A Real-Time Precise Positioning Method for Indoor Robots Based on Multi-Sensor Fusion

Zhen Xu*

Tianjin Electronic Information College, Tianjin, 300350, China
**Corresponding author: xuzhen@tjdz.edu.cn*

Abstract: *Achieving real-time, precise, and reliable localization for robots in indoor environments is a critical prerequisite for autonomous navigation and the execution of complex tasks. Single sensors are often limited by their inherent physical characteristics and susceptibility to environmental interference, making it difficult to maintain stable performance in dynamic scenes. To address this issue, this paper investigates a real-time precise localization method based on multi-sensor information fusion. By systematically analyzing the data characteristics of heterogeneous sensors, including LiDAR, visual sensors, inertial measurement units (IMUs), and wheel encoders, the method establishes accurate noise models and designs spatiotemporal synchronization and anomaly detection mechanisms, thereby laying a solid data foundation for fusion. An adaptive fusion localization model based on state space is constructed, and an algorithm that dynamically adjusts sensor weights according to online data quality is proposed. Furthermore, multi-scale feature fusion and a robust graph optimization framework are employed for pose estimation, enhancing the system's accuracy and fault tolerance. Finally, an embedded real-time localization system is designed, which ensures the efficient execution of the algorithm from the perspectives of hardware architecture, computational optimization, and resource scheduling. Experimental validation demonstrates that this method achieves centimeter-level localization accuracy in both static and dynamic indoor environments, while meeting the requirements for high real-time performance, strong stability, and fast convergence, thus providing a reliable pose estimation solution for indoor mobile robots.*

Keywords: *indoor robots; multi-sensor fusion; real-time localization; state estimation; robustness; embedded systems*

Introduction

In the field of autonomous mobile robotics, accurate and reliable real-time localization constitutes a fundamental component for executing path planning, environmental interaction, and task decision-making. Indoor environments present challenges such as structural complexity, dynamic obstacles, variable lighting conditions, and signal occlusion. These challenges impose significant limitations on the accuracy and reliability of localization methods reliant on a single information source. For instance, odometry suffers from cumulative errors; LiDAR is prone to failure in feature-sparse environments or scenes with glass walls; and visual methods are sensitive to changes in lighting and texture. Therefore, leveraging the complementary information from multiple sensors to enhance the overall performance of the localization system through fusion algorithms holds clear research necessity and engineering value. This study aims to construct a complete technical framework, encompassing data preprocessing, fusion models, to system implementation, to address the problem of high-precision, high-robustness real-time localization in indoor scenarios. This paper first models and preprocesses the characteristics of multi-sensor data. Subsequently, it designs a multi-source information fusion localization model capable of adapting to environmental changes. Finally, the integrated performance of this real-time localization system is implemented and validated on an embedded platform, providing an effective method to enhance the autonomy of indoor robots.

1. Analysis and Preprocessing of Multi-Sensor Data Characteristics

1.1 Sensor Selection and Characteristic Analysis for Indoor Localization

The performance of an indoor robot localization system is closely related to the physical characteristics and applicable boundaries of the sensors employed. Light Detection and Ranging (LiDAR) acquires high-precision, high-angular-resolution two-dimensional or three-dimensional environmental geometric contour information by emitting laser beams. Sufficiently high ranging accuracy can be achieved when the number of laser beams is adequate. However, LiDAR is highly prone to failure in the presence of transparent or strongly light-absorbing objects in the environment, and its data density attenuates with increasing distance. An Inertial Measurement Unit (IMU) provides high-frequency angular velocity and linear acceleration data of the body, possessing a completely autonomous capability for short-term motion estimation. Nevertheless, its measurement errors accumulate over time, leading to a significant drift issue[1]. Visual sensors (such as monocular, stereo, or RGB-D cameras) can capture rich environmental texture and feature information, making them suitable for feature-based localization and scene recognition. However, their performance is significantly affected by lighting variations, dynamic object interference, and texture absence. Wheel encoders estimate trajectory by recording wheel rotation angles, offering reliable short-term accuracy under non-slip planar motion models. However, their accuracy is directly constrained by factors such as wheel slippage, uneven ground, and mechanical wear. Therefore, it is necessary to quantitatively analyze the accuracy, frequency, error characteristics, and complementarity of each sensor in typical indoor environments. This analysis establishes the physical constraints and informational foundation at the sensor level for the subsequent fusion model.

1.2 Spatiotemporal Synchronization Method for Heterogeneous Sensors

The prerequisite for multi-sensor fusion is ensuring that all data streams exist within a unified and precise spatiotemporal reference frame. Temporal synchronization aims to address the timestamp inconsistency issues arising from each sensor's independent clock, as well as differing data acquisition and publication cycles. Hardware synchronization triggers all sensors to sample simultaneously by sharing an external clock pulse signal, which can achieve microsecond-level synchronization accuracy but imposes specific requirements on hardware interfaces. Software synchronization, on the other hand, performs timestamp alignment and interpolation processing on data streams from different sensors based on a master clock (typically the system clock or a high-precision sensor clock). Common methods include time alignment algorithms based on nearest-neighbor matching or linear interpolation. Spatial synchronization, also known as extrinsic calibration, requires precise determination of the rigid body transformation relationships (rotation matrix and translation vector) between each sensor relative to the robot's body coordinate system or a selected reference coordinate system. For LiDAR-camera combinations, a joint calibration method based on specific calibration board patterns can be employed. For calibrating the IMU with other sensors, it is necessary to estimate their relative attitude and installation position deviations through excitation motion. Spatiotemporal synchronization errors directly constitute systematic errors in the fusion system and must be accurately modeled and compensated[2].

1.3 Noise Models and Filtering Algorithms for Raw Data

Raw sensor observation data is inevitably contaminated by various types of noise. Establishing accurate noise statistical models is crucial for effective data preprocessing and state estimation. Inertial sensor noise is typically modeled as a combination model containing Gaussian white noise and random walk bias. LiDAR ranging noise is often treated as additive Gaussian noise related to distance, while also considering outliers caused by specular reflection. Visual feature point observation noise is related to image pyramid levels and matching accuracy. In response to these noise characteristics, corresponding filtering algorithms must be employed for preprocessing. The Kalman filter and its variants (such as the Extended Kalman Filter, EKF) are suitable for linear or linearizable systems with Gaussian noise. For non-Gaussian situations or those with significant outliers, robust filtering methods (such as those based on M-estimators) or sliding window statistical methods are more effective. Particle filters demonstrate advantages in handling nonlinear and non-Gaussian problems, but incur higher computational costs. During the construction of an actual system, it is necessary to adapt the noise model based on the characteristic parameters of the selected sensors, determine parameter identification methods, and verify the suitability and efficiency of this model in suppressing specific noise and improving the signal-to-noise ratio.

1.4 Sensor Data Quality Assessment and Anomaly Detection

In dynamic and complex indoor environments, individual sensors may experience performance degradation or instantaneous failure due to external interference or internal malfunctions. Therefore, online assessment of the reliability of each sensor's data stream is crucial for the robustness of the fusion system. Data quality assessment can be based on internal consistency checks. For example, accelerometer and gyroscope data from an IMU should conform to physical constraints under static conditions. It can also be based on external consistency checks, such as comparing the short-term estimation results from one sensor with independent observations from another sensor. Common quantitative metrics include the Signal-to-Noise Ratio (SNR), data validity probability, and Chi-square tests on innovation sequences. Anomaly detection aims to identify and eliminate gross errors or invalid data in real-time. Methods range from simple rule-based threshold judgments (e.g., exceeding physically possible ranges) and hypothesis testing based on statistical models (e.g., Gaussian Mixture Models, GMM), to anomaly pattern recognition based on machine learning (e.g., Isolation Forest, autoencoders). Designing a lightweight, low-latency quality assessment and anomaly detection module, which can provide dynamic sensor confidence weights for subsequent fusion algorithms, is a necessary component for achieving highly reliable localization.

2. Construction of a Multi-Source Information Fusion Localization Model

2.1 State-Space-Based Fusion Localization Model Architecture

The core of multi-sensor fusion localization lies in constructing a unified state-space model. This model collectively defines parameters such as the robot's pose, velocity, and systematic sensor biases as the system's state vector. The model typically adopts a theoretical framework based on Bayesian estimation, transforming the localization problem into the continuous estimation and update of the posterior probability density of the system state. The system state equation describes the robot's kinematic model, usually employing high-frequency IMU data as input for the prediction step to perform a prior estimation of the state vector. The observation equation, on the other hand, establishes the nonlinear geometric and physical relationships between the system state and various sensor measurements (such as LiDAR point cloud matching residuals, visual feature reprojection errors, and encoder odometry increments). By incorporating the observation information from heterogeneous sensors as constraints into a unified optimization objective function or recursive filtering framework, deep information fusion is achieved[3]. This architecture must clearly define the composition of the state vector, the mathematical formulation of the motion and observation models, and analyze the contributions of different sensor observations to the observability of different components of the state vector, thereby providing a rigorous theoretical foundation for subsequent fusion algorithms.

2.2 Online Estimation Algorithm for Adaptive Weighting Factors

During the fusion process, the instantaneous accuracy and reliability of different sensors vary dynamically under different environmental conditions. A fusion strategy with fixed weights struggles to adapt to such variations, potentially leading to decreased localization accuracy or even divergence. The objective of the online estimation algorithm for adaptive weighting factors is to dynamically adjust each sensor's contribution to the fusion objective function or filter update based on the real-time quality of its data. A typical method is based on the covariance matching principle of innovation sequences. This method estimates and adjusts the observation noise covariance matrix of a sensor online by comparing the consistency between the actual observation residuals and the theoretically predicted covariance, thereby indirectly altering its weight. Another approach involves directly constructing a weighting function related to data quality assessment metrics (such as the number of feature matches, point cloud matching scores, or IMU static detection variance), transforming qualitative assessments into quantitative weighting coefficients. This algorithm must meet real-time requirements and ensure the smoothness of weight updates to avoid oscillations in state estimation. Its essence is to implement a closed-loop feedback control that is aware of sensor reliability.

2.3 Multi-Scale Feature Fusion and Pose Estimation Method

To fully leverage the complementary environmental information provided by different sensors, a multi-scale feature fusion mechanism needs to be designed. On the geometric scale, LiDAR provides accurate local surface elements and edge features, while visual sensors offer rich texture corners and line segment features. Associating and fusing these heterogeneous features in space enables the construction of a more discriminative environmental representation. On the temporal scale, the fusion algorithm must handle high-frequency IMU predictions, medium-frequency visual odometry updates, and potentially low-frequency global matching corrections (such as loop closure detection based on point clouds or visual features). The pose estimation method typically employs a graph optimization-based framework. This framework treats state variables at different times as nodes and uses constraints such as IMU pre-integration, relative pose observations, and absolute pose observations as edges to construct a sparse pose graph. By solving for the maximum a posteriori probability estimate of this graph, a globally consistent trajectory is obtained. This method effectively fuses multi-scale, multi-frequency observation information and naturally supports loop closure detection to eliminate accumulated errors.

2.4 Design of Robustness and Fault Tolerance Mechanisms for the Fusion System

When confronted with situations where sensor observations become entirely unreliable due to temporary sensor failure, severe interference, or extreme environments, the fusion system must possess inherent robustness and fault tolerance capabilities. The mechanism design must be implemented across multiple levels. At the data level, based on anomaly detection results, a gating strategy can be designed to temporarily isolate sensor data streams identified as faulty, preventing them from contaminating the state estimate. At the model level, employing robust kernel functions (such as the Huber kernel or Cauchy kernel) to replace the quadratic cost function in least squares can mitigate the impact of outlier observations on the overall optimization objective. At the state estimation level, maintaining multiple parallel hypotheses or adopting multiple-model adaptive estimation methods can address sudden changes in environmental features (e.g., entering a completely new scene through a door). System-level design must incorporate state health monitoring and recovery strategies. For instance, when visual information is lost for an extended period, the system should be able to automatically degrade to a fusion mode primarily reliant on LiDAR and IMU, and perform re-localization to reintegrate into the global coordinate system once vision is restored. This hierarchical fault-tolerant design ensures the continuous availability and stability of the localization system under non-ideal conditions[4].

3. Implementation and Performance Validation of the Real-Time Localization System

3.1 Design of the Embedded Fusion Localization System Architecture

To achieve real-time autonomous localization for indoor robots, the aforementioned fusion algorithms must be deployed on the robot's embedded computing platform. This hardware architecture typically comprises a sensor interface layer, a core processing unit, and an output layer. The sensor interface layer is responsible for reliably receiving raw data streams from LiDAR, IMU, visual sensors, and encoders via specific communication protocols (such as UART, SPI, CAN, or Ethernet) and performing initial hardware timestamp labeling. The selection of the core processing unit requires balancing computational power, power consumption, and real-time requirements. A common solution is a heterogeneous computing platform, for instance, utilizing an ARM CPU to execute sensor drivers, data synchronization, and task scheduling, while employing an FPGA or GPU to accelerate computation-intensive tasks such as point cloud processing and image feature extraction. The software architecture adopts a modular design, decoupling functions like data acquisition, preprocessing, fusion computation, and state publishing into independent threads or processes, which exchange data with low latency via shared memory or message middleware. The system output layer publishes high-precision robot pose estimation results at a fixed frequency and concurrently provides system health status and covariance information for use by upper-layer navigation and decision-making modules.

3.2 Computational Optimization and Resource Scheduling for Real-Time Localization

Meeting stringent real-time constraints is the core challenge for embedded deployment, necessitating multi-level optimization of the fusion localization algorithm. At the algorithm level, incremental and sliding window optimization strategies are adopted to avoid computational delays caused by global batch optimization. Sparsity analysis and fixed-pattern precomputation are performed on Jacobian matrices in nonlinear optimization to reduce online computation load. Parts of the computation (such as specific matrix inversions) are converted into lookup table operations. At the code level, techniques like compiler optimization directives, loop unrolling, and memory alignment are employed to enhance execution efficiency. At the computational resource scheduling level, a deterministic task scheduling strategy must be designed. Based on the computation time and data dependencies of each module, CPU core resources are allocated reasonably to ensure that high-priority fusion update threads are not blocked. Tasks with longer execution times, such as global loop closure detection, are set as low-priority background threads for asynchronous execution to avoid interfering with the periodicity of the main localization thread. Through systematic optimization and scheduling, the full pipeline latency from sensor data input to pose result output is ensured to remain stable at the millisecond level.

3.3 Localization Accuracy Testing in Static and Dynamic Environments

The performance of the localization system requires quantitative evaluation through meticulously designed experiments. The experimental environment must cover typical indoor scenarios, including long corridors, open halls, structurally repetitive areas, and regions with variable lighting. Static environment testing aims to evaluate the system's absolute accuracy and repeatability precision. Using a high-precision optical motion capture system or a total station as the ground truth reference, the robot is positioned stationary at multiple known calibration points. The localization system's outputs are recorded and subjected to error statistics. Dynamic environment testing evaluates the localization accuracy and smoothness of the system during motion. The robot is controlled to follow a preset closed trajectory (e.g., a rectangle or a figure-eight). Using the trajectory recorded by the motion capture system as the ground truth, the accumulation of positional and attitude errors over time is calculated, with particular attention paid to the drift error at the point of trajectory closure[5]. Furthermore, dynamic interference must be introduced, such as adding moving pedestrians or objects within the robot's field of view, to test the system's robustness in dynamic scenes. Evaluation metrics include absolute position error, relative pose error, root mean square error, and the statistical distribution of errors.

3.4 Analysis of System Real-Time Performance, Stability, and Convergence

Beyond localization accuracy, the real-time performance, stability, and convergence of the system as a whole are critical for evaluating its engineering usability. Real-time performance analysis verifies whether the system meets the requirements of control and planning modules by measuring and statistically analyzing the period jitter of the system's localization output, the end-to-end latency from data acquisition to output, and the latency variation under high-load scenarios (such as dense point clouds or multiple visual features). Stability analysis focuses on the system's performance during long-duration operation. It involves conducting continuous operation tests for several hours or more, monitoring changes in internal state variables such as pose estimation covariance, optimization residuals, and sensor weights, to confirm the absence of a divergent trend characterized by slowly accumulating errors or unbounded growth in estimation variance[6]. Convergence testing examines the system's re-localization capability when the initial position is uncertain or tracking is lost. Given a relatively large initial pose deviation, this test observes the time and travel distance required for the system to converge to the correct pose and analyzes whether the convergence process is smooth and free of oscillation. These three aspects of quantitative analysis together constitute a comprehensive evaluation of the performance of the real-time fusion localization system.

Conclusion

This study addresses the demand for real-time, precise localization in indoor robots by proposing and implementing a complete localization method based on multi-sensor fusion. Through in-depth analysis of the data characteristics from LiDAR, visual sensors, IMUs, and encoders, a preprocessing pipeline encompassing spatiotemporal synchronization, noise filtering, and quality assessment was established, providing high-quality, consistent data input for fusion. The constructed state-space-based adaptive fusion localization model can dynamically adjust fusion weights according to the real-time reliability of sensors. Furthermore, by leveraging multi-scale feature fusion and a robust optimization framework, the accuracy and environmental adaptability of pose estimation were effectively enhanced. Implementation and optimization within an embedded system ensured the algorithm meets strict real-time constraints. Comprehensive system test results demonstrate that the proposed method achieves the expected localization accuracy across various indoor scenarios while exhibiting commendable stability and convergence. Future work may focus on exploring more lightweight deep learning-based feature extraction and fusion networks to further improve the system's adaptability in extreme environments, investigating collaborative localization among multiple robots to extend the system's application scope, and optimizing the algorithm to be compatible with lower-cost sensor configurations, thereby promoting the widespread deployment of this technology in practical applications.

References

[1] Zheng, Zhong. Research on Indoor Mobile Robot Navigation Method Based on Multi-Sensor Fusion. 2025. Changchun University of Technology, MA thesis.
[2] Li, Chenming. Research on Indoor Robot Navigation Based on Multi-Sensor Fusion. 2025. Xi'an Technological University, MA thesis.
[3] Jia, Tianpeng. Research on Multi-Sensor Fusion Localization and Navigation for Indoor Mobile Robots. 2024. Beijing University of Posts and Telecommunications, MA thesis.
[4] Zhu, Chengzhi. Research on Multi-Sensor Fusion Mapping and Path Planning Algorithms for Indoor Service Robots. 2024. Yanshan University, MA thesis.
[5] Quan, Zhenhua, Wu Bin, and Luo Liang. "Research on Sensor Fusion Localization and Tracking Algorithm for Indoor Mobile Robots." Computer Simulation, vol. 40, no. 12, 2023, pp. 481-486.
[6] Qiu, Hao. Research on Indoor Robot Navigation Technology Based on Multi-Sensor Fusion. 2023. Nanjing University of Information Science and Technology, MA thesis.