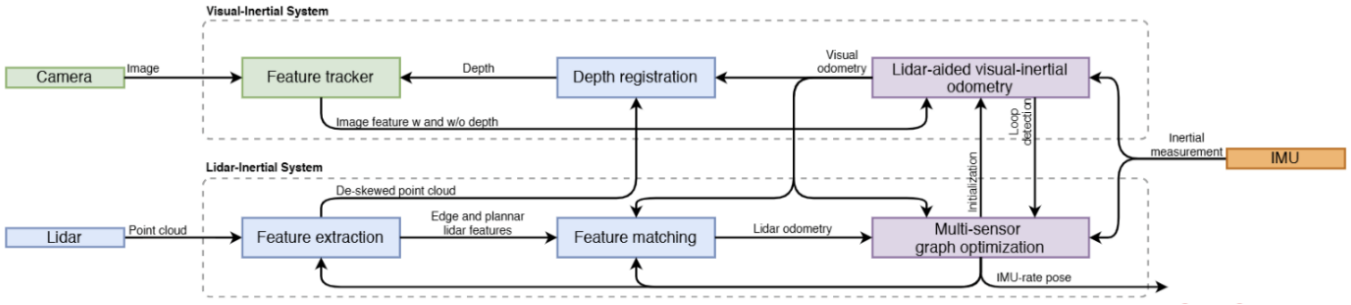

Fig. 2. Method for LVI-SAM

ROS scripts and utilities to automate the process of dataset playback, rosbag file management, and data synchronization, which streamlined the testing pipeline and enabled repeatable experiments.

In parallel, we modified and extended the visualization tools provided in the repository to support clearer and more informative qualitative analysis. These changes included the addition of trajectory overlays, error heatmaps, and visual indicators for loop closure events within RViz. We also logged intermediate states and pose estimates for quantitative analysis, enabling side-by-side comparisons of estimated and ground truth trajectories. These enhancements facilitated a more rigorous evaluation of the SLAM system's performance across different environmental challenges, such as transitions between lighting regimes and terrain types.

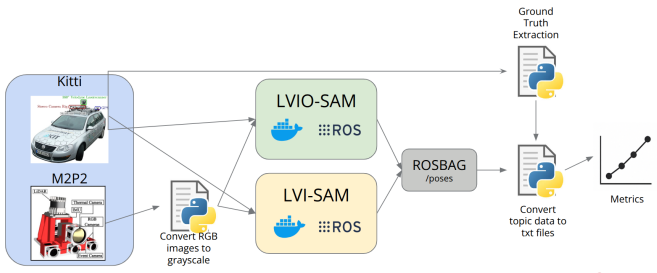

Fig. 3. Method for code implementation

V. Dataset Overview

We used the KITTI Odometry Dataset as our primary benchmark, specifically one of the ROS bag files provided in the LVIO-SAM repository. KITTI offers high-quality, time-synchronized LiDAR, stereo camera, and IMU data, making it suitable for evaluating multi-sensor SLAM systems.

To test generalization, we plan to incorporate the M2P2 Dataset [4], which contains a diverse set of indoor and outdoor environments. M2P2 Dataset offers challenging scenarios and realistic robot motion profiles. It includes RGB-D and LiDAR sensor data along with ground truth poses, making it ideal for evaluating robustness.

VI. Experiments

Our experiments aim to thoroughly evaluate and compare the performance of LVIO-SAM and LVI-SAM across structured and unstructured environments. We designed a multi-stage evaluation process involving dataset preparation and results analysis. Fig 3 details the evaluation pipeline for our implementation.

Initially, we processed raw ROSBAG data from both KITTI and M2P2 datasets. This involved converting stereo RGB images to grayscale for uniform feature extraction, renaming the topics, and extracting odometry poses into compatible text formats for quantitative evaluation.

For evaluation, we employed two widely used SLAM performance metrics:

- **Absolute Trajectory Error (ATE)**: Measures global consistency of estimated trajectories compared to ground truth.
- **Relative Pose Error (RPE)**: Measures local drift between consecutive poses, assessing motion estimation accuracy.

These metrics were computed using a custom Python script [19]. The systems were tested with consistent hyperparameter settings for fair comparison:

- IMU noise standard deviation set to 0.01.
- Visual feature tracking over a sliding window of 10 frames.
- New keyframes inserted when translation exceeds 0.5 meters or rotation exceeds 10 degrees.

KITTI dataset experiments served as the structured environment benchmark, while M2P2 experiments evaluated performance under unstructured, challenging conditions. The latter involved additional complexities, such as incomplete ground truth availability and environmental variability.

The analysis combines both quantitative error metrics and qualitative inspection of generated maps and trajectories to provide a holistic view of system performance under varying conditions. Fig. 4 and Fig. 5 show the snippets of generated maps from LVIO/LVI-SAM for KITTI and M2P2 datasets respectively.

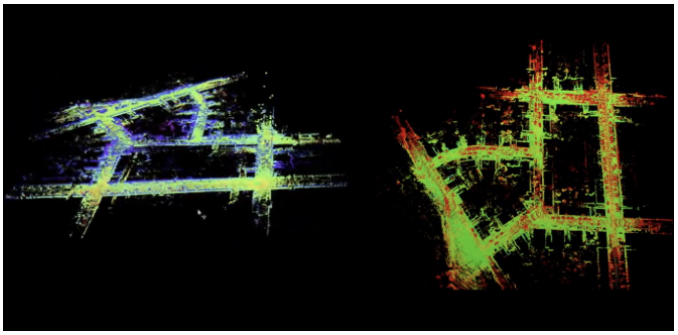

Fig. 4. Kitti Result: LVIO-SAM on the left and LVI-SAM on the right

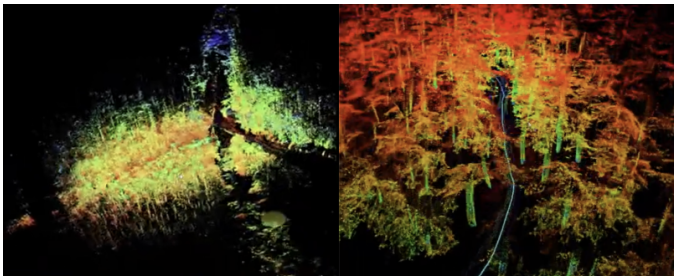

Fig. 5. M2P2 Results: LVIO-SAM on the left and LVI-SAM on the right

VII. Quantitative Evaluation of Results

Our analysis is structured around two primary datasets: the KITTI dataset representing structured urban environments, and the M2P2 dataset representing unstructured, harsh environments. This approach enables a comprehensive evaluation of the systems' robustness and generalization ability.

A. KITTI Dataset

Quantitative results for KITTI show:

- **LVI-SAM**:
  - ATE: mean = 1.27 m, std = 0.53 m, max = 2.25 m
  - RPE: mean = 1.84 m, std = 0.55 m, max = 4.09 m
- **LVIO-SAM**:
  - ATE: mean = 1.24 m, std = 0.53 m, max = 2.24 m
  - RPE: mean = 1.85 m, std = 0.56 m, max = 6.21 m

Fig. 6 shows the estimated and ground truth trajectory plots for LVI-SAM and LVIO-SAM on the KITTI Dataset.

B. M2P2 Dataset

Quantitative results for M2P2 show:

- **LVI-SAM**:
  - ATE: mean = 8.33 m, std = 3.55 m, max = 13.05 m
  - RPE: mean = 0.77 m, std = 0.36 m, max = 1.55 m
- **LVIO-SAM**:
  - ATE: mean = 8.13 m, std = 3.62 m, max = 12.51 m
  - RPE: mean = 0.75 m, std = 0.35 m, max = 1.55 m

Fig. 7 shows the estimated and ground truth trajectory plots for LVI-SAM and LVIO-SAM on the M2P2 Dataset.

Although both systems demonstrate strong performance on the KITTI dataset, LVIO-SAM slightly outperforms LVI-SAM in ATE, indicating better global trajectory alignment with ground truth. RPE values are comparable, reflecting similar local consistency. LVIO-SAM produced smoother maps with clearer loop closures at road junctions and better maintained map scaling over long distances, as visible in the trajectory plots. For the M2P2 dataset, we see a much higher ATE for LVI-SAM and LVIO-SAM compared to the ATE on KITTI datasets. Here, also, LVIO-SAM outperforms LVI-SAM in ATE.

Qualitative inspection revealed that LVIO-SAM maintained trajectory coherence even during sharp turns and feature-sparse regions, highlighting the benefit of stereo depth fusion directly into the factor graph. The plausible reasons for these results are listed as follows:

- LVIO-SAM directly obtains depth measurements for visual features compared to LVI-SAM's reliance on lidar for depth association, which can have errors when lidar scans are misaligned and stereo depth improves scale estimation for textureless environments
- LVIO SAM uses a single factor graph for jointly optimizing constraints compared to LVI-SAM's dual subsystems
- LVIO has less cross dependence and can use IMU-Stereo and IMU-Lidar combinations but LVI cannot do the same
- LVIO SAM uses a dynamic sliding window to marginalize older scans whereas LVI uses a fixed window approach that increases possibility of drift.

VIII. Challenges

During the implementation of the LVIO-SAM baseline, several challenges were encountered that impacted the ease of setup, reproducibility, and evaluation of results. First and foremost, the GitHub repository associated with the project lacked sufficient and detailed documentation, which significantly increased the initial overhead in understanding the architecture, tuning parameters, and integrating the various components of the system. Essential setup steps, such as sensor calibration, configuration of launch files, and expected input formats, were either minimally explained or entirely missing. This made it difficult to confidently deploy the system and interpret its behavior, especially when adapting it to new datasets or different sensor configurations. In addition to this limitation, the absence of readily available ground truth data for direct comparison posed another major obstacle. To compensate for this, we had to generate ground truth trajectories manually using the OXTS inertial navigation sensor data provided in the KITTI dataset, which introduced additional preprocessing steps and synchronization requirements.

Moreover, we observed that the trajectories produced by LVIO-SAM were not synchronizing properly when deployed on the KITTI dataset, despite following the provided instructions and tuning common parameters. Specifically, the transform tool used for time alignment between LiDAR and camera frames was not providing accurate transformations,

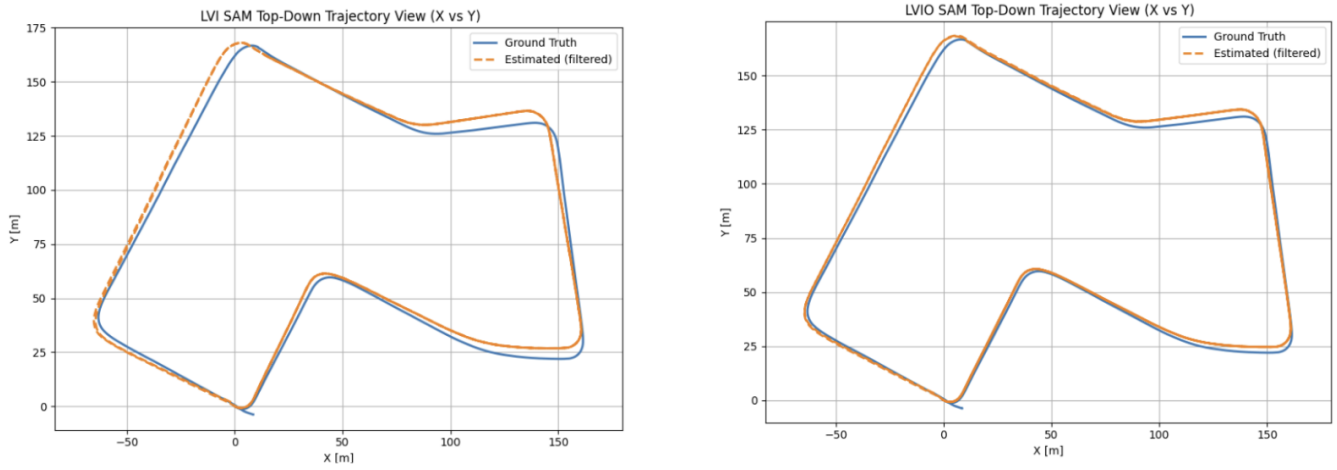

Fig. 6. LVIO-SAM results vs Ground Truth on the left and LVI-SAM results vs Ground Truth on the right for KITTI Dataset

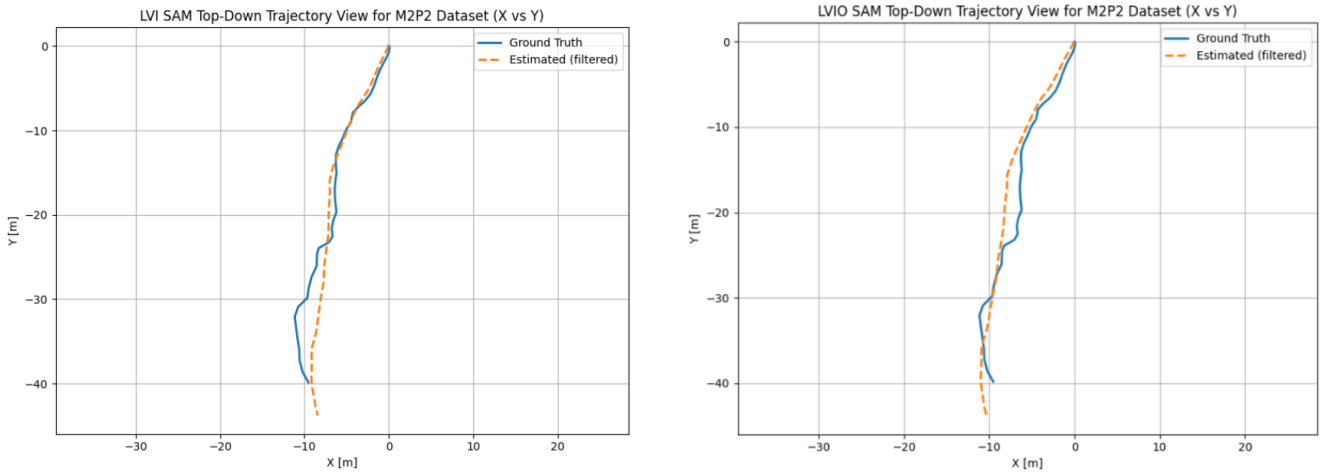

Fig. 7. LVIO-SAM results vs Ground Truth on the left and LVI-SAM results vs Ground Truth on the right for M2P2 dataset

resulting in inconsistent and sometimes divergent pose estimates. This suggested that either temporal misalignment or calibration drift might be present. Additionally, while the baseline demonstrated optimal performance on the KITTI dataset and other similar structured environments, it appeared to be sensitive to variations in scene structure and perceptual complexity, indicating a potential limitation in its adaptability and robustness when applied to less structured or more dynamic environments. This observation further emphasized the need for systems that generalize well across datasets with diverse sensory and environmental characteristics.

One other test that was conducted during the course of this work was to evaluate an alternative system known as Lvio-Fusion [17], which also proposes a tightly-coupled multi-sensor fusion strategy. However, several practical issues hindered the usability of this approach. The code dependencies in the repository were either outdated or incompletely specified, and the installation instructions lacked clarity. This made it challenging to build and run the system out of the box on a modern ROS setup. To address these issues and ensure reproducibility, we forked [18] the original repository and containerized the environment using Docker. This allowed us to isolate dependency-related problems and maintain consistent environments across machines. Additionally, we created a separate Docker container specifically for converting raw KITTI data into ROS bag files that were compatible with the expected input format of Lvio-Fusion, streamlining the preprocessing workflow and reducing manual overhead.

Although we were ultimately able to get the Lvio-Fusion system running in a controlled test environment, the output remained difficult to interpret. The system did not produce easily parsable performance metrics such as trajectory error or map quality, and there was limited diagnostic feedback to help identify failure cases. As a result, evaluating the system's effectiveness in a meaningful and quantitative way proved to be a significant challenge, and its reliability across diverse scenarios could not be fully assessed.

In addition to these software-level experiments, we also

explored the feasibility of deploying these SLAM systems on a real robotic platform. In particular, we considered using the Fetch mobile robot, which is equipped with a LiDAR, RGB-D camera, and onboard computing resources. However, a number of compatibility issues arose during this phase, especially related to rosbag playback support and real-time synchronization. These problems made it difficult to achieve seamless integration between the SLAM pipeline and the Fetch robot's data streams. As a result, real-world testing was hindered, highlighting the need for more portable and well-documented SLAM solutions that can transition smoothly from simulation or dataset-based testing to live robotic deployment scenarios.

IX. Conclusion and Future Work

We conclude that while the LVIO-SAM system demonstrates reliable and accurate performance on structured datasets such as KITTI, with well-established benchmark metrics, there remains considerable room for improvement in terms of adaptability to more complex and unstructured environments. The results obtained through our initial experiments validate that the baseline performs optimally when assumptions about scene structure, lighting, and sensor synchronization are satisfied. However, after conducting comparative analysis using the M2P2 dataset, which introduces greater diversity in terrain types and lighting conditions, it becomes evident that LVIO-SAM's robustness can be enhanced through further development. These insights justify the potential and feasibility of extending the current implementation to make it more generalizable across a wider spectrum of operational environments, especially those characterized by poor perceptual structure, irregular terrains, or degraded sensing conditions.

As a part of our future work, we aim to incorporate adaptive learning-based loop closure techniques that are capable of adjusting their behavior based on environmental variability. Unlike static feature-based loop closure mechanisms, learning-based approaches can dynamically model appearance changes, illumination variance, and structural differences, thus improving robustness in long-term or cross-domain deployments. Furthermore, we plan to investigate mechanisms that enable the SLAM system to switch sensing modalities or adjust sensor weights on-the-fly depending on high-level environmental conditions. For instance, under low-light conditions, the system could rely more heavily on LiDAR and inertial measurements, while in well-lit areas, visual odometry could be prioritized. This context-aware modality switching would allow the system to maintain consistent performance even when individual sensor streams degrade or become unreliable.

In parallel with these algorithmic advancements, a major stretch goal of our project is to implement and deploy the enhanced SLAM system on a real robotic platform, such as the Fetch mobile manipulator. This would allow us to validate the system's robustness, modularity, and real-time capability in real-world scenarios, including dynamic indoor and semi-structured outdoor environments. Deploying on a physical robot would also provide valuable insights into practical integration issues such as real-time synchronization, sensor calibration, and computational resource management. Through this, we aim to bridge the gap between dataset-driven SLAM research and real-world robotic autonomy, moving closer to fully operational, deployable solutions for harsh and unpredictable environments.

References

[1] The KITTI Vision Benchmark Suite. (n.d.). Cvlibs.net. Retrieved March 10, 2025, from https://www.cvlibs.net/datasets/kitti/

[2] M. Grupp, "evo: Python package for the evaluation of odometry and slam." 2017.

[3] X. Zhong, Y. Li, S. Zhu, W. Chen, X. Li and J. Gu, "LVIO-SAM: A Multi-sensor Fusion Odometry via Smoothing and Mapping," 2021 IEEE International Conference on Robotics and Biomimetics (ROBIO), Sanya, China, 2021, pp. 440-445, doi: 10.1109/ROBIO54168.2021.9739244.

[4] A. Datar, A. Pokhrel, M. Nazeri, M. B. Rao, C. Pan, Y. Zhang, A. Harrison, M. Wigness, P. R. Osteen, J. Ye, and X. Xiao, "M2P2: A Multi-Modal Passive Perception Dataset for Off-Road Mobility in Extreme Low-Light Conditions," arXiv preprint arXiv:2410.01105, 2024. Mourikis and Roumeliotis (2007):

[5] Mourikis, A. I., Roumeliotis, S. I. (2007). Multi-state constraint Kalman filter for vision-aided inertial navigation. Proceedings of the IEEE International Conference on Robotics and Automation, 3565–3572.

[6] Leutenegger, S., Lynen, S., Bosse, M., Siegwart, R., Furgale, P. (2015). Keyframe-based visual–inertial odometry using nonlinear optimization. The International Journal of Robotics Research, 34(3), 314–334.

[7] Qin, T., Li, P., Shen, S. (2018). VINS-mono: A robust and versatile monocular visual-inertial state estimator. IEEE Transactions on Robotics, 34(4), 1004–1020.

[8] Zhang, J., Singh, S. (2017). Low-drift and real-time lidar odometry and mapping. Autonomous Robots, 41, 401–416.

[9] Shan, T., Englot, B. (2018). LeGO-LOAM: Lightweight and ground-optimized lidar odometry and mapping on variable terrain. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 4758–4765.

[10] Graeter, J., Wilczynski, A., Lauer, M. (2018). LIMO: Lidar-monocular visual odometry. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 7872–7879.

[11] Shin, Y., Park, J., Lee, A., Kweon, I. S. (2020). Direct visual-lidar SLAM using photometric and geometric error terms. Sensors, 20(18), 5237.

[12] Shan, T., Englot, B., Ratti, C., Rus, D., Fallon, M. (2020). LIO-SAM: Tightly-coupled lidar inertial odometry via smoothing and mapping. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 5135–5142.

[13] Shan, T., Dai, B., He, D., Englot, B., Ratti, C., Rus, D., Fallon, M. (2021). LVI-SAM: Tightly-coupled lidar-visual-inertial odometry via smoothing and mapping. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 5692–5698.

[14] M2P2: Multi-Modal Passive Perception Dataset. (n.d.). Gmu.edu. Retrieved April 8, 2025, from https://cs.gmu.edu/ xiao/Research/M2P2/

[15] Dellaert F. Factor graphs and GTSAM: A hands-on introduction[R]. Georgia Institute of Technology, 2012.

[16] Z. Zhong, "LVIO-SAM," GitHub repository, https://github.com/TurtleZhong/LVIO-SAM, accessed Apr. 8, 2025.

[17] Jia, Y. (n.d.). Lvio-fusion: Lvio-fusion: A Self-adaptive Multi-sensor fusion SLAM Framework Using Actor-critic Method (IROS 2021).

[18] R. L. Catto, "Lvio-fusion," GitHub repository, Click to view, accessed Apr. 8, 2025.

[19] , GitHub repository for code, Click to view, accessed Apr. 8, 2025.