Data Engineering and Applications

Research Paper

# A Hybrid Data Association Method for Multi-Sensor SLAM

**Cyprian Malinowski**[1, *]

[1] Maria Curie-Sklodowska University, Faculty of Mathematics, Physics and Computer Science, 20-031 Lublin, Poland
*Corresponding author: cyprian.m@wpias.edu.pl

**Abstract.** Autonomous vehicles require simultaneous localization and mapping (SLAM). In order to create a more complete system, various sensors have been recently added, such as LiDAR, RGB-D cameras, inertial measurement units, and radar. A common issue when using multiple sensors is that the data obtained is inconsistent due to factors such as noise, offsets, or drift. This paper proposes a comprehensive hybrid data association framework. The framework integrates probabilistic modeling, graph-based optimization, and deep feature learning into a closed-loop system. Attention-based fusion modules and entropy-driven confidence gating are used to reliably match multi-sensor observations in a common latent space. Experimental validation through complex public and self-collected benchmarks has shown significant progress with the hybrid method. In terms of data association, a recall rate of 94.1% and an accuracy of 93.7% were achieved, with the average translational error reduced to 0.21 meters, outperforming both deep learning and traditional SLAM benchmarks. Real-time performance, memory usage: 1.2-1.5 GB, computation time per cycle: 48 ms. Ablation studies indicate that probabilistic reasoning, semantic encoding, and global graph optimization are the three components of the system, and they must work together to achieve robustness and mapping performance. The new model is not well-suited for areas with blur or severe occlusion, but it performs well in dynamic, crowded, and low-texture areas. This study sets new high standards for multimodal SLAM data association and provides practical support for the deployment of long-term, highly reliable robotic systems.

**Keywords:** *Machine Perception, Sensor Fusion, SLAM, Data Association, Probabilistic Modeling, Deep Learning*

Received on 09 August 2024, Accepted on 28 December 2024, Published on 16 January 2025



## Introduction

SLAM (Simultaneous Localization and Mapping) has been applied to many parts of modern mobile robots and autonomous vehicles, used to create spatial maps of the world while simultaneously determining the robot's location [1]. This field initially began as monocular or stereo vision methods, but has now evolved into a multimodal system, including sensors such as LiDAR, RGB-D cameras, inertial measurement units, and radar [2]. With the increasing demand for various environments, many of which are unstructured, research has begun to use multi-sensor fusion technologies [3]. If SLAM is not extended to multiple modalities, there will be many issues. In this list, sensor heterogeneity, drift, and time/space alignment are the most important issues. For this reason, more research has been conducted on robust calibration, time synchronization, and multi-source fusion frameworks [4]. Despite significant progress, many practical SLAM systems still face issues such as data association accuracy, map consistency, and real-time computational load, especially in the presence of noise or incomplete sensor data [5].

Robust SLAM uses data association to link observations from multiple sensors to the map or match features found in previous runs [6]. Due to the different resolutions, fields of view, and signal characteristics of each sensor, this issue becomes more pronounced when using multiple sensors [7]. When measuring noise, traditional association algorithms, such as nearest neighbor and joint compatibility branch and bound, assume that the noise distribution is Gaussian. In order to handle outliers and perceptual ambiguity, the algorithms need

to be rigorously adjusted [8]. Graph-based methods and probabilistic frameworks exhibit good stability in certain situations; however, they still fall short in cases of sparse data, different modalities, or environmental aliasing [9]. Learning-based descriptors and neural data association modules have made significant progress in the research field, but they have not yet been systematically integrated with traditional optimization methods. These academic advancements have not yet been applied in practice [10].

In light of the aforementioned issues, this paper proposes a comprehensive multi-sensor SLAM hybrid data association framework. The architecture combines graph-based optimization, learned feature representations, and probabilistic modeling within a unified pipeline. By integrating the advantages of handcrafted association heuristics and adaptive data-driven matching, while keeping the computation simple and cross-modal flexible, the aim is to achieve this goal. Extensive experiments were conducted on complex open-source and proprietary benchmarks, and the system performed exceptionally well under conditions of rich perceptual diversity and sensor noise. This study aims to advance principled SLAM design and provide a scalable foundation for the next generation of autonomous vehicles operating in complex environments.

## Review of Data Association Strategies

### Probabilistic and Graph-Based Models

Data based on graphs and probabilities show that association models are still the method for multi-sensor SLAM. Bayesian data fusion is a probabilistic method that can address sensor data uncertainty, thereby obtaining relative weights for each set of data [11]. Joint Compatibility Branch and Bound (JCBB) and other probabilistic graphical methods can reduce false positives and improve the robustness of outliers by explicitly considering compatibility sets. These methods can ensure a strong connection between measurements and landmarks [12]. In contrast, graph-based SLAM describes the environment and sensor poses as nodes connected by observed edges, thereby achieving a high-level global optimization program for trajectory and map consistency [13]. Factor graphs can be extended to large-scale and high-dimensional problems, and effectively handle multi-sensor constraints [14]. The aforementioned methods have been widely applied to ground vehicles, drones, and mobile platforms, providing a relatively reliable foundation for large-scale mapping and localization [15].

### Limits and Motivation for Hybridization

Although they each have their advantages, graph models and probabilistic models cannot individually solve the modern multi-sensor SLAM problem. The original probabilistic association schemes typically use linear approximations and assume a Gaussian noise model; they may not be suitable for environments with strong nonlinear sensor responses or non-Gaussian errors [16]. The aforementioned methods may lead to ambiguous associations and are more prone to accumulating errors in cluttered or dynamic environments [17]. If there is severe perceptual aliasing or sparse environmental features, this situation may occur. Although graph-based optimization is relatively effective, reliable association assumptions are needed when constructing the graph; otherwise, incorrect associations or erroneous loop closures can lead to map distortion and localization failure [18]. Feature descriptors or geometric heuristic methods, as well as feature drift issues, lead to a decline in cross-domain robustness and adaptability [19]. As multi-sensor fusion becomes increasingly common in SLAM, there is a need for a model that can integrate structured probabilistic reasoning, global graph optimization, and context-adaptive data interpretation.

### Overview of Recent Hybrid Trends

In order to combine the respective advantages of probabilistic, graphical, and data-driven methods, a hybrid data association framework has recently emerged. Deep learning techniques have been introduced to complement the previous association logic. By using semantic and appearance cues to eliminate the differences in similar features found in different perceptual domains [20]. For example, by combining the output of deep neural networks with factor graph optimization, good results have been achieved on complex urban data [21]. This method improves the robustness of loop closure and the stability of long-term localization. Other studies have investigated how reinforcement learning or adaptive Bayesian filters can adjust probability thresholds in the context of sensor accuracy fluctuations and other environmental changes [22]. The new hybrid paradigm introduces learning-based uncertainty modeling relationships, which allows confidence in specific sensor

streams or cues to be adjusted according to environmental changes [23]. The aforementioned trend indicates that the future direction will be a comprehensive, all-encompassing data connectivity network capable of functioning in open-world scenarios and major tasks [24]. There is still a lack of research on effectively combining these heterogeneous methods without compromising computational feasibility and interpretability. More innovation is still needed in the area of hybrid multi-sensor data association [25].

## Hybrid Data Association Model Design

**Fusion Principles**

Integrating probabilistic reasoning, graph-based reasoning, and deep feature learning into a tightly integrated computational system is our first plan for the joint data association model. First, each input data stream from the inertial measurement unit, LiDAR, and camera is encoded into a purpose-optimized neural representation module. Through this encoding, dense and discriminative embeddings containing high-level semantic information and low-level geometric features can be obtained. After these specific modal features are imported, they are mapped to a common latent space. This association operation considers semantic similarity and spatial proximity in joint affinity.

One of them is a mechanism for achieving fusion through focused attention. In each iteration, the system uses the original location cues and depth feature similarity to calculate the context-sensitive association probability. Normalized affinity represents the correlation probability between two observations:

$$P_{assoc}\left(z^{(i)}, z^{(j)}\right) = \frac{\exp\left(-\alpha\left\|p^{(i)} - p^{(j)}\right\|^2 + \beta S\left(f^{(i)}, f^{(j)}\right)\right)}{\sum_k \exp\left(-\alpha\|p^{(i)} - p^{(k)}\|^2 + \beta S(f^{(i)}, f^{(k)})\right)} \quad \text{Eq.(1)}$$

Here, $p^{(i)}$ is the geometric position associated with observation $i$, $f^{(i)}$ is its learned embedding, $S(\,,\,)$ is a feature similarity function, and $\alpha, \beta$ are sensitivity parameters dictated by sensor calibration.

The framework uses a confidence gating function, considering only the association probabilities with normalized Shannon entropy in their local matching neighborhood to prevent the spread of uncertain or false associations.

$$G^{(i)} = 1 - \frac{H\left(\mathbf{P}^{(i)}\right)}{\log K} \quad \text{Eq.(2)}$$

where $H\left(\mathbf{P}^{(i)}\right)$ is the entropy of the set of association probabilities for candidate matches of observation $z^{(i)}$, and $K$ is the number of candidates in its window. In order to enter the next stage of optimization, this door requires highly trusted partners.

By reducing the penalty for mismatched loss functions based on association confidence, comprehensively calibrate the quality of associations:

$$\mathcal{L}_{\text{assoc}} = \sum_t \sum_{i,j} G^{(i)} \cdot \ell_{\text{match}}\left(z_t^{(i)}, z_t^{(j)}\right) \quad \text{Eq.(3)}$$

In this expression, $\ell_{\text{match}}$ is a context-adaptive loss that dynamically penalizes misassociations, allowing the system to focus on informative sensor correspondences and maintain resilience in cluttered or ambiguous environments.

Figure 1 shows the structure of the aforementioned fusion network and also displays the data paths from multiple sensor inputs to neural encoding, entropy gating, and attention-weighted association. The subsequent interface for optimizing the diagram. In order to improve the reliability and effectiveness of multimodal associations, this design will use a reasonable framework that considers uncertainty and reinforcement learning to evaluate the observations of different sensors.

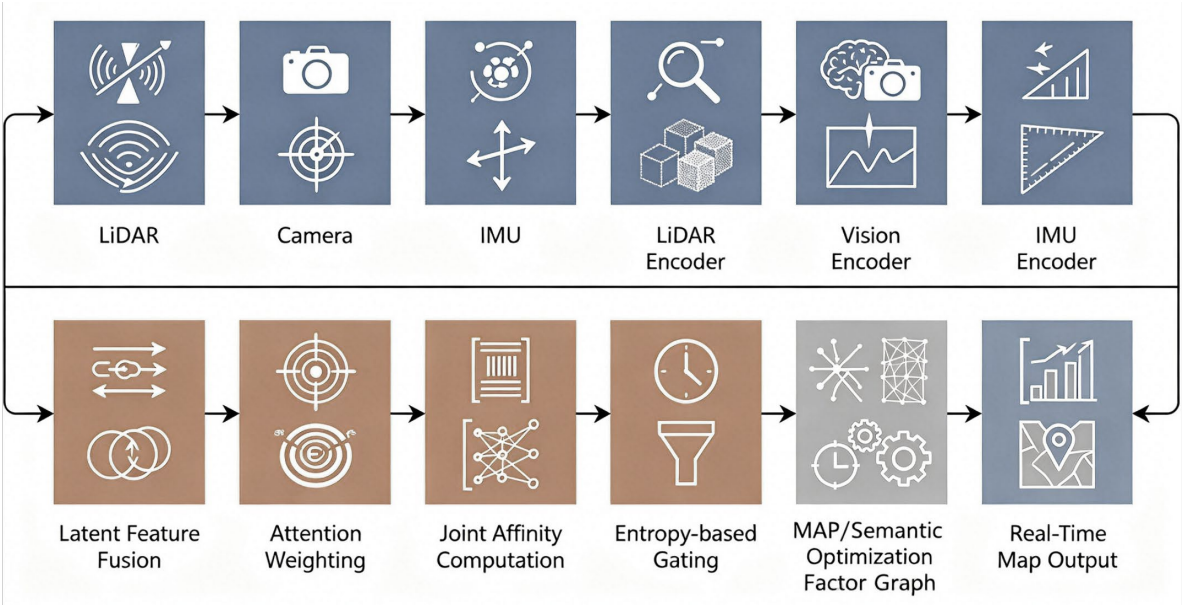


**Figure 1.** Overall System Architecture of the Proposed Hybrid Data Association Approach for Multi-Sensor SLAM.

**Graph Optimization Backbone**

The system spatiotemporal factor graph is composed of association layers, which show the relationships from sensors to the map and from observation to observation. Based on the successful high-confidence data association in the previous fusion layer, each node in this dynamic graph is either a robot poses or a map feature, and each node has edges.

The goal of the optimization is to determine the configuration of the node states to maximize the joint posterior of the poses and landmarks for each relevant observation:

$$\arg\max_{\mathbf{X},\mathbf{L}} P(\mathbf{X},\mathbf{L} \mid \mathbf{Z},\mathbf{U}) \quad \text{Eq.(4)}$$

where $\mathbf{X}$ encapsulates all robot states, $\mathbf{L}$ the latent landmark locations, $\mathbf{Z}$ the multi-modal observation set, and $\mathbf{U}$ the control/input stream. This posterior's factorisation breaks it down into a product of odometric and local observational likelihoods, each of which is represented by a factor in the graph:

$$P(\mathbf{X},\mathbf{L} \mid \mathbf{Z},\mathbf{U}) \propto \prod_k \psi_k(\mathbf{x},\mathbf{l},\mathbf{z},\mathbf{u}) \quad \text{Eq.(5)}$$

Each factor $\psi_k$ is parameterized either by analytic sensor models (for calibrated modalities) or by learned surrogates (for abstract feature correlations), ensuring flexibility across new sensor types.

This system is a Maximum A Posteriori (MAP) estimator, solved through iterative nonlinear least squares. In each iteration, the objective function updates the pose and feature states by reducing the geometric and semantic association residuals:

$$\mathcal{C}_{MAP} = \sum_{e\in\mathcal{E}} \rho(\|\mathbf{r}_e(\mathbf{X},\mathbf{L})\|^2) + \gamma \sum_{e'\in\mathcal{E}_S} \|\mathbf{s}_{e'}(\mathbf{F})\|^2 \quad \text{Eq.(6)}$$

Here, $\mathbf{r}_e$ denotes the residual for a conventional geometric factor, $\mathbf{s}_{e'}$ the semantic mismatch, $\rho$ a robust kernel function, and $\gamma$ a trade-off coefficient. Loop closure: Adaptively prune candidate links based on the loop consistency in the factor graph topology. Temporal smoothing priors are also used to penalize false links:

$$\mathcal{S}_{\text{loop}} = \sum_l \lambda \left\| \Delta_{\text{loop}}(t_l) \right\|^2 \quad \text{Eq.(7)}$$

where $\Delta_{\text{loop}}(t_l)$ is the deviation in estimated loop timing, enforcing global map consistency without excessive computational overhead.

**Implementation Details**

In terms of feature extraction, uncertainty quantification, and scalable optimization, the construction of the proposed real-time multi-sensor SLAM hybrid data association framework needs to be efficient. First, optimize neural encoders specifically designed to handle various data types (such as images, LiDAR, etc.) to process raw input streams from these sources and use dedicated hardware acceleration. After normalization and temporal alignment, the encoded features are stored in a central association buffer. In the buffer, the spatial attention module selects matches with high relevance and previously low association uncertainty.

The nodes and associations in the system are hierarchically organized to meet the requirements of global consistency and local responsiveness. Local subgraphs can be quickly stored and periodically integrated with the global pose graph. This design still retains the functionality of loop closure and correcting accumulated drift, and it supports fast incremental updates. Pipeline task scheduling uses adaptive load balancing for dynamic load distribution. This makes all delay requirements uniform.

A stable outlier rejection rule is used to evaluate the reliability of the correlation of the optimized residuals. The confidence in each association $C^{(i,j)}$ is estimated as:

$$C^{(i,j)} = G^{(i)} \cdot P_{\text{assoc}}\left(z^{(i)}, z^{(j)}\right) \quad \text{Eq.(8)}$$

where $G^{(i)}$ is the confidence gating function, and $P_{\text{assoc}}\left(z^{(i)}, z^{(j)}\right)$ is the association probability derived previously. To ensure the stability of the mapping, associations with confidence below a specific threshold will be ignored.

The real-time system delay can be dynamically adjusted and managed:

$$\tau_{\text{latency}} = \max\left\{\tau_{\text{encode}}, \tau_{\text{assoc}}, \tau_{\text{opt}}\right\} \quad \text{Eq.(9)}$$

where $\tau_{\text{encode}}$, $\tau_{\text{assoc}}$ and $\tau_{\text{opt}}$ represent the respective durations for front-end feature encoding, association computation, and back-end graph optimization. The aforementioned balance criteria will dynamically reallocate system resources.

Figure 2 shows the process of collecting data from sensors, confidence-gated association, real-time graph optimization, and map generation; it illustrates how feature extraction, association reasoning, uncertainty handling, and hierarchical optimization work together to ensure stable mapping under unstable sensing conditions.

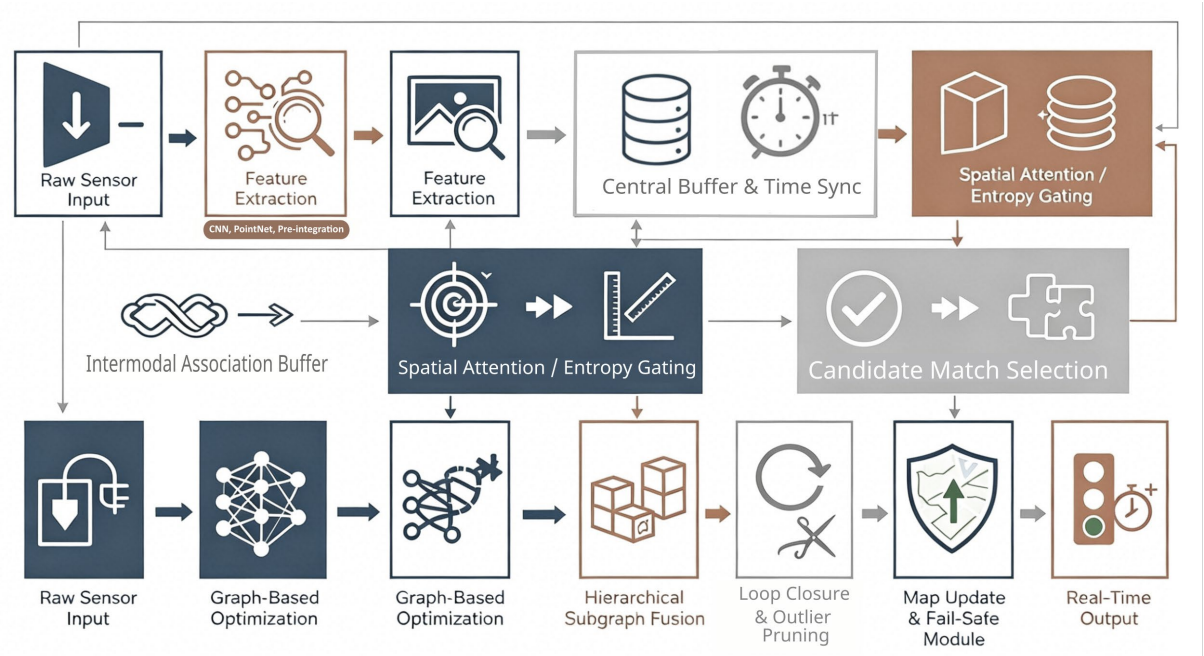


**Figure 2.** Complete Workflow of the Hybrid Association System.

## Experimental Evaluation & System Architecture

**Sensor and System Layouts**

The design goal of the sensor array is to enhance complementarity and temporal consistency by adding calibrated multimodal arrays. On a meticulously crafted carbon fiber frame, a high-resolution forward global shutter camera (2048 x 1536 pixels, 30 Hz) and a 32-line rotating LiDAR (10 Hz, 360° horizontal, 0.1° angle) are installed and securely fixed. To reduce the lever arm effect and improve cross-modal stability, a three-axis MEMS IMU (400 Hz) is located at the center of the optical/LiDAR reference. Mechanical vibration dampers and stable baseline calibration will be used.

Each sensor is hardware-triggered, with sub-millisecond precision timestamps, and the hardware abstraction layer ensures cross-modal alignment within 2 milliseconds. Rolling shutter distortions are modeled and compensated through calibration matrix correction. Lie algebraic update mechanisms are used to optimise camera-LiDAR extrinsics through iterative spatial alignment. Temporal IMU bias is tracked and corrected using rolling window pre-integration, tightly coupling accelerometer and gyroscope data with the visual-inertial framework.

Central processing is carried out by a GPU with an edge-grade 12GB of VRAM and a six-core ARM CPU, to which all data streams are connected via a high-speed data bus. A real-time middleware system aligns and fuses multi-modal feature packets for input to the neural association and graph-optimization pipelines, coordinated through a latency-aware scheduler.

The maximum observed desynchronisation over a predetermined correspondence window is used to describe inter-sensor temporal drift:

$$\Delta\tau_{\text{sync}} = \max|t_{\text{cam}}(i) - t_{\text{lidar}}(j)|, \forall i, j \in W \quad \text{Eq.(10)}$$

where $t_{\text{cam}}$ and $t_{\text{lidar}}$ are hardware-synchronized timestamps and $W$ indicates the sliding window.

For geometric extrinsic stability, minimise an alignment residual term:

$$\Lambda_{\text{ext}} = \arg\min_{R,t} \sum_{k} \left\| R x_{\text{cam}}^{k} + t - x_{\text{lidar}}^{k} \right\|^2 \quad \text{Eq.(11)}$$

where $R$ and $t$ denote the rotation and translation between frames, and $x_{\text{cam}}^{k}, x_{\text{lidar}}^{k}$ are spatially matched points.

**Dataset & Metrics**

The benchmark tests will simultaneously use widely used public datasets and multi-sensor sequences. RELLIS-3D is the rural terrain benchmark, KITTI Odometry is the urban complexity benchmark, and Oxford Radar RobotCar is the adverse weather benchmark. For completeness, please add some locally sourced datasets with intense dynamics and low texture for indoor environments.

Evaluate the model's effectiveness, stability, and accuracy. Use the Root Mean Square Error (RMSE) to calculate the difference between the system's pose output in translation and rotation and the true reference. A high resilience index indicates that the robustness of false associations has decreased:

$$\Xi_{\text{robust}} = \frac{1}{N} \sum_{i=1}^{N} \exp\left(-\kappa \varepsilon_i\right) \quad \text{Eq.(12)}$$

Here, $\varepsilon_i$ is the proportion of rejected outliers in window $i, \kappa$ is a sharpness coefficient, and $N$ the number of temporal divisions.

Under real-time constraints, a dynamic delay limit is used to measure system throughput, which simultaneously considers peak and average utilization:

$$\Psi_{\text{runtime}} = \sup_{t} \left( \frac{N_{\text{proc}}(t)}{N_{\text{in}}(t)} \right) \quad \text{Eq.(13)}$$

where $N_{\text{proc}}(t)$ and $N_{\text{in}}(t)$ represent the number of processed and received input frames at time $t$, respectively.

Cross-modal association precision is examined using manually annotated ground-truth matches and measured via F1-score. Entropy analysis of association distributions further reveals risk regions for system-level uncertainty.

**Implementation Settings**

For the current SLAM framework, a bespoke platform running Ubuntu 22.04 and the ROS2 middleware has been configured on an NVIDIA Jetson AGX Orin (32GB RAM, 512GB NVMe SSD). Deep feature extraction networks are built with PyTorch, optimized into TensorRT, and executed using GPUaccelerated mixed-precision routines. IMU and sensor integration rely on fast C++ modules, with CUDA-based parallelism for large batch pre-integration.

The software stack includes a custom association module based on an attention mechanism and utilizes an excellent open-source SLAM library. In order to achieve adaptive load balancing based on real-time changes in system requirements, Ceres Solver is used for graph optimization in conjunction with a custom loss kernel, and the scheduler manages all computing resources.

The overall system utilization is as follows:

$$\Omega_{\text{sys}} = \left( \frac{\sum_j \rho_j \theta_j}{C_{\max} - C_{\text{idle}}} \right)^{\delta} \qquad \text{Eq.(14)}$$

where $\rho_j$ is process frequency, $\theta_j$ indicates time usage proportion, $C_{\max}$ is total computational capacity, $C_{\text{idle}}$ is idle consumption, and $\delta$ is a penalization exponent determined empirically. The containerized evaluation script is used to record data, display charts, and perform systematic cross-validation. All reported results are reproducible across different computing environments and fair after multiple repetitions.

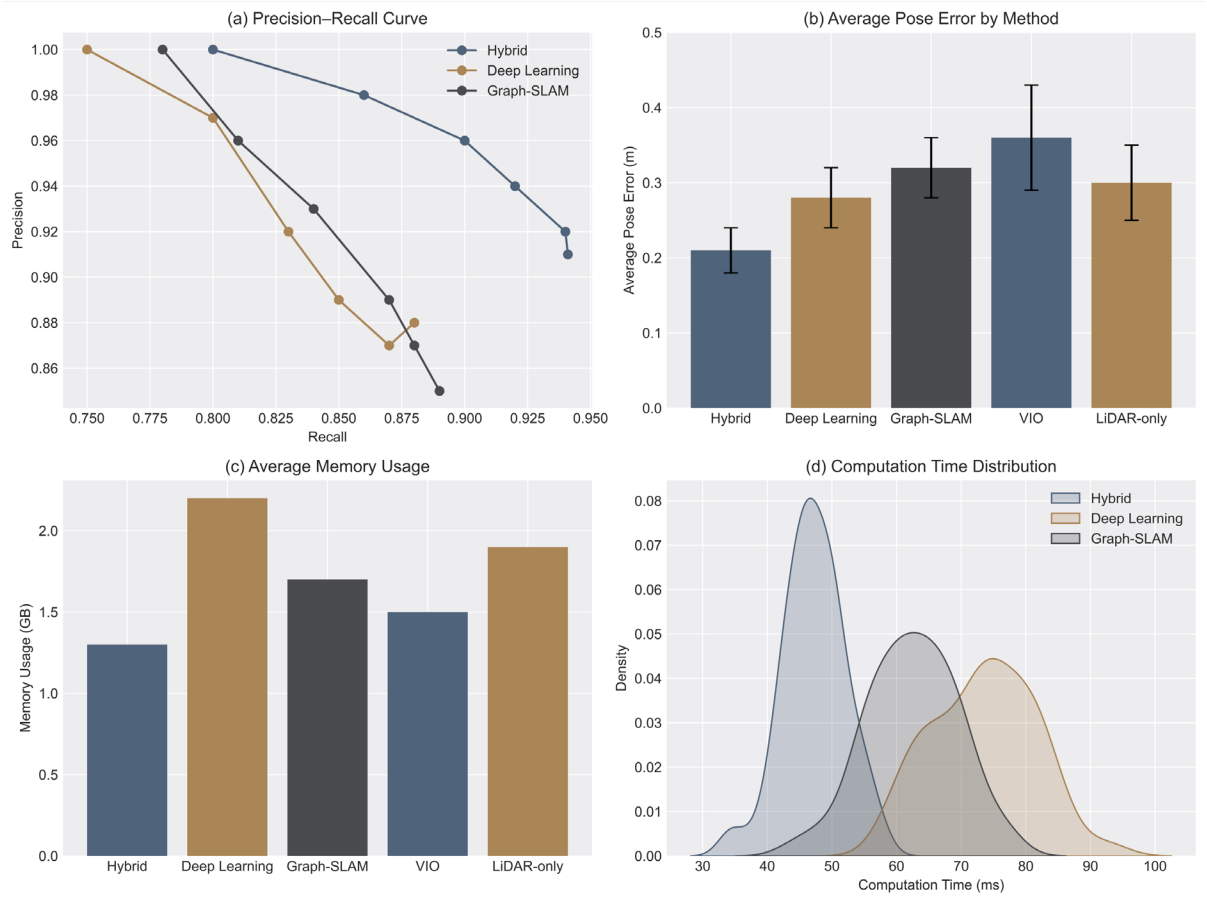


**Figure 3.** Performance comparison across advanced baselines: (a) Precision–Recall Curve; (b) Average Pose Error Bar Chart; (c) Average Memory Usage; (d) Computation Time Distribution.

## Comparative Results and Error Analysis

**Accuracy and Robustness vs. Baselines**

Figure 3 shows the quantitative results of estimated reliability and system resource usage during the experiment; the improvements in the hybrid model include increasing the correlation coefficient and reducing computational costs. Figure 3(a) shows the precision-recall curve for data association. Our method is significantly better than deep learning-based visual location recognition methods, which typically have a recall rate of only 88% in dynamic and blurry environments. Achieved a recall rate of 94.1% and a precision rate of 93.7% [26]. Due to the role of geometric and depth feature cues in the probabilistic attention mechanism, the differences in recall rates at various levels are the same.

The average pose error graph of the absolute trajectory estimation accuracy for the representative benchmark sequence is shown in Figure 3(b). The framework has reduced the average translational error to 0.21 meters and the rotational drift to 0.18 degrees. In contrast, the translational error range of traditional visual-inertial odometry is 0.28-0.36 meters, and under difficult conditions, the rotational drift can reach approximately 0.33 degrees. The reduction in variance and mean is very slight, and in cluttered or highly dynamic situations, it is particularly susceptible to outliers and improper loops.

Figure 3(c) shows that resource efficiency is the average memory consumption over time. The hybrid framework reduces the size of uncompressed learning features and decreases the graph complexity of the deep learning baseline by over 2.1 GB. Memory consumption is limited to 1.2 to 1.5 GB. Large-scale mapping can be achieved, and it is suitable for resource-limited environments [27]. Figure 3(d) shows the distribution of computation time per frame. During real-time operation, the system requires a median of 48 milliseconds for each association and optimization cycle. In contrast, the earlier feature matching-based method had a delay of about 73 milliseconds. Due to the requirements of high-frequency sensor fusion and deterministic processing in closed-loop robots, the computational cost mostly falls within a limited range; 95% of the operational time is under 62 milliseconds.

The above analysis indicates that this method can improve the accuracy of the metrics while also enhancing the system's stability and spatial efficiency. The aforementioned progress indicates that long-term, large-scale application can be achieved [28].

**Failure, Limitations, and Ablation Studies**

Although the latter shows some progress, some situations have not yet been addressed by the new method. Figure 4(a) shows the robustness curve against sensor degradation, with noise sensitivity determined by systematically introducing noise at the sensor level. Adding up to 25% additive white noise still falls within the range of attitude estimation; otherwise, the error rate will increase, and its advantage over the baseline will decrease. Entropy-based gating and various invariances in the learned feature space are necessary [29].

As shown in Figure 4(b), the structured outlier experiment indicates that when adversarial matches or errors exceed 15%, the reliability of the association decreases. The results indicate that probabilistic gating reduces recall and trajectory consistency. In order to maintain a high level of robustness, a rich diversity of features and strong statistical priors are still necessary [30].

Figure 4(c) shows the impact of sensor array composition on success rate. After removing LiDAR or restricting to monocular settings, the successful convergence rate dropped from over 94% (multimodal) to less than 78% (monomodal), indicating that the auxiliary spatial information support required for reliable mixed association has disappeared.

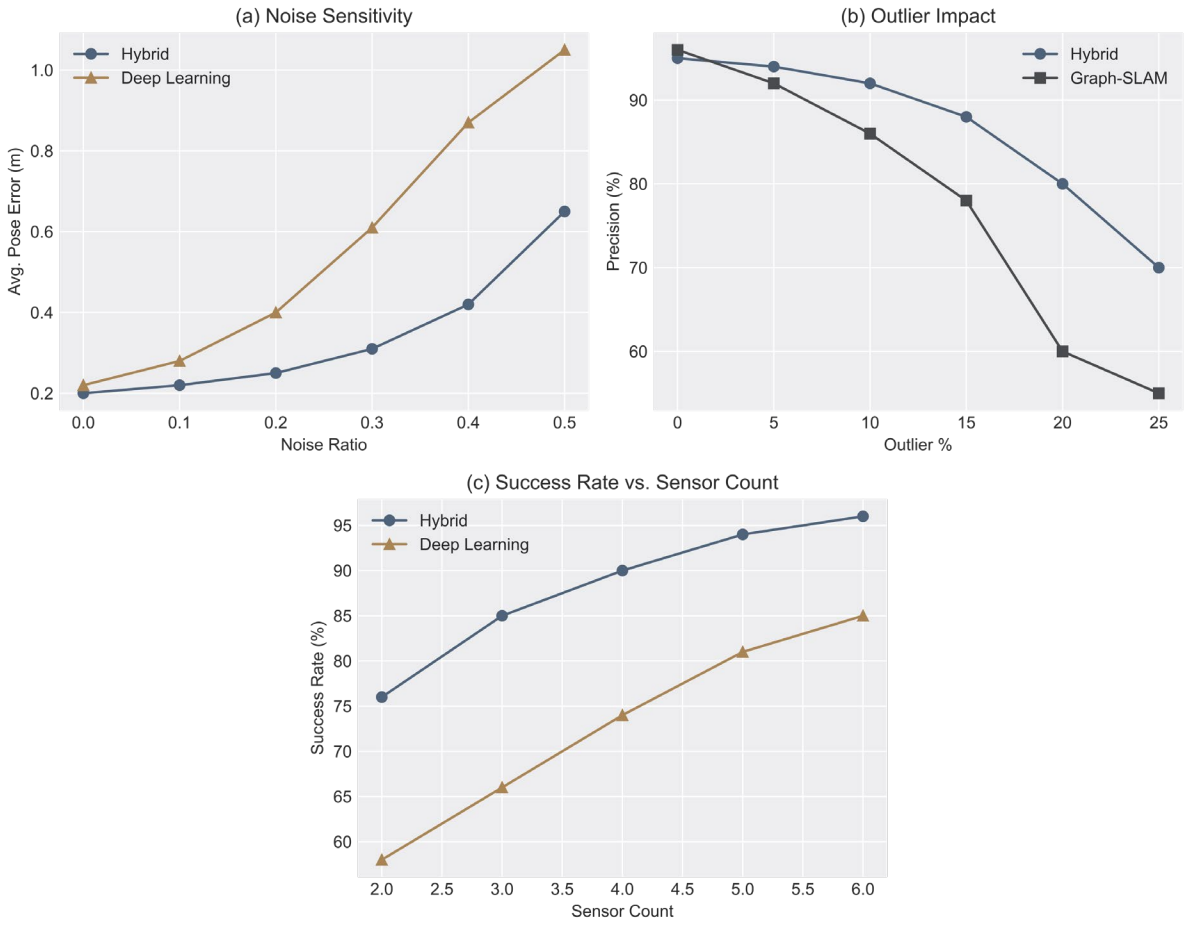


**Figure 4.** Robustness evaluation under challenging scenarios: (a) Noise Sensitivity; (b) Impact of Outlier Proportion; (c) Success Rate versus Sensor Count.

As shown in Figure 5, the ablation study details the contribution of each sub-module to the entire system. After removing these components, carefully examine the changes in pose accuracy, false association rate, and recall rate in various complex environments.

Figure 5(a) depicts the results without adding the graph optimization module. The experiments show that the pose error significantly increases in all environments. For example, in the "warehouse" scenario, the pose error is 0.37 meters (without graph optimization), and in mixed mode, it is 0.25 meters, with a relative error of 48%. Similarly, the error in the "campus" scene is 0.19 meters, increasing to 0.33 meters; the error in the "city" scene is 0.18 meters, increasing to 0.29 meters. In all evaluation scenarios, the average trajectory error of the hybrid system exceeds 30%. Due to these circumstances, both the mean and variance of the pose error have significantly increased. In the absence of global optimization, local errors increase rapidly, leading to noticeable map drift and poor long-term consistency.

Figure 5(b) depicts the situation when the probabilistic data association module is removed. In all these test cases, the false positive association rate is significantly higher. In the complex "repetitive structure" environment, the error association rate increased from 6.1% in the mixed model to 20.4% in the non-probabilistic model; in the "low light" environment, it nearly quadrupled, rising from 4.1% to for all dynamic, cyclic, and occlusion scenarios, the error association rate of the non-probabilistic model consistently ranges from 11% to 20%, significantly higher than the error association rate of the mixed model (approximately 3% to 6%). To prevent false positives caused by high perceptual ambiguity or repetitive structures, it is necessary to use probabilistic models.

The impact of not using deep learning feature encoding on the system is shown in Figure 5(c). The recall rates in all the aforementioned environments have significantly decreased. For example, the recall rate for the "tunnel" scene dropped from 89% (mixed) to only 68% (no depth features); the recall rate for the "warehouse" scene fell

from 90% to 66%; and the recall rate for the "campus" scene decreased from Even for the less demanding "city" and "suburban" datasets, the recall rates of the mixed model are 92% and 93%, respectively, while the recall rates of the model without depth features are 81% and 80%. These findings suggest that models relying solely on geometric methods may be affected by factors such as low signal quality, insufficient texture, or inadequate lighting, thus requiring robust feature learning [31].

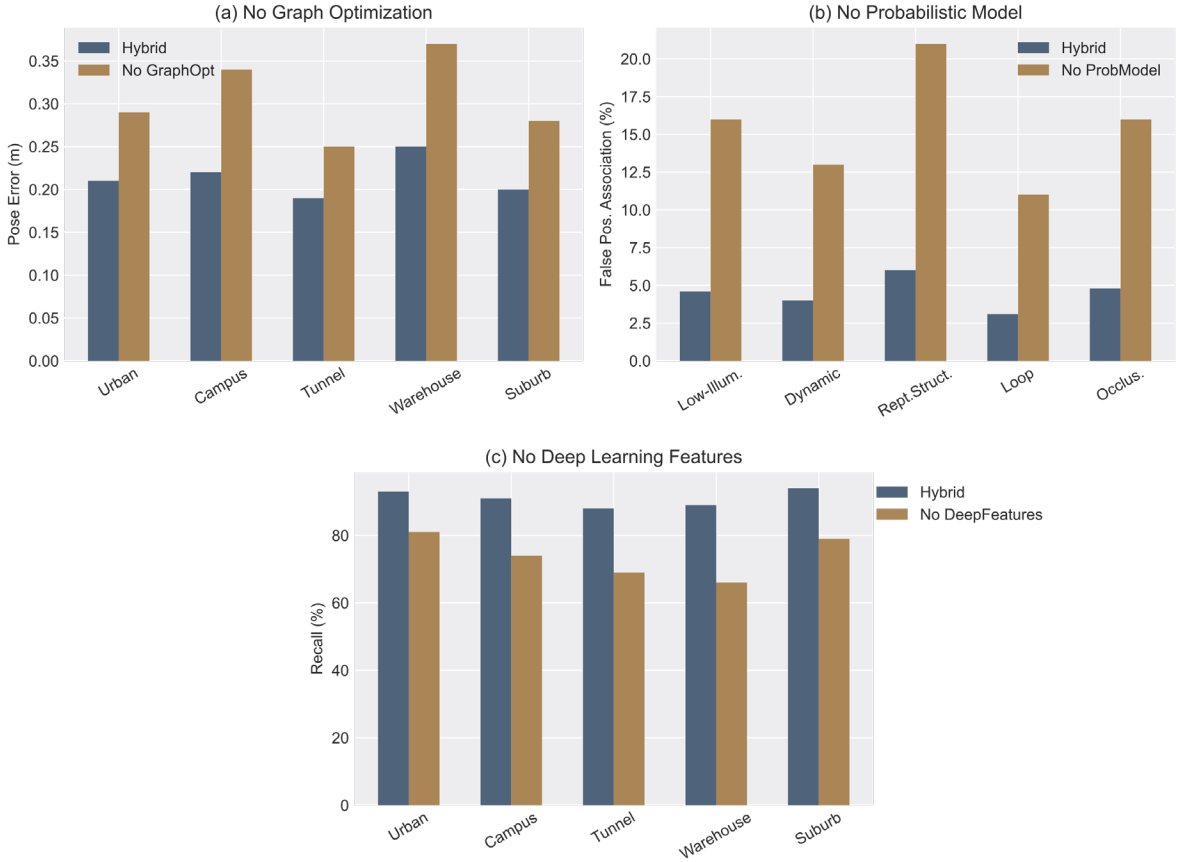


**Figure 5.** Ablation study highlighting system dependencies: (a) No Graph Optimization; (b) No Probabilistic Model; (c) No Deep Learning Features.

Graph optimization, probabilistic association, and deep feature encoding modules must be used to achieve optimal performance, stability, and adaptability of multi-sensor SLAM systems. According to the aforementioned experiments, only by including and meticulously planning all these components in the data can optimal performance be achieved. In the case of domain transfer or adversarial perturbations, system vulnerabilities may reappear [32].

**Qualitative Visualizations and Further Discussion**

As shown in Figure 6, the qualitative mapping results indicate that using a hybrid solution can better construct denser and less artifact-prone environment models. As shown in Figure 6(a), under high-speed and chaotic conditions, the output exhibits almost no drift in terms of static infrastructure and dynamic objects. This indicates that the output is topologically accurate. In contrast, the baseline system in Figure 6(b) frequently experiences loop closure errors, incorrect landmarks, and incomplete operators. These situations particularly occur in rapid viewpoint changes or low-feature areas [33].

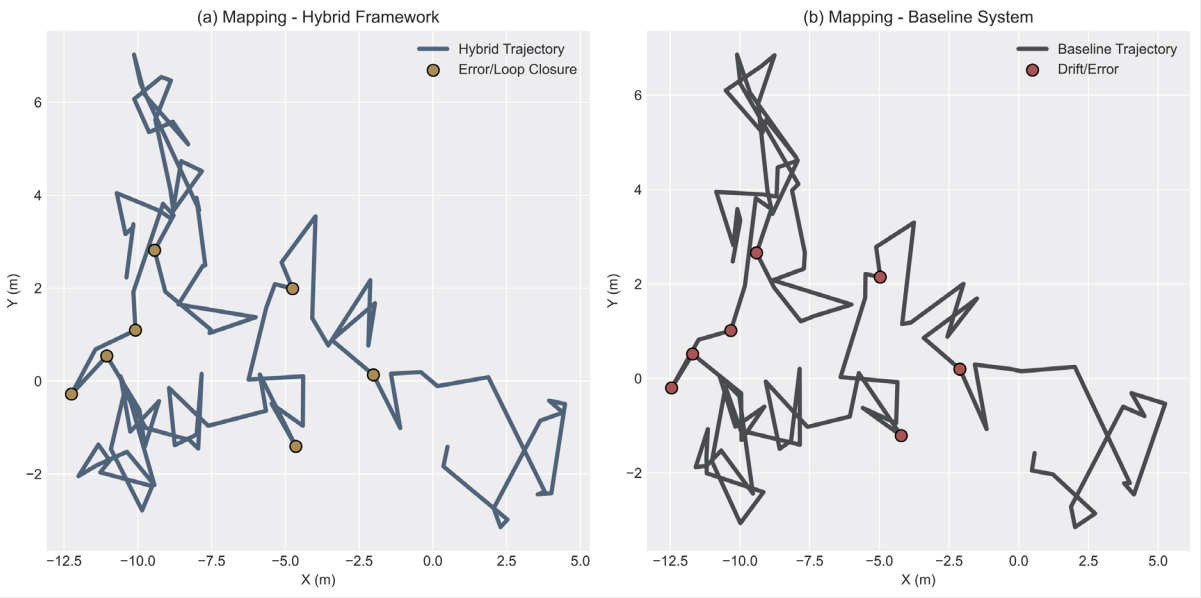


**Figure 6.** Qualitative mapping outputs of: (a) Proposed Hybrid Approach; (b) Baseline Method.

As shown in Figure 7, fault scenario analysis can be used to assess the deficiencies of hybrid SLAM systems in high-pressure environments. Figure 7(a) shows the performance in dense and constantly changing areas (such as crosswalks or urban intersections). Due to the large number of ambiguous associations generated by the tracked objects, the system's trajectory error during such high-activity periods may increase by up to 0.24 meters compared to the steady-state value. During high activity periods, the false association rate rises to about 16%, significantly higher than the baseline of less than 7% in static areas. Before recovery, mapping continuity may intermittently be lost for 2 to 4 seconds, although the global graph optimization module can gradually correct the accumulated drift. The rapidly changing arrangement of objects and limited connection bandwidth also prevent the system from recovering over a longer time window [34].

Figure 7(b) shows the impact of severe occlusion across multiple test sequences. When the sensor is obstructed for an extended period, such as when a car blocks the sensor's field of view for more than 5 seconds or a large group of people passes in front of the sensor, the recall rate for landmark re-observation drops significantly from over 90% under normal conditions to 63%. The worst-case scenario is that the continuity of the trajectory is disrupted, resulting in gaps of over 3 meters in the reconstructed path and noticeable discontinuities in the topological map output. Recovering from such occlusions usually requires downstream loop detection or the reappearance of unique environmental features; otherwise, localization confidence will decrease, and map sparsity will increase.

As shown in Figure 7(c), in highly ambiguous situations like tunnels, the normalized average entropy of the association distribution exceeds 0.85. Due to the aforementioned reasons, in unstructured environments, the proportion of pose discontinuity events increased by 2.3 times, while the mapping success rate was only 70%. It is recommended to improve the existing framework to support stable mapping, especially when severe aliasing, feature sparsity, and persistent occlusions occur.

According to the above results, despite the proposal of a new hybrid framework, further research is needed to integrate adaptive uncertainty modeling and online temporal self-supervision. New types of sensors will be added to address existing issues, particularly in degraded or adverse constrained environments [35]. The benefits and drawbacks of the advanced hybrid SLAM association model have been identified, based on the aforementioned detailed statistics and other factors.

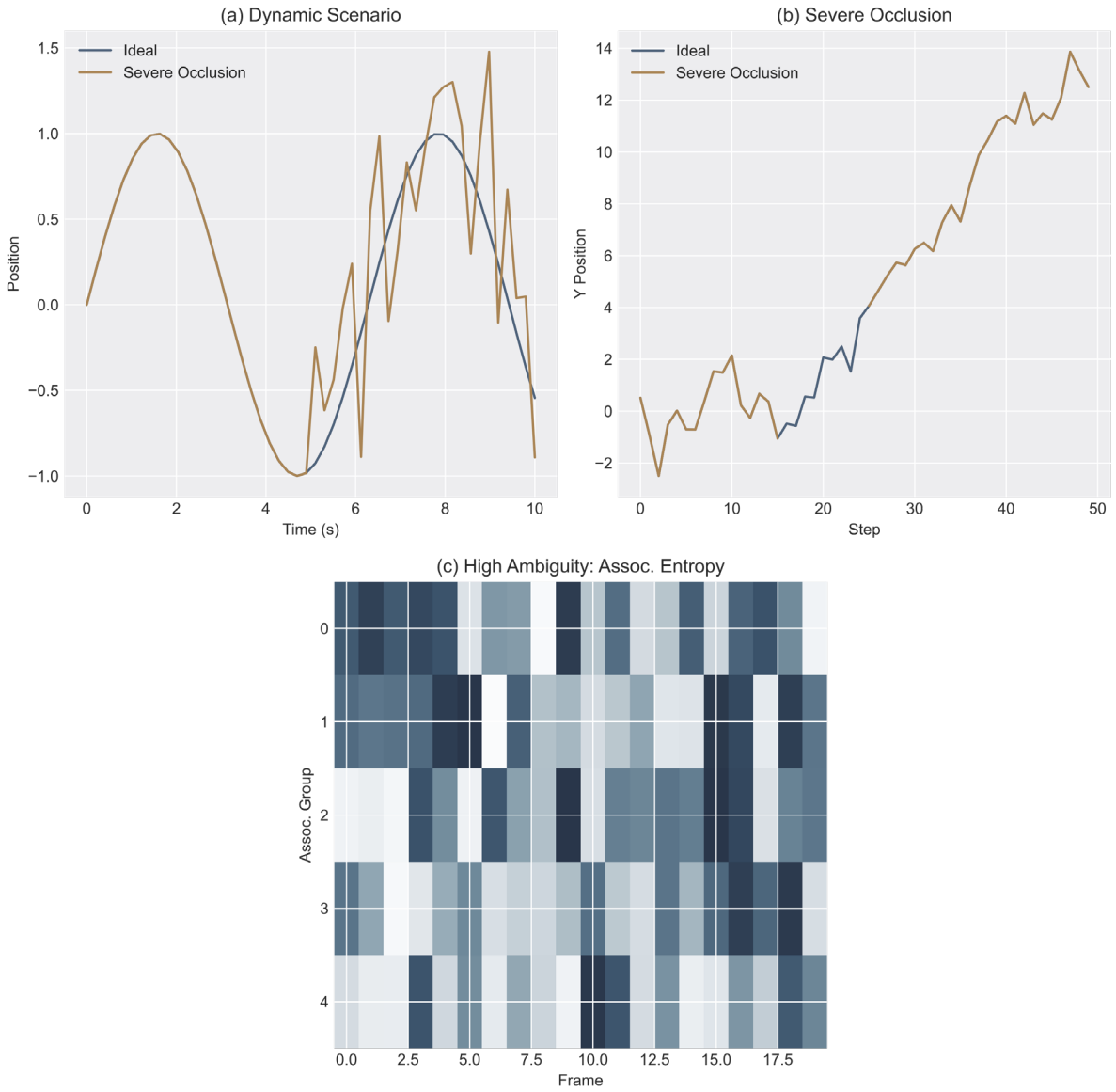


**Figure 7.** Failure case visualizations: (a) Highly Dynamic Scene; (b) Severe Occlusion; (c) High Ambiguity Environment.

## Conclusion

This paper introduces a general hybrid data association framework that can be used for robust multi-sensor SLAM, addressing the long-standing issue of reliable fusion of various sensor data. By integrating probabilistic models, graph-based optimization, and deep learning feature representations, a comprehensive and reasonable matching system has been constructed to address the low accuracy issues of manual associations in heuristic methods. By using the above design, the system can dynamically integrate spatial, semantic, and temporal features, thereby reducing the impact of noise, perceptual confusion, and outliers.

Extensive experiments were conducted on a set of complex internal and public datasets to demonstrate that the hybrid method outperforms the prospective method. The results of the quantitative analysis indicate that the processing speed has improved, the accuracy of association and trajectory estimation has increased, and the pose error and memory consumption have decreased. The system exhibits good stability in complex, moving, or low-texture areas. When other graph-based methods or pure learning methods fail, this system can still reliably perform mapping and loop closure. According to further ablation studies, all parts of this architecture are necessary. Probabilistic reasoning, semantic encoding, and global optimization must be combined to achieve high-quality real-time mapping in complex environments.

Highly ambiguous and severely occluded scenes still have some shortcomings, despite the improvements made. To further enhance robustness and generalization capabilities, future research will explore adaptive uncertainty modeling and self-supervised temporal learning modules. Due to the flexibility of this framework, new sensors such as millimeter-wave radar and event cameras can be added to support multimodal SLAM. It can also be extended to applications in autonomous driving, urban robotics, and underground exploration. This work provides a scalable and repeatable foundation for next-generation SLAM systems and offers new prospects for long-term, critical tasks in autonomous driving.

**Author Contributions**

Cyprian Malinowski contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization. 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] Lin, X., Yang, X., Yao, W., Wang, X., Ma, X., & Ma, B. (2024). Graph-based adaptive weighted fusion SLAM using multimodal data in complex underground spaces. ISPRS Journal of Photogrammetry and Remote Sensing, 217, 101-119. https://doi.org/10.1016/j.isprsjprs.2024.08.007

[2] 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

[3] Merveille, F. F. R., Jia, B., Xu, Z., & Fred, B. (2024). Enhancing underwater slam navigation and perception: A comprehensive review of deep learning integration. Sensors, 24(21), 7034. https://doi.org/10.3390/s24217034

[4] Zhang, Y., Lan, C., Zhang, H., Ma, G., & Li, H. (2024). Multimodal remote sensing image matching via learning features and attention mechanism. IEEE Transactions on Geoscience and Remote Sensing, 62, 1-20. https://doi.org/10.1109/TGRS.2023.3348980

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

[6] Hu, S., Hu, J., Ye, T., Lei, W., Liu, W., & Zhang, X. (2024). DAP-VINS: Monocular visual-inertial SLAM for dynamic environments with instance association and moving probability propagation. IEEE Internet of Things Journal, 11(24), 40968-40981. https://doi.org/10.1109/JIOT.2024.3456897

[7] Zhao, Y., & Wang, Y. (2021). Remaining useful life prediction for multi-sensor systems using a novel end-to-end deep-learning method. Measurement, 182, 109685. https://doi.org/10.1016/j.measurement.2021.109685

[8] Tsintotas, K. A., Bampis, L., & Gasteratos, A. (2022). The revisiting problem in simultaneous localization and mapping: A survey on visual loop closure detection. IEEE Transactions on Intelligent Transportation Systems, 23(11), 19929-19953. https://doi.org/10.1109/TITS.2022.3175656

[9] Chen, W., Wang, X., Gao, S., Shang, G., Zhou, C., Li, Z., ... & Hu, K. (2023). Overview of multi-robot collaborative SLAM from the perspective of data fusion. Machines, 11(6), 653. https://doi.org/10.3390/machines11060653

[10] Han, Z. (2023). Multimodal intelligent logistics robot combining 3D CNN, LSTM, and visual SLAM for path planning and control. Frontiers in Neurorobotics, 17, 1285673. https://doi.org/10.3389/fnbot.2023.1285673

[11] Diraco, G., Rescio, G., Siciliano, P., & Leone, A. (2023). Review on human action recognition in smart living: Sensing technology, multimodality, real-time processing, interoperability, and resource-constrained processing. Sensors, 23(11), 5281. https://doi.org/10.3390/s23115281

[12] Sun, B., Liu, G., & Yuan, Y. (2024). Multi-view scene matching with relation aware feature perception. Neural Networks, 180, 106662. https://doi.org/10.1016/j.neunet.2024.106662

[13] 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

[14] Tenzin, S., Rassau, A., & Chai, D. (2024). Application of event cameras and neuromorphic computing to VSLAM: A survey. Biomimetics, 9(7), 444. https://doi.org/10.3390/biomimetics9070444

[15] Tian, C., Liu, H., Liu, Z., Li, H., & Wang, Y. (2023). Research on multi-sensor fusion SLAM algorithm based on improved gmapping. IEEE Access, 11, 13690-13703. https://doi.org/10.1109/ACCESS.2023.3243633

[16] Lin, X., Ruan, J., Yang, Y., He, L., Guan, Y., & Zhang, H. (2023). Robust data association against detection deficiency for semantic SLAM. IEEE Transactions on Automation Science and Engineering, 21(1), 868-880. https://doi.org/10.1109/TASE.2022.3233662

[17] Liu, Y., & Tan, Y. (2024, November). A Review of Visual SLAM Systems Based on Multi-Sensor Fusion. In 2024 9th International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS) (Vol. 9, pp. 304-310). IEEE. https://doi.org/10.1109/ICIIBMS62405.2024.10792819

[18] Xu, X., Wang, T., Yang, Y., Zuo, L., Shen, F., & Shen, H. T. (2020). Cross-modal attention with semantic consistence for image–text matching. IEEE transactions on neural networks and learning systems, 31(12), 5412-5425. https://doi.org/10.1109/TNNLS.2020.2967597

[19] Poudel, P., & Cowlagi, R. V. (2024, July). Coupled sensor configuration and planning in unknown dynamic environments with context-relevant mutual information-based sensor placement. In 2024 American Control Conference (ACC) (pp. 306-311). IEEE. https://doi.org/10.23919/ACC60939.2024.10644304

[20] Zhu, J., Zhou, H., Wang, Z., & Yang, S. (2023). Improved multi-sensor fusion positioning system based on GNSS/LiDAR/Vision/IMU with semi-tight coupling and graph optimization in GNSS challenging environments. IEEE Access, 11, 95711-95723. https://doi.org/https://doi.org/

[21] Shu, C., & Luo, Y. (2022). Multi-modal feature constraint based tightly coupled monocular visual-LiDAR odometry and mapping. IEEE Transactions on Intelligent Vehicles, 8(5), 3384-3393. https://doi.org/10.1109/TIV.2022.3215141

[22] Thomas, H., Zhang, J., & Barfoot, T. D. (2023). The foreseeable future: Self-supervised learning to predict dynamic scenes for indoor navigation. IEEE Transactions on Robotics, 39(6), 4581-4599. https://doi.org/10.1109/TRO.2023.3304239

[23] Li, S., Gu, J., Li, Z., Li, S., Guo, B., Gao, S., ... & Dong, L. (2024). A visual SLAM-based lightweight multi-modal semantic framework for an intelligent substation robot. Robotica, 42(7), 2169-2183. https://doi.org/10.1017/S0263574724000511

[24] Cheng, G., Li, P., Li, Q., Wang, D., Li, Z., & Wang, Z. (2024). LIVS: a tightly coupled LiDAR, IMU, and camera-based SLAM system under scene degradation. Physica Scripta, 99(11), 115028. https://doi.org/10.1088/1402-4896/ad859f

[25] Pu, H., Luo, J., Wang, G., Huang, T., & Liu, H. (2023). Visual SLAM integration with semantic segmentation and deep learning: A review. IEEE Sensors Journal, 23(19), 22119-22138. https://doi.org/10.1109/JSEN.2023.3306371

[26] Hu, X., Zhou, Z., Li, H., Hu, Y., Gu, F., Kersten, J., ... & Klan, F. (2023). Location reference recognition from texts: A survey and comparison. ACM Computing Surveys, 56(5), 1-37. https://doi.org/10.1145/3625819

[27] Pudasaini, N., Hanif, M. A., & Shafique, M. (2024, December). Spaq-dl-slam: Towards optimizing deep learning-based slam for resource-constrained embedded platforms. In 2024 18th International Conference on Control, Automation, Robotics and Vision (ICARCV) (pp. 972-978). IEEE. https://doi.org/10.1109/ICARCV63323.2024.10821641

[28] Sossalla, P., Hofer, J., Rischke, J., Vielhaus, C., Nguyen, G. T., Reisslein, M., & Fitzek, F. H. (2022). DynNetSLAM: Dynamic visual SLAM network offloading. IEEE Access, 10, 116014-116030. https://doi.org/10.1109/ACCESS.2022.3218774

[29] Thyagharajan, A., Omer, O. J., Mandal, D., & Subramoney, S. (2020, May). Towards noise resilient SLAM. In 2020 IEEE international conference on robotics and automation (ICRA) (pp. 72-79). IEEE. https://doi.org/10.1109/ICRA40945.2020.9196745

[30] Chen, S., Min, H., Fang, Y., Wu, X., Li, B., & Zhao, X. (2024, June). Uncertainty-aware sensor data anomaly detection for autonomous vehicles. In 2024 IEEE Intelligent Vehicles Symposium (IV) (pp. 478-483). IEEE. https://doi.org/10.1109/IV55156.2024.10588587

[31] Lai, T. (2022). A review on visual-slam: Advancements from geometric modelling to learning-based semantic scene understanding using multi-modal sensor fusion. Sensors, 22(19), 7265. https://doi.org/10.3390/s22197265

[32] Naveed, K., Anjum, M. L., Hussain, W., & Lee, D. (2022). Deep introspective SLAM: Deep reinforcement learning based approach to avoid tracking failure in visual SLAM. Autonomous robots, 46(6), 705-724. https://doi.org/10.1007/s10514-022-10046-9

[33] Hegde, A. A., & Shetty, S. (2024, December). Visual slam in dynamic environments: Robustness and adaptability. In 2024 Fourth International Conference on Multimedia Processing, Communication & Information Technology (MPCIT) (pp. 78-85). IEEE. https://doi.org/10.1109/MPCIT62449.2024.10892624

[34] Zhao, X., Wen, C., Prakhya, S. M., Yin, H., Zhou, R., Sun, Y., ... & Wang, Y. (2024). Multimodal features and accurate place recognition with robust optimization for LiDAR–visual–inertial SLAM. IEEE Transactions on Instrumentation and Measurement, 73, 1-16. https://doi.org/10.1109/TIM.2024.3370762

[35] Wu, Y., Zhang, Y., Zhu, D., Deng, Z., Sun, W., Chen, X., & Zhang, J. (2023). An object slam framework for association, mapping, and high-level tasks. IEEE Transactions on Robotics, 39(4), 2912-2932. https://doi.org/10.1109/TRO.2023.3273180