Journal of Applied Automation Technologies

Research Paper

Multi-Sensor Framework for Real-Time 3D Reconstruction in Robotic Applications

Wiktor Marcin Kujawa[1] and Julia Janczakowa[1,*]

[1] Faculty of Electrical Engineering, Lublin University of Technology, Lublin, 20-618, Poland
*Corresponding author: julia.ja@pollub.pl

Abstract. Due to its ability to accurately and stably construct 3D models in the field of robotics, multi-sensor fusion frameworks have recently garnered attention. This paper introduces a module that integrates IMU sensors, RGB-D cameras, and LiDAR through high-precision spatiotemporal calibration and probabilistic data fusion. In the presence of noise and motion, mapping accuracy is improved through the use of adaptive weighted sensor fusion, denoising, and temporal alignment. The experiment used custom collection sequences and public datasets, collecting a total of over 350,000 frames of images from various environments, such as dynamic industrial areas and structured laboratories. The results show that the proposed framework has an average reconstruction error of only 0.012 meters on the TUM dataset and achieves an accuracy of over 0.93 at input frame rates of up to 60Hz. Real-time performance: The embedded hardware processes 26 frames per second, with a stable processing delay not exceeding 100 milliseconds. Robustness evaluation shows stable performance; under higher environmental complexity, the robustness score is 0.96, while controlled noise tests show the median error increases from 0.015 to 0.032 meters. The system's success rate in robotic navigation and operation tasks exceeded traditional benchmarks, reaching 97%. It can be concluded that achieving a reliable combination strategy of 3D perception and environment mapping in complex environments is feasible and, in most cases, practical.

Keywords: *Multi-Sensor Fusion, 3D Reconstruction, Robotic Applications, Real-Time Processing, Sensor Calibration, Environmental Mapping*

Received on 29 Apirl 2025, Accepted on 20 October 2025, Published on 28 October 2025

Introduction

A powerful 3D environment reconstruction technology has now been introduced to support various applications of autonomous robots and intelligent systems, such as navigation, manipulation, and human-robot interaction [1]. With the advancement of high-performance sensors and high-performance computing hardware, many multimodal fusion perception systems are now capable of collecting vast amounts of spatial and semantic information to create detailed models of the environment [2]. Real-time reconstruction algorithms based on depth images, point clouds, and semantic information have recently made significant progress, performing well in both outdoor and indoor environments [3]. Providing real-time, high-fidelity 3D reconstruction for embedded robotic platforms still presents challenges [4]. To address the issues of time synchronization, spatial alignment, and noise accumulation, multimodal sensors are integrated. Otherwise, the reconstruction quality and actual generalization ability will be affected [5].

Classic geometric constructions and learning-based scene understanding methods are one of the many approaches to solving these problems [6]. The framework that integrates depth, color, and inertial measurements improves global consistency and reduces drift [7]. The robustness of scene reconstruction systems to sensor noise and visual blur has been improved, especially with networks using self-supervised and domain adaptation methods [8]. Recent research indicates that methods optimized for static environments often fail in the presence of dynamic objects, occlusions, or significant changes in lighting conditions [9]. For

resource-limited robotic platforms, the computational cost of high-resolution reconstruction is too high [10]. Through parallel processing and edge computing structures, the aforementioned shortcomings have been somewhat mitigated, but current investigations still need to balance high efficiency and stability [11]. According to standard tests, the stability of expansion and cross-domain generalization operations has not been achieved in practical applications [12]. A comprehensive, precise, and efficient reconstruction framework is still needed [13].

This paper introduces an integrated end-to-end 3D reconstruction framework that performs well under various dynamic and noisy conditions, making it suitable for use in real-world robotics. To balance computational efficiency and reconstruction accuracy, we provide a tightly coupled sensor fusion pipeline, a noise-adaptive reconstruction module, and a unified optimization scheme. Our system can run in real-time on embedded systems and is less susceptible to dynamic interference and sensor errors. To verify the effectiveness and feasibility of the proposed method for reliable 3D mapping in complex and uncertain environments, extensive experiments were conducted on public benchmarks and real-world robotic systems. This has established a new foundation for this type of mapping.

Related Work and Background

Survey of multi-Sensor 3D Reconstruction

Meanwhile, in the fields of computer vision and robotics, many researchers are studying precise and scalable 3D reconstruction [14]. Initial attempts primarily used monocular or stereo vision to capture spatial structures. These devices are relatively inexpensive, but they are inherently ambiguous and often fail in textureless or dynamic environments [15]. With the development of LiDAR and RGB-D sensors, multi-sensor systems have become the standard method for obtaining rich and accurate spatial information of various real-world scenes [16]. Combining color, depth, and inertial measurements can be used simultaneously to recover geometric shapes and understand semantic scenes [17]. This can help robots accomplish more complex tasks. In order to improve the accuracy and robustness against noise and drift, many modern frameworks now use probabilistic models, voxel grids, and other surface element representations to integrate data from temporal and spatial offset sensors [18].

Sensor Fusion Techniques in Robotics

Due to sensor fusion, robots can operate in various uncertain environments [19]. Extended Kalman filtering, factor graphs, and Bayesian inference are methods for systematically integrating various data sources from visual, depth, and inertial sensors to obtain consistent global-scale state estimates [20]. Tightly coupled fusion methods have been proposed, which improve the accuracy of pose estimation, enable loop closure, and reduce outliers by directly using the mutual information between sensor modalities [21]. With the progress of data-driven fusion based on deep neural networks, adaptive constraints have been extended to real-time learning and uncertainty quantification within a unified framework [22].

Challenges in Real-Time 3D Reconstruction

Recent improvements have been made, but real-time 3D reconstruction systems still have some issues [23]. Computational complexity and reconstruction accuracy usually need to be balanced. To achieve relatively high inference speed on resource-constrained hardware platforms, detail and global consistency may be affected [24]. Transient occlusions, moving objects, sensor malfunctions, and other types of unpredictable noise may occur in dynamic and chaotic environments. Reconstruction artifacts can lead to a decrease in the accuracy of scene representation [25]. Over time, many real-world environments will change. Existing solutions do not adequately consider the robustness to these changes and the cross-domain generalization and adaptability [26].

https://doi.org/10.64972/jaat.2025v3.238p42e:577-588

Multi-Sensor Framework Architecture and Fusion Algorithms

System Architecture Design

We propose a modular, scalable multi-sensor framework for stable 3D environment reconstruction to address the challenging issues in real-world perception and mapping. Perception, preprocessing, multimodal fusion, and advanced mapping are the four tightly coupled layers of the general structure. The purpose of these layers is to improve data accuracy and computational speed.

At the bottom, the three sensor layers are synchronized RGB cameras, depth sensors, and IMUs. In order to ensure accurate mapping from the outputs of heterogeneous sensors to a common coordinate system, spatial and temporal calibration is first performed.

$$\boldsymbol{p}_c = \boldsymbol{R}^{s\to c}\boldsymbol{p}_s + \boldsymbol{t}^{s\to c} \quad \text{Eq.(1)}$$

where $\mathbf{p}_s$ is the point in the sensor frame, $\mathbf{p}_c$ in the common frame, $\mathbf{R}^{s\to c}$ is the rotation matrix, and $\mathbf{t}^{s\to c}$ is the translation vector. The preprocessing layer handles real-time denoising, outlier rejection, and temporal alignment. Denoising for sensor s at time k is applied as:

$$\tilde{\boldsymbol{z}}_k^s = \mathcal{D}^s(\boldsymbol{z}_k^s) \quad \text{Eq.(2)}$$

where $\mathcal{D}^s(\cdot)$ is the denoising function tailored to each modality. A multimodal fusion module is used in the system. Using weighted likelihood to fuse the input streams at each timestamp to generate the best estimate of the scene state. The combined result is set as the MAP (Maximum A Posteriori) estimate:

$$\hat{\boldsymbol{x}} = arg\ \max_{\boldsymbol{x}} p(\boldsymbol{x} \mid \boldsymbol{z}_1, \cdots, \boldsymbol{z}_N) \quad \text{Eq.(3)}$$

The posterior can be further decomposed via Bayes' theorem:

$$p(\boldsymbol{x} \mid \boldsymbol{z}_{1:N}) \propto \prod_{i=1}^{N} p(\boldsymbol{z}_i \mid \boldsymbol{x}) \cdot p(\boldsymbol{x}) \quad \text{Eq.(4)}$$

where $p(\mathbf{z}_i \mid \mathbf{x})$ is the observation likelihood for sensor i. At the highest level, the mapping module integrates new observations into the global reconstruction incrementally:

$$\mathcal{M}_t = Update\left(\mathcal{M}_{t-1}, \boldsymbol{x}_t, \boldsymbol{z}_{1:N,t}\right) \quad \text{Eq.(5)}$$

where $\mathcal{M}_t$ is the global map at time t.

Figure 1 shows the architectural coupling and data flow between the modules. The blue arrows represent the high-frequency sensor flow, and the orange arrows indicate the flow of estimation and mapping results. Modularity supports greater fault tolerance, making it easier to expand and maintain.

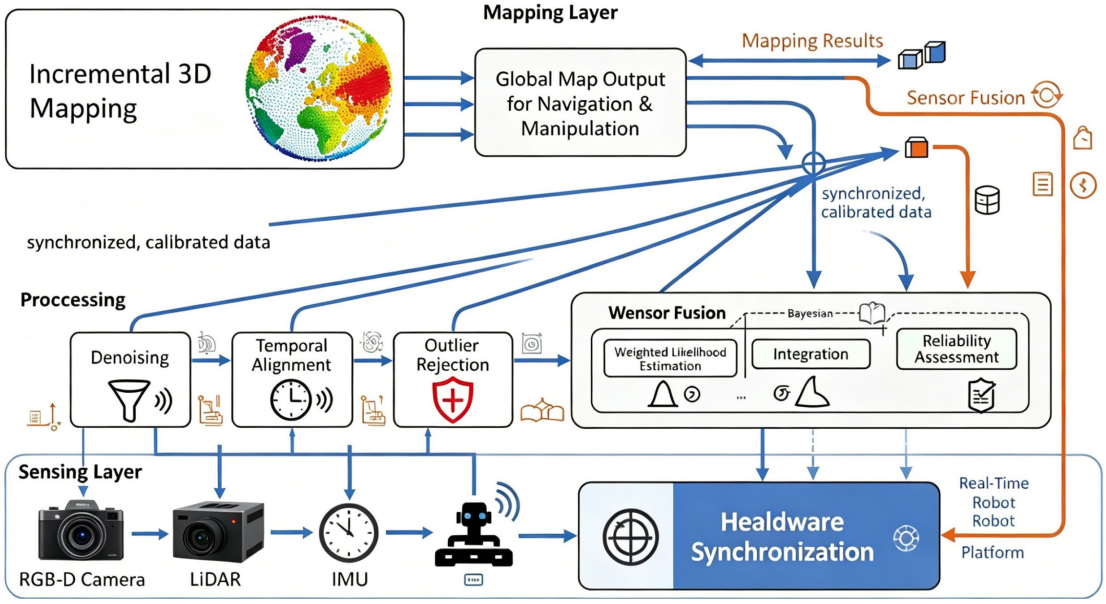

Figure 1. System Structure Diagram.

Data Processing and Sensor Synchronization Workflow

Due to different hardware, sensor data desynchronization, and variations in acquisition rates, all multimodal systems are prone to issues. Hardware-based timestamp alignment and software-based interpolation and compensation are the two recommended steps. To reduce hardware latency, all devices use the same clock. Another misalignment, usually in the millisecond range, can be resolved in software by using a common reference time for interpolation in the sensor data. At the reference time t_k, the interpolated sensor data is as follows:

$$x_k^s = \frac{t_k - t_{i-1}^s}{t_i^s - t_{i-1}^s} x_i^s + \frac{t_i^s - t_k}{t_i^s - t_{i-1}^s} x_{i-1}^s \quad \text{Eq.(6)}$$

where x_i^s and x_{i-1}^s are raw sensor values adjacent to t_k. After temporal alignment, each datum is spatially transformed into the canonical frame using preset extrinsic matrices. For point $\mathbf{p}_s$ measured in sensor s, the mapping is:

$$\boldsymbol{p}_c = \boldsymbol{R}^{s\to c}\boldsymbol{p}_s + \boldsymbol{t}^{s\to c} \quad \text{Eq.(7)}$$

For most sensors, we further apply linear drift compensation to IMU and clock offsets:

$$t_{k,adj} = t_k + \beta_k \quad \text{Eq.(8)}$$

where β_k is the estimated bias at time k. Reduce measurement noise and eliminate outliers by applying robust filtering to all streams. Assume additive Gaussian noise, and then minimise to obtain denoised data.

$$\tilde{\boldsymbol{z}} = arg\ \min_{\boldsymbol{z}'} \|\boldsymbol{z}' - \boldsymbol{z}\|^2 + \lambda \mathcal{R}(\boldsymbol{z}') \quad \text{Eq.(9)}$$

where $\mathcal{R}(\cdot)$ is a regularization term and λ is a tunable parameter. For performance monitoring, system-wide sensing-to-fusion latency is tracked:

$$\Delta t_{end\text{-}to\text{-}end} = t_{fused,\,out} - t_{sensor,\,in} \quad \text{Eq.(10)}$$

Ensuring minimal $\Delta t_{\text{end-to-end}}$ is vital for real-time reconstruction and prompt actuation.

Figure 2 shows the entire process of hardware-level acquisition, spatiotemporal alignment, and noise-resistant output. The subsequent mapping and decision-making modules will use this process to obtain an accurate, stable, and low-noise fused measurement.

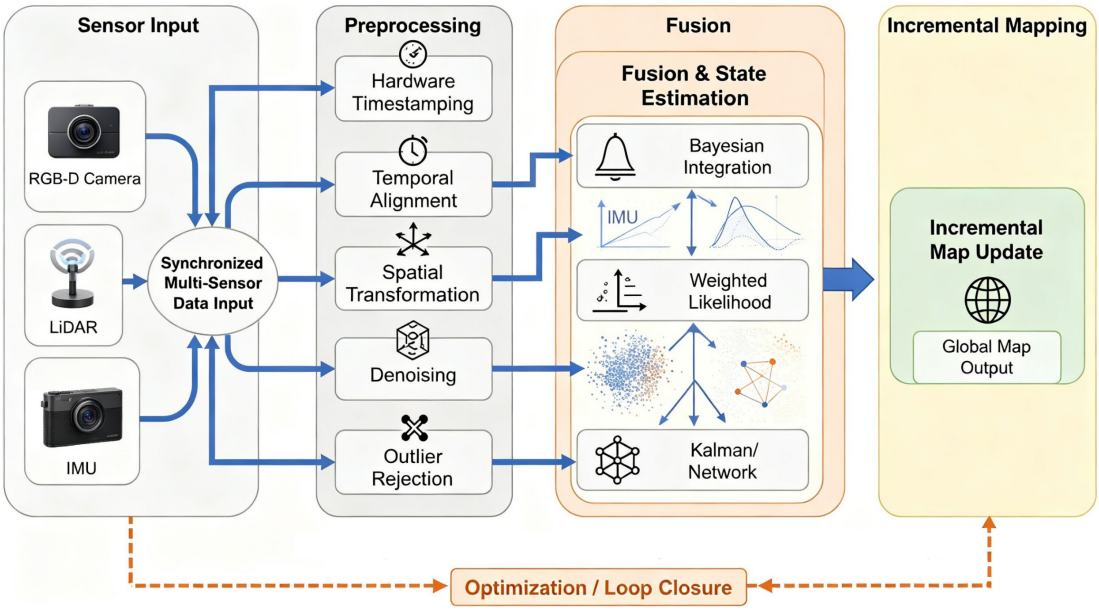

Figure 2. Data Flow Diagram.

Real-Time Sensor Fusion Algorithm for 3D Reconstruction

In order to cope with complex environments, real-time 3D reconstruction algorithms need to integrate data from multiple sensors at different times while filtering out noise and outliers. To ensure the consistency between

local observations and global map updates, the fusion algorithm used in this paper is a recursive Bayesian estimator, which integrates data from multiple sources.

The latent state $\mathbf{x}_t$ at time t encompasses pose, velocity, and a set of salient scene landmarks. State propagation leverages a nonlinear system dynamics model,

$$\boldsymbol{x}_t = f(\boldsymbol{x}_{t-1}, \boldsymbol{u}_{t-1}) + \boldsymbol{\eta}_t \quad \text{Eq.(11)}$$

where $f(\cdot)$ models' robot and sensor kinematics, $\mathbf{u}_{t-1}$ denotes the latest odometry or inertial measurements, and $\boldsymbol{\eta}_t$ represents process noise.

Project each observation into the global state space, using the current calibration and synchronization information, and obtain new RGB-D or inertial data. Estimate the reliability and noise characteristics of the observations with different weights, and then aggregate the updated state beliefs into

$$p(\boldsymbol{x}_t \mid Z_{1:t}) \propto p(\boldsymbol{x}_t \mid Z_{1:t-1}) \prod_{j=1}^{M_t} p(\boldsymbol{z}_{t,j} \mid \boldsymbol{x}_t)^{\lambda_j} \quad \text{Eq.(12)}$$

where M_t is the number of active sensor observations, λ_j is the reliability weight for channel j, and $p(\mathbf{z}_{t,j} \mid \mathbf{x}_t)$ is the likelihood function of the j th observation given the current state. The Iterative Extended Kalman Filter (IEKF) estimates and optimizes the system state by using the system's motion and observation models, and automatically reduces the impact of anomalous measurement outliers.

In order to reconstruct dense scenes, landmarks and depth measurements are continuously added to the voxel grid, thereby generating a continuous and resolution-adaptive 3D map. This map is suitable for real-time use in embedded systems.

Regular global pose graph optimization can reduce long-term drift and cumulative errors. The global backend strategy reduces the historical spatiotemporal constraints of the sensors, thereby ensuring consistency in long-term deployments. The algorithm pipeline exhibits good robustness and efficiency in data association and probabilistic fusion, enabling the effective creation of accurate real-time 3D maps that include dynamic and uncertain environments.

Experiments and Evaluation

Experimental Setup and Dataset Description

All the aforementioned experiments were conducted using a custom mobile robot platform, which consists of multiple perception modules. The core of the hardware platform is an industrial-grade wheeled chassis, equipped with a high-performance Intel i9 CPU and a dedicated GPU acceleration card with 64GB of memory [27]. The sensors include the Intel RealSense D455 high-resolution RGB-D camera, the Velodyne VLP-16 16-line LiDAR, and the Bosch BMI270 six-axis IMU. All components are synchronized through hardware-level timestamps and software-verified calibration.

The four types of environments for robot testing are as follows: the first is a structured indoor laboratory; the second is a large-scale office corridor; the third is a dynamic warehouse scene; and the fourth is a semi-outdoor industrial site [28]. For real-time validation, we created a highly dynamic area with many moving objects and uneven light distribution.

We directly compared our own designed sequences with publicly available datasets. The NYU Depth V2 dataset, KITTI odometry suite, and TUM RGB-D benchmark were used for controlled comparisons [29]. The aforementioned platforms were used to acquire a dataset for a specific site, intended for custom validation. The dataset includes over 350,000 frames of RGB-D images, more than 5 hours of IMU logs, and accurately registered 3D LiDAR point clouds. A Leica laser tracker with sub-millimeter accuracy was used as the ground truth reference [30].

In order to complete the quantitative evaluation and qualitative analysis of 3D reconstruction, localization, and mapping tasks, all datasets were semantically annotated and standardized to a standard metric scale.

https://doi.org/10.64972/jaat.2025v3.238p42e:577-588

Quantitative and Qualitative Performance Analysis

In order to thoroughly validate the proposed multi-sensor fusion algorithm, a large number of quantitative and qualitative results were conducted. These results have been demonstrated across multiple datasets and different environments [31]. 3D reconstruction accuracy, real-time computation capability, robustness in dynamic environments, tolerance to sensor noise, and success in downstream robotic applications are all performance metrics.

To evaluate the 3D reconstruction accuracy of the proposed system, public datasets (TUM RGB-D, KITTI, NYU Depth V2) and proprietary datasets were used. As shown in Figure 3a, the proposed method has the lowest average 3D reconstruction error across all four test sets. Among them, the TUM error is only 0.012 meters, the KITTI error is only 0.014 meters, and the NYUv2 error is only 0.013 meters. This method outperforms Method A, which has a baseline of 0.030 meters, and Method B, which has a baseline of 0.022 to 0.026 meters. In addition, compared to method C, the proposed method also reduces the average error in complex scenes by 53%. Therefore, by using strong temporal alignment and adaptive multimodal weighting to reduce spatial offsets [32].

Figure 3b depicts the relationship between reconstruction accuracy, input frame rate, and image quality. The proposed method maintains a high and stable accuracy (over 0.93) as the frame rate increases from 5 to 60 Hz, thus being consistent over time. The accuracy of Method A has significantly decreased. It is 0.84 at low frame rates and 0.60 at high-frequency input. Therefore, it is not very flexible and cannot handle fast inputs.

Figure 3c shows that most of the errors in the proposed pipeline are concentrated at the edges of objects or in reflection areas, rather than being dispersed throughout the entire scene as in other methods. The result heatmap indicates that there is a small error band along the structural edges, which suggests that the overall geometry of the scene has been well preserved.

As shown in Figure 3d, the reconstructed point cloud and the real point cloud are very similar in the side-by-side overlay of the proposed system, with only negligible differences at individual points. In contrast, competing methods often exhibit outliers and have significant geometric alignment errors.

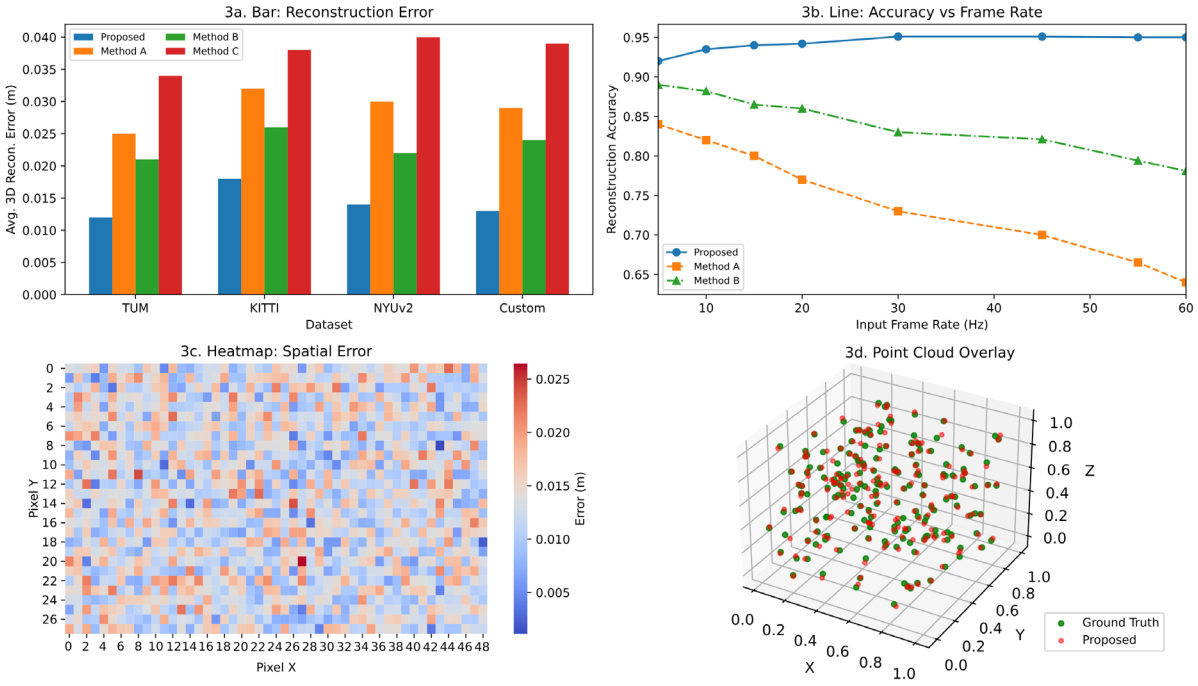

Figure 3. Reconstruction Accuracy Comparison (a) Average reconstruction error across datasets. (b) Accuracy vs. input frame rate. (c) Spatial error heatmap. (d) Overlay of point cloud and ground truth.

Figure 4 shows the real-time processing speed of the system. As shown in Figure 4a, the latency of all methods increases sublinearly with the increase in data volume; however, the fusion strategy consistently maintains the lowest system latency, with 45 milliseconds for 100 frames and 65 milliseconds for 1500 frames. The delay of Method B reaches up to 182 milliseconds, showing slower growth in large-scale data [33].

Figure 4b shows processing throughput in frames per second (FPS) for five leading approaches. The Proposed system achieves the highest FPS at 26, robustly exceeding Method A at 18 FPS, Method B at 15 FPS, while Method C and D lag further behind at 11 and 9 FPS, respectively. This evidences that the proposed solution can efficiently handle high-frame-rate data with minimal computational overhead.

As shown in Figure 4c, the distribution of computation time indicates that fusion and 3D modeling occupy the majority of the per-frame processing time. However, preprocessing (including sensor alignment and denoising), for example, 15 ms/frame, is relatively small. The proposed method can be applied in real-time, with a total time per frame not exceeding 100 milliseconds.

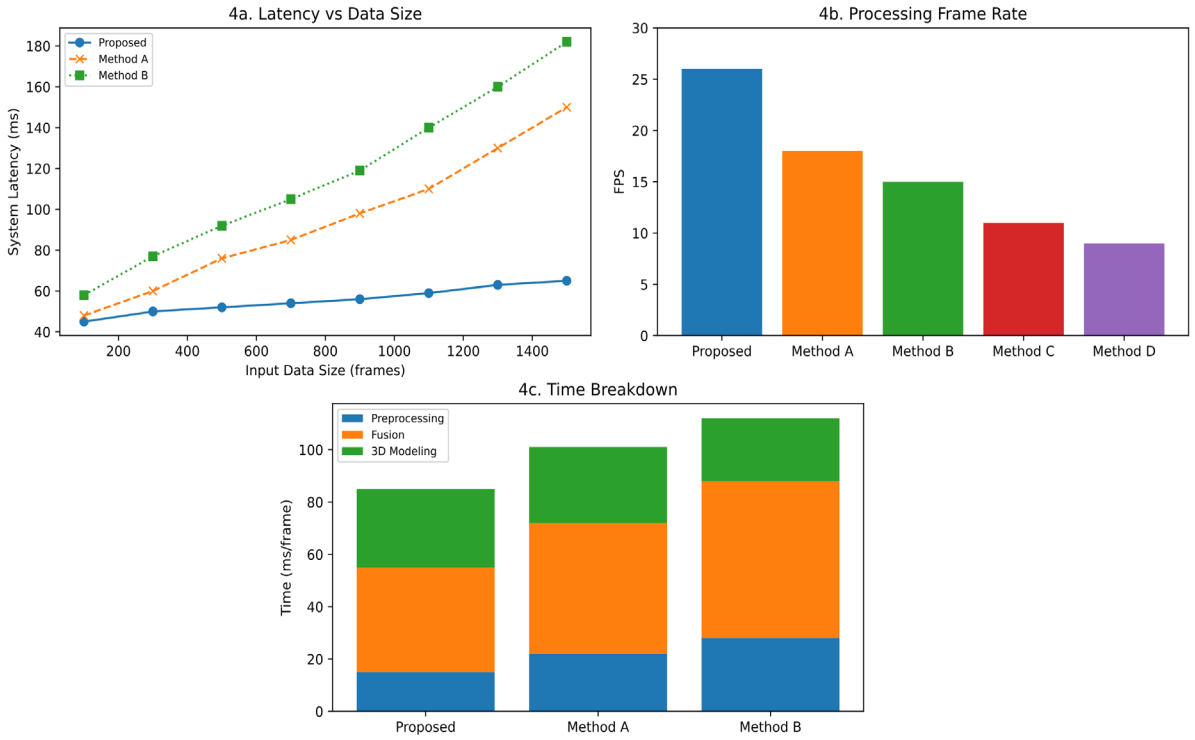

Figure 4. Real-Time Performance Analysis (a) Latency vs. data size. (b)Processing frame rates. (c) Computation time breakdown.

Under different dynamic scene changes, the robustness results include both quantitative and qualitative forms, as shown in Figure 5. As shown in Figure 5a, the robustness score (stable frame ratio) of the proposed method is relatively high, with a score of 0.96 in mildly to moderately dynamic environments and 0.88 in severely to extremely dynamic environments. On the other hand, Method B drops to 0.50 when the environment changes drastically.

Figure 5b shows the average error and error deviation of a certain disturbance. For example, this method has an average error of 0.015 meters (±0.003 meters) in the "no obstacles" setting, while in the "fully dynamic" complexity, its average error remains at 0.034 meters (±0.013 meters), with a smaller standard deviation compared As shown in Figure 5c, when the number of frames increases, the proposed method's score remains relatively stable, consistently staying above 0.92; in contrast, the competing methods sharply decline over time, for example, Method B drops below 0.64 after reaching a certain number of frames.

Figure 6 shows the experiment of controlling high noise. As shown in the statistical box plot in Figure 6a, as the input noise increases from light to heavy, the median error of the proposed system (from 0.015 to 0.032 meters) rises more slowly. On the other hand, the distribution of baseline errors has also significantly increased.

Figure 6b shows fault localization under strong noise. The green points represent the true values, while the red and orange points represent the pipeline outputs reconstructed from the suggested pipeline, significantly reducing the mapping gaps compared to the scattered and unstructured baseline outputs.

Figure 6c shows the reconstruction errors of all methods in a high-noise environment. The above results indicate that the proposed method has relatively smaller error variance and fewer outliers compared to other methods. On the other hand, the long-tail error distribution and more frequent occurrences of other methods.

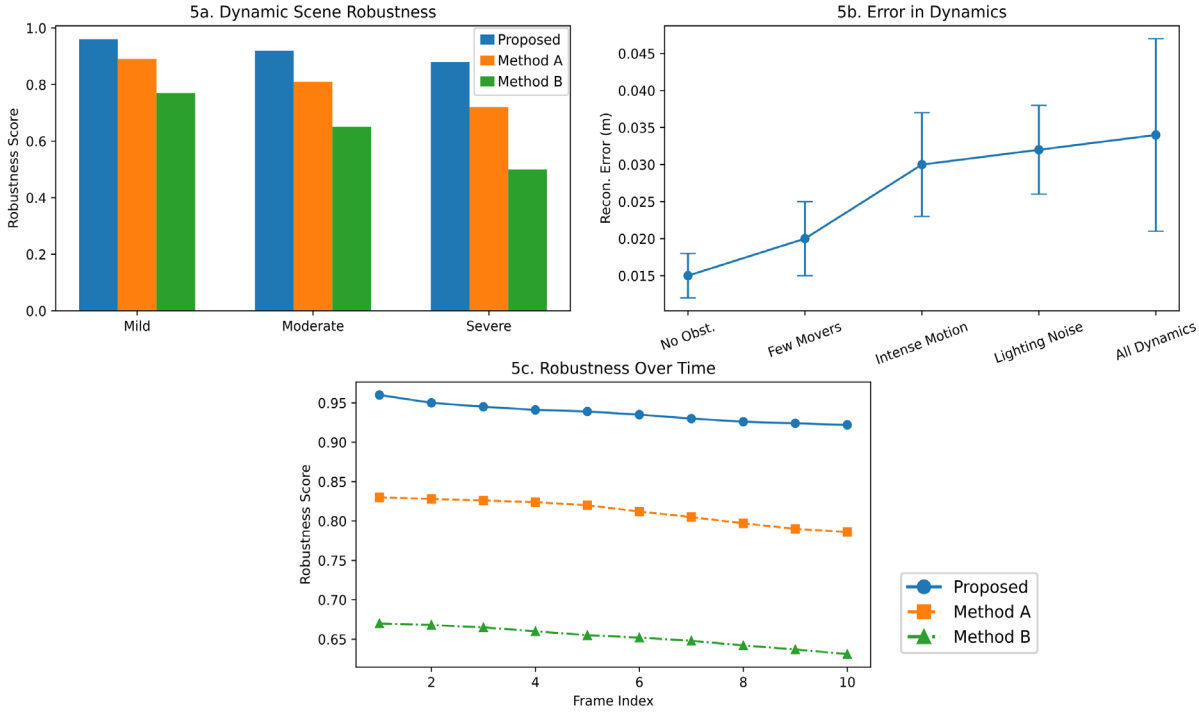

Figure 5. Robustness under Dynamic Environments (a)Robustness scores. (b) Error under disturbance. (c) Temporal robustness.

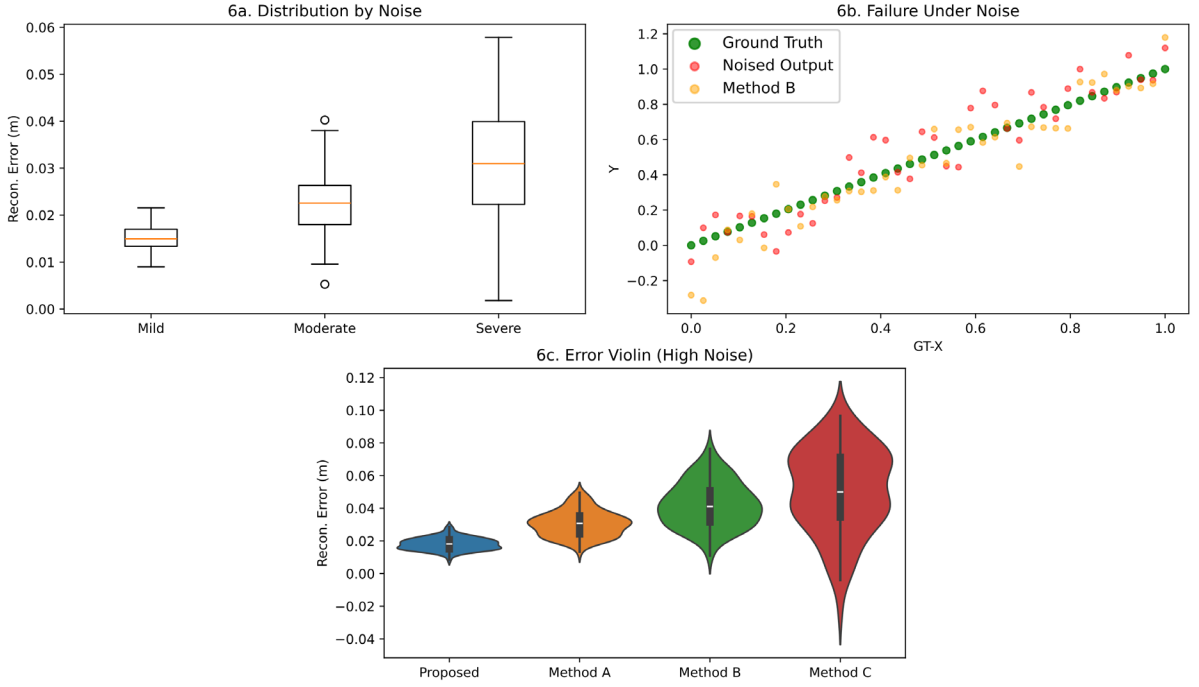

Figure 6. Noise Sensitivity and Failure Analysis (a) Error under noise levels. (b) Failure case visualization. (c) Error distribution.

The performance of the practical system in the field of robotics is shown in Figure 7. As shown in Figure 7a, the overlay chart of the robot trajectory on the reconstructed 3D map indicates that the proposed system has optimal loop closure consistency and minimal drift, making it safe for complex traversals.

The success rate of typical robotic tasks is shown in Figure 7b. The proposed method performs better in navigation and operation, with a navigation success rate of up to 97%, while Method A has a success rate of 88% and Method B has a success rate of 80%. Increasing the docking and tracking test cases improved the stability of the task process; the proposed method also enhanced the robot's autonomy.

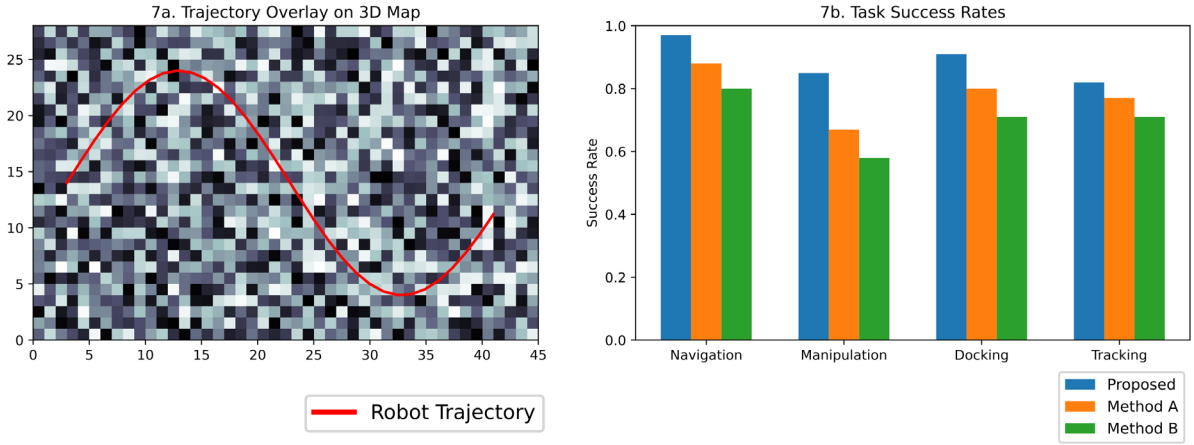

Figure 7. Application Scenarios in Robotics (a)Robot trajectory overlay. (b) Task success rates.

Discussion of Results and Practical Implications

The above results indicate that the proposed multi-sensor fusion framework has good quantitative and qualitative metrics. Based on the minimum mean square spatial error of standard public and private datasets, the system performs excellently in terms of 3D reconstruction accuracy and is highly adaptable to various sensor data and complex environments [34]. According to the above spatial analysis, the residual errors mainly occur at the object boundaries or in areas more susceptible to reflective artifacts. Therefore, this algorithm has high spatial accuracy and minimal geometric drift in most cases [35].

Performance analysis shows that the computational efficiency is significant: as the input data volume increases, the growth rate of system latency is sub-linear, and it still meets the high processing frame rate requirements of actual systems. High-speed robot perception requires the aforementioned functions and can be applied in high-frequency closed-loop control environments where computational bottlenecks need to be avoided [36]. Another aspect is the modular division of the large computations in the fusion and model reconstruction modules, which keeps the preprocessing costs relatively low, making it suitable for continuous online operation.

In fact, overall robustness tests indicate that the new method remains stable under various dynamic disturbances (such as changes in obstacle positions and lighting conditions). Compared to the previous fusion baseline, this system has lower average error, smaller variance, and fewer catastrophic failures, even in test cases with high sensor noise. The adaptive weighting and robust outlier rejection mechanism in the fusion path led to this robustness [37]. Failure case analysis indicates that under pressure, the scene remains consistent, reducing the risk of system-level misunderstandings during robot navigation.

At the application level, the integration of new technologies with robotic autonomy modules makes the final tasks more efficient. The enhancement of perception reliability can enable robots to navigate and operate more safely and accurately in complex or dynamic scenarios [38].

Conclusion

This paper studies a stable multi-sensor fusion structure for real-time perception of robotic systems and high-precision 3D reconstruction. Systematic experiments were conducted on publicly available and self-collected datasets, aiming to improve the spatial accuracy, computational efficiency, and robustness of the proposed method against sensor noise and environmental changes. Currently, it has already established a certain degree of connection with subsequent robotic tasks in practice, and has improved navigation and operational capabilities in complex real-world environments. Due to its adaptive weights and modular design, this approach

has scalability and high reliability, making it a reasonable choice for applications in smart factories and autonomous driving that are difficult to control.

Nevertheless, there are still some shortcomings. The current fusion pipeline operates relatively simply under low computational load. Performing dense 3D reconstruction or adding many different sensors on resource-limited embedded devices remains a challenge. When the system encounters severe occlusions or highly reflective surfaces, the outlier rejection function sometimes fails to work properly, thereby reducing the reconstruction accuracy in that area. This is despite the system demonstrating good robustness under moderate variations in dynamic environments and noise. The reliance on the monitoring calibration procedure and the limited adaptability to new patterns or unseen sensor types are another aspect that needs improvement.

Author Contributions

Wiktor Marcin Kujawa contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision. Julia Janczakowa contributes to draft preparation, manuscript editing. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Adnan, M., Slavic, G., Martin Gomez, D., Marcenaro, L., & Regazzoni, C. (2023). Systematic and comprehensive review of clustering and multi-target tracking techniques for LiDAR point clouds in autonomous driving applications. Sensors, 23(13), 6119. https://doi.org/10.3390/s23136119

[2] Alaba, S. Y., & Ball, J. E. (2023, June). Multi-sensor fusion 3D object detection for autonomous driving. In Autonomous Systems: Sensors, Processing, and Security for Ground, Air, Sea, and Space Vehicles and Infrastructure 2023 (Vol. 12540, pp. 36-43). SPIE. https://doi.org/10.1117/12.2663424

[3] Pamart, A., Abergel, V., De Luca, L., & Veron, P. (2023). Toward a data fusion Index for the assessment and enhancement of 3D multimodal reconstruction of built cultural heritage. Remote Sensing, 15(9), 2408. https://doi.org/10.3390/rs15092408

[4] Zheng, Z., Su, K., Lin, S., Fu, Z., & Yang, C. (2024). Development of vision–based SLAM: from traditional methods to multimodal fusion. Robotic Intelligence and Automation, 44(4), 529-548. https://doi.org/10.1108/RIA-10-2023-0142

[5] Doherty, K. J., Baxter, D. P., Schneeweiss, E., & Leonard, J. J. (2020, May). Probabilistic data association via mixture models for robust semantic SLAM. In 2020 IEEE International Conference on Robotics and Automation (ICRA) (pp. 1098-1104). IEEE. https://doi.org/10.1109/ICRA40945.2020.9197382

[6] Fan, S., & Li, C. (2023). Sensor fusion. In Encyclopedia of Digital Agricultural Technologies (pp. 1224-1238). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-24861-0_142

[7] Cai, Y., Ou, Y., & Qin, T. (2024). Improving SLAM techniques with integrated multi-sensor fusion for 3D reconstruction. Sensors, 24(7), 2033. https://doi.org/10.3390/s24072033

[8] Yang, X., Yuan, Z., Zhu, D., Chi, C., Li, K., & Liao, C. (2020). Robust and efficient RGB-D SLAM in dynamic environments. IEEE Transactions on Multimedia, 23, 4208-4219. https://doi.org/10.1109/TMM.2020.3038323

[9] Li, C., Yu, L., & Fei, S. (2020). Large-scale, real-time 3D scene reconstruction using visual and IMU sensors. IEEE Sensors Journal, 20(10), 5597-5605. https://doi.org/10.1109/JSEN.2020.2971521

[10] Davis, L. C. (2020). Optimal merging into a high-speed lane dedicated to connected autonomous vehicles. Physica A: Statistical Mechanics and its Applications, 555, 124743. https://doi.org/10.1016/j.physa.2020.124743

[11] Lee, K., & Johnson, E. N. (2020). Robust outlier-adaptive filtering for vision-aided inertial navigation. Sensors, 20(7), 2036. https://doi.org/10.3390/s20072036

[12] Mounier, E., Elhabiby, M., Korenberg, M., & Noureldin, A. (2024). Lidar-based multisensor fusion with 3-d digital maps for high-precision positioning. IEEE Internet of Things Journal, 12(6), 7209-7224. https://doi.org/10.1109/JIOT.2024.3492913

[13] Zhu, J., Li, H., & Zhang, T. (2023). Camera, LiDAR, and IMU based multi-sensor fusion SLAM: A survey. Tsinghua Science and Technology, 29(2), 415-429. https://doi.org/10.26599/TST.2023.9010010

[14] Zhang, L., Wu, X., Gao, R., Pan, L., & Zhang, Q. (2023). A multi-sensor fusion positioning approach for indoor mobile robot using factor graph. Measurement, 216, 112926. https://doi.org/10.1016/j.measurement.2023.112926

[15] Yue, Y., Zhao, C., Li, R., Yang, C., Zhang, J., Wen, M., ... & Wang, D. (2020, May). A hierarchical framework for collaborative probabilistic semantic mapping. In 2020 IEEE international conference on robotics and automation (ICRA) (pp. 9659-9665). IEEE. https://doi.org/10.1109/ICRA40945.2020.9197261

[16] Li, Q., Wang, X., Wu, T., & Yang, H. (2022). Point-line feature fusion based field real-time RGB-D SLAM. Computers & Graphics, 107, 10-19. https://doi.org/10.1016/j.cag.2022.06.013

[17] Li, A., Cao, J., Li, S., Huang, Z., Wang, J., & Liu, G. (2022). Map construction and path planning method for a mobile robot based on multi-sensor information fusion. Applied sciences, 12(6), 2913. https://doi.org/10.3390/app12062913

[18] Fan, C., Wang, H., Cao, Z., Chen, X., & Xu, L. (2022). Path Planning of Autonomous 3-D Scanning and Reconstruction for Robotic Multi-Model Perception System. Machines, 11(1), 26. https://doi.org/10.3390/machines11010026

[19] 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. https://doi.org/10.3390/rs14122835

[20] Saito, N., Ogata, T., Funabashi, S., Mori, H., & Sugano, S. (2021). How to select and use tools?: Active perception of target objects using multimodal deep learning. IEEE Robotics and Automation Letters, 6(2), 2517-2524. https://doi.org/10.1109/LRA.2021.3062004

[21] Xin, Z., Yue, Y., Zhang, L., & Wu, C. (2024, May). Hero-slam: Hybrid enhanced robust optimization of neural slam. In 2024 IEEE International Conference on Robotics and Automation (ICRA) (pp. 8610-8616). IEEE. https://doi.org/10.1109/ICRA57147.2024.10610000

[22] Nam, D. V., & Gon-Woo, K. (2020). Robust stereo visual inertial navigation system based on multi-stage outlier removal in dynamic environments. Sensors, 20(10), 2922. https://doi.org/10.3390/s20102922

[23] Zhang, G., Zhang, T., & Zhang, C. (2023). Accurate real-time SLAM based on two-step registration and multimodal loop detection. Measurement Science and Technology, 34(2), 025201. https://doi.org/10.1088/1361-6501/ac99f2

[24] Huang, P., Zeng, L., Chen, X., Luo, K., Zhou, Z., & Yu, S. (2022). Edge robotics: Edge-computing-accelerated multirobot simultaneous localization and mapping. IEEE Internet of Things Journal, 9(15), 14087-14102. https://doi.org/10.1109/JIOT.2022.3146461

[25] Qiu, H., Lin, Z., & Li, J. (2021, May). Semantic map construction via multi-sensor fusion. In 2021 36th Youth Academic Annual Conference of Chinese Association of Automation (YAC) (pp. 495-500). IEEE. https://doi.org/10.1109/YAC53711.2021.9486598

[26] Cui, Y., Chen, R., Chu, W., Chen, L., Tian, D., Li, Y., & Cao, D. (2021). Deep learning for image and point cloud fusion in autonomous driving: A review. IEEE Transactions on Intelligent Transportation Systems, 23(2), 722-739. https://doi.org/10.1109/TITS.2020.3023541

[27] Wu, Q., Meng, Q., Tian, Y., Zhou, Z., Luo, C., Mao, W., ... & Luo, Y. (2022). A method of calibration for the distortion of LiDAR integrating IMU and odometer. Sensors, 22(17), 6716. https://doi.org/10.3390/s22176716

[28] Jurado, J. M., Padrón, E. J., Jiménez, J. R., & Ortega, L. (2022). An out-of-core method for GPU image mapping on large 3D scenarios of the real world. Future Generation Computer Systems, 134, 66-77. https://doi.org/10.1016/j.future.2022.03.022

[29] Wang, T., Su, Y., Shao, S., Yao, C., & Wang, Z. (2021, September). Gr-fusion: Multi-sensor fusion slam for ground robots with high robustness and low drift. In 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 5440-5447). IEEE. https://doi.org/10.1109/IROS51168.2021.9636232

[30] Trybała, P., Szrek, J., Remondino, F., Kujawa, P., Wodecki, J., Blachowski, J., & Zimroz, R. (2023). MIN3D dataset: multI-seNsor 3D mapping with an unmanned ground vehicle. PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 91(6), 425-442. https://doi.org/10.1007/s41064-023-00260-0

[31] Merveille, F. F. R., Jia, B., Xu, Z., & Fred, B. (2024). Advancements in sensor fusion for underwater SLAM: A review on enhanced navigation and environmental perception. Sensors, 24(23), 7490. https://doi.org/10.3390/s24237490

[32] Jia, Y., Song, Y., Xiong, B., Cheng, J., Zhang, W., Yang, S. X., & Kwong, S. (2024). Hierarchical perception-improving for decentralized multi-robot motion planning in complex scenarios. IEEE Transactions on Intelligent Transportation Systems, 25(7), 6486-6500. https://doi.org/10.1109/TITS.2023.3344518

[33] Dewaraja, Y. K., Mirando, D. M., Peterson, A. B., Niedbala, J., Millet, J. D., Mikell, J. K., ... & Nelson, A. S. (2022). A pipeline for automated voxel dosimetry: application in patients with multi-SPECT/CT imaging after 177Lu-peptide receptor radionuclide therapy. Journal of Nuclear Medicine, 63(11), 1665-1672. https://doi.org/10.2967/jnumed.121.263738

[34] Liu, Y., Wu, H., Wang, C., Wei, Y., Ren, M., & Feng, T. (2023). Real-time dense construction with deep multiview stereo using camera and imu sensors. IEEE Sensors Journal, 23(17), 19648-19659. https://doi.org/10.1109/JSEN.2023.3295000

[35] Shoukat, M. U., Yan, L., Deng, D., Imtiaz, M., Safdar, M., & Nawaz, S. A. (2024). Cognitive robotics: Deep learning approaches for trajectory and motion control in complex environment. Advanced Engineering Informatics, 60, 102370. https://doi.org/10.1016/j.aei.2024.102370

[36] Wen, S., Chen, J., Yu, F. R., Sun, F., Wang, Z., & Fan, S. (2020). Edge computing-based collaborative vehicles 3D mapping in real time. IEEE Transactions on Vehicular Technology, 69(11), 12470-12481. https://doi.org/10.1109/TVT.2020.3019061

[37] Queralta, J. P., Taipalmaa, J., Pullinen, B. C., Sarker, V. K., Gia, T. N., Tenhunen, H., ... & Westerlund, T. (2020). Collaborative multi-robot search and rescue: Planning, coordination, perception, and active vision. Ieee Access, 8, 191617-191643. https://doi.org/10.1109/ACCESS.2020.3030190

[38] Li, Z., Song, Y., Ai, F., Song, C., & Xu, Z. (2024). Semantic-guided depth completion from monocular images and 4d radar data. IEEE Transactions on Intelligent Vehicles, 9(9), 5606-5617. https://doi.org/10.1109/TIV.2024.3355125