

Multi-Sensor Fusion in Robotics: Core Frameworks, Practical Implementations and Development Trends

Guoxuan Cheng *

School of Marine Engineering, Harbin Institute of Technology, Weihai, China

* Corresponding Author Email: 2023212678@stu.hit.edu.cn

Abstract. As a core technology for enhancing environmental perception and autonomous decision-making in robots, multi-sensor fusion has advanced significantly in both theoretical and engineering aspects from 2015 to 2025. This paper reviews research progress through bibliometrics and case studies, identifying a three-stage evolution in algorithms: from Kalman Filter (KF)-based linear systems achieving around 40% accuracy improvement, to Particle Filter (PF) handling non-Gaussian noise with 85% localization accuracy, to deep learning-enabled end-to-end fusion reaching 92.7% classification accuracy. It analyzes common vision-LiDAR-IMU (Inertial Measurement Unit) configurations and compares traditional methods like EKF (Extended Kalman Filter) and UKF (Unscented Kalman Filter) with newer approaches such as CNN (Convolutional Neural Network) and GNN (Convolutional Neural Network). For example, LiDAR-vision SLAM (Simultaneous Localization and Mapping) attains 0.12 m RMSE (Root Mean Square Error). In applications including industrial robots, service robots, and autonomous driving, multi-sensor fusion improves localization accuracy to ± 0.05 mm, dynamic obstacle avoidance success to 97%, and reduces mapping error to 2.3%. The study also addresses ongoing challenges like spatiotemporal registration, heterogeneous data fusion, and computational efficiency, while highlighting emerging trends such as federated learning and spiking neural networks. It offers theoretical and practical guidance for algorithm selection and optimization in multi-sensor robot systems.

Keywords: Multi-sensor fusion; robots; algorithms.

1. Introduction

1.1. Background and Significance

Multi-sensor fusion algorithms play a crucial role in intelligent systems, greatly enhancing environmental perception capabilities. A single sensor has obvious limitations: for example, vision is affected by illumination, LiDAR (Light Detection and Ranging) is difficult to identify transparent objects, and IMU has cumulative errors. By fusing multi-source data through algorithms such as KF and Bayesian estimation, a more stable and reliable environmental model can be constructed. Studies have shown that the combination of LiDAR and vision can reduce the localization error of mobile robots by more than 40%, and it also has excellent performance in dynamic obstacle avoidance. This technology has promoted the development of fields such as autonomous driving and industrial inspection; for instance, Boston Dynamics' Atlas robot can adjust its posture in real time in complex terrain. In the field of medical rehabilitation, lower limb exoskeletons integrating electromyographic and force sensors can accurately identify patients' movement intentions; in intelligent manufacturing, the fusion of GNSS (Global Navigation Satellite System) and tilt sensors controls the construction accuracy within ± 2 mm. According to the report of the International Federation of Robotics, the failure rate of industrial robots adopting multi-sensor fusion has decreased significantly. Technological progress has also driven the upgrading of the industrial chain, such as the reduction in the cost of millimeter-wave radar and the increase in the shipment volume of MEMS (Micro-Electro-Mechanical Systems) inertial sensors. This study aims to analyze the theoretical framework of fusion algorithms, evaluate their performance in scenarios such as SLAM and human-robot interaction, and provide methodological support for solving problems such as sensor spatiotemporal registration and heterogeneous data fusion.

1.2. Purpose and Structure

Multi-sensor fusion algorithms play a critical role in robotic perception systems, and their development directly influences the accuracy with which intelligent systems interpret their environment and make decisions. This paper consolidates authoritative domestic and international literature from the past five years, focusing on the evolution of algorithmic principles and trends in application expansion. Particular emphasis is placed on areas such as KF optimization, deep learning-based feature-level fusion, and real-time performance in multi-target tracking, highlighting the ongoing shift from traditional probabilistic statistics to data-driven paradigms.

The search strategy employed three keyword categories: robot types (mobile, service, industrial), sensor fusion terms (multi-sensor fusion, information fusion), and specific algorithms (KF, PF, deep learning fusion). Comprehensive coverage was ensured through nested Boolean search operators. Literature screening prioritized relevance (core fusion algorithms such as KF), academic rigor (e.g., well-designed studies like Hong's 2019 work), and timeliness (preference for publications within five years, while retaining foundational theories) [1,2]. Evaluation incorporated both quantitative metrics (e.g., citation count and journal impact, as exemplified by Wang, L. B.'s influential 2022 study) and qualitative assessments (theoretical innovation and application value, such as Wu's 2021 research on neural network-based fusion) [3, 4]. Qian's team's image fusion evaluation system was included for objective standards.

The content covers three dimensions: basic theory, technical methods, and application cases. At the theoretical level, this paper analyzes the JDL (Joint Directors of Laboratories) model, Bayesian estimation, and D-S (Dempster-Shafer) evidence theory; at the technical method level, it compares the adaptability of KF (including EKF and UKF), PF, CNN, and GNN in heterogeneous data fusion; at the application case level, it selects mobile robot SLAM, lower limb exoskeleton gait recognition, and industrial sorting systems, and quantitatively evaluates key indicators such as localization error and recognition rate.

The structure adopts a "theory-method-application-trend" framework. The first chapter lays a mathematical foundation and constructs a model architecture; the second chapter reviews the mechanisms and applicable boundaries of mainstream algorithms, and discusses the innovative achievements of neural networks and classical filtering algorithms; the third chapter verifies the robustness of algorithms by combining twelve types of application scenarios; the final chapter summarizes the problems of multi-modal data spatiotemporal registration and nonlinear modeling limitations, and looks forward to the potential of new technologies such as federated learning and spiking neural networks in cross-domain collaborative perception.

2. Theories and Evolution

2.1. Core Theoretical Framework

Multi-sensor fusion technology combines information theory, probability statistics, and artificial intelligence, and uses the complementarity and redundancy of sensor data to improve the accuracy of environmental perception. Its framework includes the data layer, feature layer, and decision layer: the data layer processes original observations, such as using KF for state estimation; the feature layer extracts multi-source features and realizes spatial mapping through neural networks; the decision layer integrates the independent judgments of each sensor, and D-S evidence theory or Bayesian networks are commonly used.

Extended KF (EKF) addresses nonlinear problems, PF handles non-Gaussian noise, and Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks are used for target recognition and temporal fusion, respectively. In robot systems, multi-modal sensors work together to construct spatial cognition: for example, 2D LiDAR is used for plane scanning, depth cameras obtain RGB-D data, and Inertial Measurement Units (IMUs) provide pose information. Hardware adopts a time synchronization mechanism, and functions rely on algorithm chains; for

instance, the SLAM module fuses LiDAR and vision to construct maps, and the obstacle avoidance module combines ultrasound and infrared to realize path planning.

The key of this technology is to overcome the limitations of a single modality: vision sensors perform well in texture-rich scenarios but are limited by illumination; millimeter-wave radar has strong penetration but is difficult to identify categories. Fusion algorithms adjust confidence through adaptive weighting to improve localization accuracy. In industry, intelligent excavators integrate multiple sensors to achieve precise control of buckets; service robots maintain a high obstacle avoidance success rate in low-light environments; autonomous driving adopts a three-level architecture of LiDAR, vision, and millimeter-wave radar to realize long-distance obstacle recognition.

2.2. Evolution Process

Multi-sensor fusion algorithms originated in the 1960s. The emergence of KF marked the entry of this technology into a systematic stage. It realizes the optimal estimation of linear systems through a recursive method and is widely used in aerospace and navigation fields. In the 1980s, to address the problem of nonlinear systems, Extended KF (EKF) and Unscented KF (UKF) emerged, which are used to handle nonlinearity and model uncertainty. However, EKF is highly dependent on models, which limits its application in complex environments.

At the beginning of the 21st century, PF (PF) became a key technology for handling nonlinear and non-Gaussian systems, and achieved breakthroughs especially in mobile robot localization and SLAM. Based on the Monte Carlo method, PF uses a large number of samples to approximate the posterior distribution, making it suitable for multi-modal and non-Gaussian noise environments. With the improvement of computing power and the diversification of sensors, fusion research has gradually shifted to distributed architectures. By combining local estimation and global fusion, the robustness and scalability of the system are enhanced, which is suitable for large-scale networks and multi-robot systems.

In recent years, deep learning has driven fusion technology towards intelligence. Methods based on CNN and Recurrent Neural Network (RNN) have shown excellent performance in image, speech, and inertial data processing. Graph optimization and factor graph techniques have been extensively adopted in SLAM, significantly enhancing localization accuracy and computational efficiency. Currently, multi-sensor fusion is evolving toward multi-modality, adaptability, and end-to-end learning, thereby offering more robust support for robotic perception and decision-making in complex and dynamic environments.

Contemporary research in multi-sensor fusion algorithms exhibits strong interdisciplinary traits, with deep learning, edge computing, and multi-modal perception emerging as central themes. Recently, neural network-based fusion methods—such as those utilizing CNNs and Recurrent Neural Networks (RNNs)—have achieved notable progress in robot localization and navigation. These approaches effectively integrate data from vision, LiDAR, and inertial measurement units (IMUs), thereby improving perceptual performance in challenging settings. As highlighted in research by Wang, Z. (2022), deep neural networks are capable of efficiently extracting high-level semantic features from multi-source data to facilitate cross-modal fusion [5].

The security robot system introduced by Jiang et al. (2023) incorporates edge computing, leveraging edge intelligence modules for localized data fusion, which substantially improves response speed and system stability [6]. Additional development trends include strengthening algorithmic robustness and optimizing computational efficiency. Conventional methods like KF and PF are increasingly combined with deep learning to form hybrid fusion architectures. For instance, the algorithm developed by Gao and Bai (2023) reduces localization errors by over 15% in dynamic environments [1].

Distributed sensor networks are widely deployed in multi-robot systems, facilitating collaborative decision-making and information sharing across nodes. Research by Zhang, S. L. (2021) demonstrated that distributed fusion-based indoor navigation systems can effectively extend

perceptual coverage [7]. Advances in 5G and IoT technologies have further increased sensor data transmission rates and reduced processing latency, promoting applications in industrial automation and intelligent transportation.

Nevertheless, several challenges remain, including high computational complexity, difficulties in spatiotemporal synchronization and calibration, and limited robustness in dynamic environments. Many existing algorithms are challenging to deploy on resource-constrained devices, while timestamp discrepancies and coordinate misalignment can lead to distorted fusion outcomes.

3. Fusion Approaches and Application Technologies for Robots

3.1. Fusion Approaches

In the field of robotic multi-sensor fusion, the KF remains a fundamental and widely-used technique. Operating under the assumption of a linear Gaussian system, it integrates sensor measurements with model-based predictions through a recursive prediction-update procedure to achieve optimal state estimation in the minimum mean square error sense. In 2023, Gao used Extended KF (EKF) to fuse LiDAR and IMU data, reducing the localization error of mobile robots to less than 0.15 m [1].

In contrast, PF is based on Monte Carlo sampling, which breaks through linear limitations and approximates the posterior probability distribution using weighted particle swarms. Wang (2021) verified its robustness under non-Gaussian noise; when vision fails, the fusion of odometer and ultrasonic data still maintains a localization accuracy of 85% [8].

In addition, neural networks realize end-to-end data fusion with their nonlinear mapping capabilities. Wu (2021) constructed a deep fusion network to fuse visual, infrared, and sonar data, achieving a classification accuracy of 92.7% in service robot scene recognition, which is significantly better than traditional methods [4].

When evaluating models, Root Mean Square Error (RMSE) is a key indicator. Wang, L. B. (2022) achieved a LiDAR-vision fusion RMSE of 0.12 m in the SLAM system [3]. Data consistency relies on timestamp alignment and spatial registration; Qiu (2023) controlled the synchronization deviation within 5 ms through hardware triggering [9]. Optimization strategies such as adaptive noise covariance adjustment are also widely used; for example, Hong (2019) dynamically corrected the process noise parameters of KF in construction machinery, significantly improving control accuracy [2].

3.2. Application Scenarios and Comparison

Multi-sensor fusion technology is widely used in robot systems, covering fields such as industrial automation, service robots, and autonomous driving. Taking DJI ROMO sweeping robot as an example, it adopts the fusion technology of cameras and LiDAR: LiDAR is used to obtain environmental structure information, and cameras are combined to capture color and texture features. The SLAM algorithm is used to construct high-precision maps, realizing centimeter-level mapping and dynamic obstacle recognition, thereby improving cleaning efficiency and safety. The system uses KF to reduce the impact of sensor noise.

In the industrial field, the autonomous navigation assembly robot launched by ABB integrates ultrasonic, visual, and force sensors. By fusing multi-source information, it achieves sub-millimeter-level assembly accuracy. It adopts a parallel fusion strategy and PF algorithm, which improves stability and response speed in dynamic environments, and has been widely used in automobile manufacturing and electronic assembly lines.

The AI navigation system equipped on Ecovacs sweeping robots fuses infrared, gyroscope, and visual data, and uses the fuzzy similarity algorithm to enhance the robustness of path planning. Experiments show that its navigation success rate in complex home environments exceeds 97%, the path deviation is controlled within 5 cm, and it is superior to traditional Bayesian methods in terms

of computational efficiency and energy consumption. It is suitable for embedded platforms with limited resources and provides reliable perception support for indoor service robots.

4. Current Challenges and Future Direction

4.1. Current Challenges

Multi-sensor fusion algorithms face many challenges in complex and dynamic environments. Environmental interference, such as illumination changes, rain, and snow, will affect visual sensors; electromagnetic interference will cause IMU errors; ultrasonic sensors are prone to false detection in multi-reflection environments. All these will increase data uncertainty and affect fusion quality.

The spatiotemporal misalignment between sensors will also reduce fusion effects; temporal misalignment may lead to cumulative errors in the SLAM system, affecting navigation accuracy. Traditional KF performs poorly under nonlinear and non-Gaussian noise and is easily disturbed by outliers; although PF can handle non-Gaussian distributions, its high computational cost makes it difficult to meet real-time requirements.

In recent years, deep learning fusion methods have attracted attention, but they rely on a large amount of labeled data and have poor interpretability. When data is missing or sensors fail, existing fault-tolerant mechanisms are insufficient, which puts higher requirements on the adaptability of algorithms. At the system level, the computing power of embedded platforms is limited, making it difficult to process high-dimensional data in real time, resulting in delays. Multi-source data transmission has high bandwidth requirements and is prone to forming communication bottlenecks; for example, ROS (Robot Operating System) may experience packet loss or delays under multi-sensor concurrency.

In addition, sensor aging and cumulative calibration errors will also affect long-term stability. Therefore, constructing a fusion system with online calibration, dynamic resource scheduling, and anomaly detection capabilities is crucial for improving adaptability and reliability in complex environments.

4.2. Future Directions

Deep learning-multi-sensor fusion integration is key. Feature-level fusion via deep neural networks breaks traditional data correlation limits—e.g., CNN LiDAR-vision fusion achieves centimeter-level localization and reduces KF delay to 0.8ms.

Future research should focus on breakthroughs in core algorithms of multi-sensor fusion, optimize the architecture of deep neural networks, and promote lightweight development. For example, using Transformer to process temporal data of heterogeneous sensors to reduce computational delay, and existing studies have shown its potential. At the same time, an adaptive fusion mechanism should be constructed to dynamically adjust sensor weights according to the environment, and relevant literature provides a theoretical basis for environmental modeling.

At the hardware level, it is necessary to establish standardized interface protocols and realize plug-and-play of multi-source sensors by learning from the ROS 2 framework to improve the compatibility of industrial robots. The security architecture introduces a triple redundancy verification mechanism and constructs a fault diagnosis tree combined with the ISO 13849 standard; cases show that it can reduce the failure probability to below 10⁻⁶.

Multi-sensor fusion will profoundly affect the robot industry ecosystem. Under the background of Industry 4.0, it is expected that the penetration rate of multi-sensor robots in global intelligent manufacturing will reach 35% by 2025. The fusion of millimeter-wave radar and visual sensors can achieve an assembly accuracy of ± 0.05 mm. In terms of medical robots, electronic skin technology is expected to make up for the lack of haptic feedback in the Da Vinci system. In emergency rescue, the SLAM system fusing UWB (Ultra-Wideband) and LiDAR has a localization error of less than 0.3 m in GPS-free environments.

With the deployment of 5G-Advanced, multi-robot collaborative perception has become a trend, requiring a transmission delay of less than 5 ms to realize collision avoidance. The above technical breakthroughs are expected to promote the service robot market to exceed 100 billion US dollars by 2030.

5. Conclusions

By collaboratively processing heterogeneous data, multi-sensor fusion algorithms have significantly improved robots' environmental perception capabilities. This technical framework is based on KF, PF, and deep learning, which effectively improves the localization accuracy and robustness of mobile robots. Breakthroughs have promoted the leapfrog development of robot system performance. Additionally, redundant capabilities to handle sensor failures are conferred upon robots by this technology. The core of multi-sensor fusion resides in the reconstruction of the robot perception paradigm. Not only does it surmount the limitations of a single sensor, but it also boosts the development of robots in the direction of autonomy and intelligence. Current research trends point out that cross-modal feature fusion and end-to-end learning frameworks will break through the perception bottleneck in complex environments to a greater extent, thereby laying a solid technical foundation for embodied intelligent systems.

References

- [1] Gao, Q., Bai, J. N. Research on Mobile Robot Localization Algorithm Based on Multi Sensor Fusion [J]. *Automation and Instrumentation*, 2023, (03):295-300.DOI:10.14016/j.cnki.1001-9227.2023.03.295.
- [2] Hong, L., Li, X., Wang, S. L., Wang, J. C. Research on Intelligent Construction System of Construction Machinery Based on Multi Sensor Fusion [J]. *Modern Manufacturing Engineering*, 2019, (05):132-138.
- [3] Wang, L. B. Research on SLAM algorithm for ground robots based on multi-sensor fusion. Shenyang Institute of Automation, Chinese Academy of Sciences, 2022.
- [4] Wu, L. S. Research and application of intelligent multi-sensor fusion algorithm based on neural network. Institute of Microelectronics, Chinese Academy of Sciences, 2021.
- [5] Wang, Z. Research on vision-tactile perception fusion system for robot grasping task. School of Artificial Intelligence, University of Chinese Academy of Sciences, 2020.
- [6] Jiang, H. Y., Wu, T., Chen, H., Sun, J. H., Ren, J. X., Li, M. S. Design and Application of a Multi Sensor Fusion Security Robot [J]. *Modern Electronic Technology*, 2023, 46 (21): 183-186.
- [7] Zhang, S. L. Research on localization and navigation of indoor mobile robot based on multi-sensor fusion. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, 2021.
- [8] Wang, Z. Research on localization method of mobile robot based on multi-sensor fusion. Shenyang University of Technology, 2021.
- [9] Qiu, H. Research on indoor robot navigation technology based on multi-sensor fusion. Nanjing University of Information Science & Technology, 2023.