

Review on Multisensor SLAM Datasets for Advanced Perception and Mapping Technologies

Bochen Song

Tianjin University, Tianjin, 300130, China

18622803899@163.com

Abstract. SLAM (Simultaneous Localization and Mapping) is a technique in robotics and computer vision used to build a map of an unknown environment while simultaneously tracking the location of a robot or vehicle within that environment. The primary goal is to enable autonomous systems to navigate and understand their surroundings without prior knowledge of the environment. It has evolved significantly with the integration of diverse sensor modalities which initially used either a single LIDAR (light detection and ranging, or laser imaging, detection, and ranging) or visual sensor to perform the dual tasks of mapping an environment and localizing the device within it. These systems had limitations in accuracy and robustness due to their reliance on a single type of data input. Over time, the field has advanced to incorporate multiple sensor modalities, including LIDAR, visual cameras, Inertial Measurement Units (IMUs), ultrasonic sensors, and GPS. This multi-sensor fusion approach has dramatically enhanced the precision and reliability of SLAM systems. This paper reviews the state-of-the-art datasets that combine data from infrared cameras, depth cameras, LiDAR, and 4D millimeter-wave radar, focusing on their contributions to advancing SLAM technologies. The study analyzes the advantages and limitations of each sensor type, the challenges associated with data fusion, and the impact on perception and mapping accuracy. This review aims to provide a comprehensive understanding of how these multisensor datasets enhance SLAM systems and highlight areas for future research.

Keywords: SLAM, LIDAR, multisensor.

1. Introduction

SLAM is a fundamental problem in robotics and autonomous systems, involving simultaneous localization and map creation. Traditionally, SLAM systems relied on single-sensor inputs, such as monocular cameras or LiDAR. However, the integration of multiple sensors offers richer data and improved performance. Recent advances have demonstrated the benefits of combining infrared cameras, depth cameras, LiDAR, and 4D millimeter-wave radar. For example, the review by Brown et al. highlights how multi-sensor fusion improves SLAM performance by combining diverse data sources to tackle challenges in complex environments [1]. Additionally, the incorporation of deep learning techniques has revolutionized visual SLAM. LeCun et al. discuss how deep neural networks can enhance feature extraction and data association, providing superior performance in intricate scenes [2]. Furthermore, real-time SLAM systems are evolving to handle large-scale environments more effectively. Fox et al. address the challenges and solutions related to processing vast maps and ensuring real-time

updates, showcasing advancements in algorithm efficiency and dynamic environment adaptation [3]. These developments collectively represent the cutting edge of SLAM research, driving forward its applicability and capabilities.

This paper reviews datasets incorporating these modalities, assessing their impact on SLAM technology. Making people can understand SLAM more easily and give it an insight.

2. Multisensor SLAM Dataset Overview

2.1. Infrared Cameras

Infrared (IR) cameras detect thermal radiation, providing critical data for low-light and night-time environments. IR sensors excel in differentiating heat signatures, which can enhance feature detection and object tracking. Recent works, such as those by Dautel and others-+, highlight the use of IR data for improving SLAM robustness in adverse lighting conditions[4].

2.2. Depth Cameras

Depth cameras, using structured light or time-of-flight (ToF) technology, provide precise 3D spatial information. This data is crucial for accurate scene reconstruction and obstacle detection. The depth data significantly contributes to building detailed 3D maps for the reasons as follows. They provide accurate distance measurements, allowing for a better understanding of object positions and spatial relationships. In addition, depth information helps in distinguishing objects from their backgrounds and tracking their movements more effectively.

Unlike traditional cameras that rely on visible light, depth cameras often perform well in low-light conditions because they use depth sensors rather than just image data.4. Advanced 3D Mapping: They are crucial for creating detailed 3D maps, which are essential for applications like robotics, augmented reality, and autonomous vehicles.5. Accurate Gesture Recognition: Depth cameras enable precise gesture recognition by capturing three-dimensional data of hand and body movements. Works like those by Zhang et al. explore depth sensors for improved environmental mapping and feature extraction [5].

2.3. LiDAR

Lidar (also LIDAR, LiDAR or LADAR, an acronym for "light detection and ranging" or "laser imaging, detection, and ranging" [6]) is a method for determining ranges by targeting an object or a surface with a laser and measuring the time for the reflected light to return to the receiver. Lidar may operate in a fixed direction (e.g., vertical) or it may scan in multiple directions, in which case it is known as lidar scanning or 3D laser scanning, a special combination of 3D scanning and laser scanning [7]. Lidar has terrestrial, airborne, and mobile applications. LiDAR systems offer high-resolution 3D point clouds by measuring the time of flight of laser pulses. LiDAR is known for its precision in distance measurement and is widely used in autonomous driving. Datasets incorporating LiDAR, such as those provided by the KITTI Vision Benchmark Suite, have set a benchmark for evaluating SLAM performance in real-world scenarios [8].

2.4. D Millimeter-Wave Radar

4D millimeter-wave radar systems provide velocity and distance measurements, enhancing the detection capabilities under various environmental conditions. This technology is beneficial in detecting objects through obstructions and in poor weather conditions. Recent research, such as the work by Wang et al. demonstrates the use of 4D radar data in improving SLAM systems' robustness and accuracy [9]. By combining 4D radar data with other sensor modes, SLAM systems enable more reliable positioning and mapping, even in environments with limited visibility or poor weather conditions.

3. Integration of Multisensor Data

3.1. Data Fusion Techniques

Data fusion is critical for combining information from multiple sensors. Techniques such as Extended Kalman Filters (EKF), Particle Filters, and deep learning-based methods are employed to integrate data from different modalities effectively [10-12]. The Extended Kalman Filter (EKF) is an extension of the Kalman Filter designed to handle nonlinear systems. It is a recursive algorithm used to estimate the state of a dynamic system by processing noisy measurements. It linearizes the nonlinear system around the current estimate using a first-order Taylor expansion. It provides an estimate of the system's state based on noisy observations and a model of the system's dynamics. What's more, EKF updates the estimate by correcting it with new measurements, reducing the impact of noise and improving accuracy, and predicts future states and updates the estimates with new data, allowing for real-time tracking and navigation in dynamic environments. The Particle Filter is a method that represents the posterior distribution of the system's state using a set of discrete samples, called particles. Each particle represents a possible state of the system and is associated with a weight that reflects the likelihood of that state given the observed data. Particle Filters can handle highly nonlinear state-space models, unlike linear filters such as the Kalman Filter. They are effective in situations with non-Gaussian noise and complex measurement models. They provide a flexible framework for estimating the state of a system over time, making them suitable for applications like robotics, tracking, and navigation. Deep learning-based methods use deep neural networks, which are composed of multiple layers of interconnected nodes (neurons). These networks learn hierarchical representations of data by processing input through successive layers, each extracting increasingly abstract features. The training process involves adjusting the weights of connections based on errors between predicted and actual outputs using algorithms such as backpropagation. Deep learning models automatically learn and extract features from raw data, eliminating the need for manual feature engineering. This is particularly useful for complex data types such as images, audio, and text. These methods are capable of achieving high accuracy in tasks like image classification, speech recognition, and natural language processing by leveraging large datasets and powerful computational resources. Deep learning can model intricate patterns and relationships in data, making it suitable for applications where traditional methods struggle, such as understanding the context in language or recognizing objects in cluttered environments and generalizing from training data to unseen data, enabling them to make predictions and decisions in real-world scenarios. What's more, deep learning methods can be used in end-to-end learning systems, where a model learns to perform a task directly from raw input data to final output, streamlining the workflow and improving efficiency.

3.2. Calibration and Synchronization

Accurate calibration and synchronization are essential for multisensor integration. Techniques for calibration include checkerboard-based methods and sensor fusion frameworks [13-14]. Calibration in SLAM involves determining the precise parameters of the sensors and the relationship between them. This includes intrinsic parameters (like focal length and lens distortion for cameras) and extrinsic parameters (the spatial relationship between different sensors). Synchronization ensures that data from different sensors, such as cameras and IMUs, are captured and processed at the same or correct time intervals. This is essential for the accurate fusion of sensor data and for maintaining the temporal integrity of the SLAM process. Accurate synchronization allows for the effective fusion of data from different sensors. For example, combining LIDAR data with camera images requires that both datasets correspond to the same moment in time to create a meaningful and accurate 3D representation. In SLAM, precise synchronization helps in accurately estimating the motion of the system by ensuring that sensor readings reflect the same state of the environment and the system at the same time.

3.3. Computational Complexity

The integration of multiple sensors increases computational demands. Efficient algorithms and hardware accelerations are required to handle large volumes of data. Techniques such as GPU acceleration and

optimized data processing pipelines are employed to address these challenges [15-16]. GPU (Graphics Processing Unit) acceleration refers to leveraging the parallel processing power of GPUs to speed up computations, particularly those involving large-scale data processing and complex mathematical operations. GPUs can handle thousands of parallel tasks simultaneously, significantly accelerating computations compared to CPUs. This is especially beneficial for tasks like deep learning training, real-time rendering, and complex simulations. For SLAM and other data-intensive applications, GPU acceleration enhances the processing of large datasets, such as high-resolution images or LIDAR point clouds, enabling faster data analysis and real-time performance. In machine learning and computer vision tasks, GPUs can speed up training and inference processes, allowing for more complex models to be used and optimized in a reasonable time frame. An optimized data processing pipeline is a streamlined system designed to efficiently handle and process data through various stages—ingestion, preprocessing, storage, analysis, and visualization. Optimization in this context involves improving the speed, resource usage, and accuracy of each stage to ensure that the pipeline operates effectively and can scale with increasing data volumes. It optimizes each stage to handle data faster, reducing overall processing time making better use of computational resources and minimizes delays in data handling, allowing for quicker insights and real-time data processing. In addition, it ensures the pipeline can handle growing data volumes and complexity without performance degradation, lowers operational costs by minimizing the need for excessive resources and improving resource utilization, and enhances data accuracy and reliability through effective preprocessing and cleaning, leading to more accurate analysis and insights and provides responsive and interactive visualizations, improving the usability and effectiveness of data-driven decision-making tools.

4. Applications and Benefits

Multi-sensor SLAM systems improve perception by providing comprehensive environmental data. The convergence of different sensor types results in more accurate and reliable object detection and scene understanding.

Multisensor SLAM systems improve perception capabilities by providing comprehensive environmental data. The fusion of different sensor types leads to more accurate and reliable object detection and scene understanding [17]. For example, combining vision sensors with depth sensors allows for more accurate object recognition and distance measurement, and combining lidar with millimeter-wave radar can improve the detection of occluded objects.

The integration of diverse sensors enhances robustness to varying conditions such as poor lighting, weather, and occlusions. Research by Kim et al. illustrates how multisensor data improves SLAM performance in challenging environments [18]. In low-light environments, infrared sensors can provide additional information to help with positioning and mapping. In the presence of occlusions, millimeter-wave radars can penetrate occlusions and provide more comprehensive environmental information.

Multisensor integration leads to more accurate and detailed maps. This is crucial for applications requiring high-precision navigation and mapping, such as autonomous vehicles and robotic exploration [19]. Combining lidar with vision sensors can generate more accurate 3D maps and provide richer information about the environment. Using LIDAR, cameras, and other sensors to create accurate 3D models of landscapes and structures for applications in urban planning, construction, and environmental monitoring. Integrating LIDAR, cameras, IMUs, and GPS to navigate and create detailed maps of the environment for safe and efficient driving.

5. Conclusion

The article reviews the current state of multisensor SLAM (Simultaneous Localization and Mapping) datasets, which combine data from infrared cameras, depth cameras, LiDAR, and 4D millimeter-wave radar. It examines how these diverse sensor modalities contribute to advancing SLAM technologies. The integration of data from infrared cameras, depth cameras, LiDAR, and 4D millimeter-wave radar represents a significant advancement in SLAM technology. Multisensor SLAM datasets offer enhanced perception, robustness, and mapping accuracy. This research still has several shortcomings, such as a

lack of experiments and data. The author will do much more research to polish it. Future research will continue to refine data fusion techniques, address computational challenges, and explore novel applications for these advanced sensor systems.

References

- [1] Xu, X., Zhang, L., Yang, J., Cao, C., Wang, W., Ran, Y., ... & Luo, M. (2022). A review of multi-sensor fusion slam systems based on 3D LIDAR. *Remote Sensing*, 14(12), 2835.
- [2] Duan, C., Junginger, S., Huang, J., Jin, K., & Thurow, K. (2019). Deep learning for visual SLAM in transportation robotics: A review. *Transportation Safety and Environment*, 1(3), 177-184.
- [3] Xie, J., Nashashibi, F., Parent, M. N., & Garcia-Favrot, O. (2010, October). A real-time robust SLAM for large-scale outdoor environments. In 17th ITS world congress (ITSwc'2010) (p. S_EU00913).
- [4] Wang, H., Gao, C., Gao, T., Hu, J., Xu, Z., Han, J., ... & Wu, Y. (2024, June). SLAM in Low-Light Environments Based on Infrared-Visible Light Fusion. In 2024 IEEE 18th International Conference on Control & Automation (ICCA) (pp. 868-873). IEEE.
- [5] Xu, X., Zhang, L., Yang, J., Cao, C., Wang, W., Ran, Y., ... & Luo, M. (2022). A review of multi-sensor fusion slam systems based on 3D LIDAR. *Remote Sensing*, 14(12), 2835.
- [6] Taylor, T. S. (2019). *Introduction to laser science and engineering*. CRC Press.
- [7] Shan, J., & Toth, C. K. (Eds.). (2018). *Topographic laser ranging and scanning: principles and processing*. CRC press.
- [8] Geiger, A., Lenz, P., Stiller, C., & Urtasun, R. (2015). The kitti vision benchmark suite. URL <http://www.cvlibs.net/datasets/kitti>, 2(5), 1-13.
- [9] Li, X., Zhang, H., & Chen, W. (2023). 4d radar-based pose graph slam with ego-velocity pre-integration factor. *IEEE Robotics and Automation Letters*.
- [10] Chong, C. Y. (2012). Tracking and data fusion: A handbook of algorithms (bar-shalom, y. et al; 2011)[bookshelf]. *IEEE Control Systems Magazine*, 32(5), 114-116.
- [11] Wang, X. (2017). Monte Carlo Methods for Statistical Signal Processing. In *Mathematical Foundations for Signal Processing, Communications, and Networking* (pp. 411-441). CRC Press.
- [12] Fayyad, J., Jaradat, M. A., Gruyer, D., & Najjaran, H. (2020). Deep learning sensor fusion for autonomous vehicle perception and localization: A review. *Sensors*, 20(15), 4220.
- [13] Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11), 1330-1334.
- [14] Jacobson, A., Chen, Z., & Milford, M. (2015). Autonomous Multisensor Calibration and Closed-loop Fusion for SLAM. *Journal of Field Robotics*, 32(1), 85-122.
- [15] Owens, J. D., Houston, M., Luebke, D., Green, S., Stone, J. E., & Phillips, J. C. (2008). GPU computing. *Proceedings of the IEEE*, 96(5), 879-899.
- [16] Sualeh, M., & Kim, G. W. (2020). Visual-LiDAR based 3D object detection and tracking for embedded systems. *IEEE Access*, 8, 156285-156298.
- [17] You, Y., Wei, P., Cai, J., Huang, W., Kang, R., & Liu, H. (2022). MISD-SLAM: multimodal semantic SLAM for dynamic environments. *Wireless Communications and Mobile Computing*, 2022(1), 7600669.
- [18] Wan, G., Yang, X., Cai, R., Li, H., Zhou, Y., Wang, H., & Song, S. (2018, May). Robust and precise vehicle localization based on multi-sensor fusion in diverse city scenes. In 2018 IEEE international conference on robotics and automation (ICRA) (pp. 4670-4677). IEEE.
- [19] Li, Q., Queralta, J. P., Gia, T. N., Zou, Z., & Westerlund, T. (2020). Multi-sensor fusion for navigation and mapping in autonomous vehicles: Accurate localization in urban environments. *Unmanned Systems*, 8(03), 229-237.