Dean&Francis

ISSN 2959-6157

A Review of Multimodal Sensor Technologies and Fusion Methods for Intelligent Robots

Ziyu Xie[1, *]

[1]Faculty of Science and Technology
University: University of Macau
City: Macau Country: China
Postcode: 230000
*Corresponding author: dc42780@um.edu.mo

Abstract:

With the rapid development of artificial intelligence and robotics, intelligent robots are gradually transitioning from structured environments to open and complex scenarios, posing unprecedented requirements for the depth, breadth, and precision of environmental perception. Single-modal sensors can no longer satisfy the requirements of complex tasks, and multimodal sensor fusion technology has become a key approach to enhance the robot's environmental perception, state estimation, and decision-making capabilities. This paper presents a systematic review of multimodal sensing technologies and fusion methodologies for intelligent robots. First, the paper outlines the principles and advances of various core sensors; then it delves into key technologies from signal preprocessing to fusion algorithms spanning classical filtering and deep learning; and it synthesizes their application performance in typical scenarios such as navigation, operation, and human-robot collaboration. Finally, confronting current challenges, the paper envisions the future trends of sensing technologies towards intelligence, flexibility, and chip-based development, aiming to offer insights for researchers in the field.

Keywords: intelligent robot; multimodal sensor; sensor fusion; environmental perception; machine learning

1. Introduction

Robot perception serves as the cornerstone of automation and intelligence, and constitutes the fundamental means for robots to perceive their own states, operational objects, and surrounding environments[1]. From robotic arms in industrial assembly lines to household service robots, and further to exploration robots operating in extreme environments such as the deep sea and space, the intelligence level of these robots largely relies on the performance of their perception systems. In recent years, the advancements in sensor technology, computing hardware, and artificial intelligence algorithms have jointly promoted the leap in robot perception capabilities.

The current domestic and international research landscape indicates that robot perception is evolving from reliance on a single sensor to multimodal collaborative perception[2][3]. Although single vision, laser, or inertial measurement unit (IMU) each have their own advantages, they also have inherent limitations, such as monocular vision's susceptibility to illumination variations, lidar's challenges in perceiving transparent objects, and IMU having cumulative errors. Through the fusion of complementary multi-source sensory data, the robustness, accuracy, and reliability of the perception system can be substantially enhanced. For example, visual-inertial odometry (VIO) integrates rich visual environmental cues from cameras with high-frequency motion state measurements from IMUs and has emerged as a standard technology for mobile robot localization and navigation[4]. This paper seeks to present a systematic and comprehensive review of multimodal sensor technologies and their fusion methods for intelligent robots. First, this paper systematically classifies core sensors and elaborates on their technological advancements; second, it delves into the processing and fusion algorithms of multimodal data; then, it analyzes the practical applications of fusion technologies in typical robotic scenarios; finally, it explores prevailing challenges and envisions future development trends. The innovation of this paper resides in the tight integration of emerging sensing technologies (such as event cameras, electronic skin E-Skin, flexible sensors) and state-of-the-art fusion algorithms (especially deep learning and Transformer models) in recent years, as well as the in-depth discussion of systemic issues including standardization and safety within the field of robotic perception.

2. Classification and Technological Progress of Core Robot Sensors

As the foundational component enabling robots to interact with the physical environment, the performance of the robotic perception system directly dictates the intelligence level and application potential of robots. Based on differences in perceptual modalities and physical working mechanisms, core robotic sensors can be categorized into the following main types, with all related technologies witnessing rapid advancement in recent years.

2.1 Visual Sensors

In terms of visual sensing, robots mainly rely on optical sensors to capture environmental information. Monocular and binocular cameras represent the most basic configurations: monocular cameras are cost-effective and easy to deploy, yet fail to directly obtain depth information; binocular cameras compute parallax based on stereo vision principles to estimate depth, and exhibit high effectiveness in texture-rich scenes[5]. Depth cameras further extend 3D perception capabilities, covering technical routes such as time-of-flight (ToF), structured light and stereo vision. These devices can output high-resolution depth maps in real time and have been extensively applied in navigation and 3D reconstruction[6]. Event cameras, as an emerging dynamic visual sensor, respond to asynchronous brightness changes, have high dynamic range and microsecond-level latency, and are particularly suitable for high-speed motion scenes[7]. In addition, multispectral and hyperspectral imaging technologies have played an important role in vertical fields such as agriculture and environmental monitoring by capturing wide-band spectral information[8].

2.2 Force-Torque and Tactile Sensors

Force-torque and tactile sensing enable robots to achieve physical interaction and fine operation. Six-dimensional force/torque sensors are capable of simultaneously measuring forces and torques in three orthogonal directions, and have been widely adopted in robotic arm tasks such as assembly and grinding[9]. Tactile sensors encompass diverse types, including piezoresistive, capacitive, and piezoelectric sensors. In recent years, the advancement of flexible electronics and stretchable materials has facilitated the advancement of electronic skin (E-Skin) technology, endowing robots with distributed tactile perception capabilities across their surfaces[10]. These sensors not only enhance the operational intelligence of robots but also play a key safety role in human-machine collaboration[11].

2.3 Inertial and Motion Sensors

Inertial and motion sensors mainly include inertial measurement units (IMUs), which consist of gyroscopes, accelerometers, and magnetometers, used to estimate the robot's attitude, angular velocity, and acceleration[3]. Despite the presence of drift issues in IMUs, their high-frequency response characteristics render them indispensable in multi-sensor fusion systems, such as complementing cameras in visual-inertial odometry (VIO) to enhance the

ISSN 2959-6157

robustness and continuity of state estimation[4].

2.4 Distance and Proximity Sensors

Distance and proximity sensors endow robots with the capability to detect surrounding obstacles and nearby objects. Common types encompass ultrasonic, infrared, lidar (LiDAR), and millimeter-wave radar. LiDAR enables the generation of high-precision point cloud maps and serves as a core sensor for autonomous driving and high-end mobile robots[12], while millimeter-wave radar maintains stable operation in adverse weather conditions such as fog and rain[14]. Ultrasonic and infrared sensors are typically employed for near-range obstacle avoidance and object presence detection due to their lower cost.

2.5 Environmental and Biochemical Sensors

Environmental and biochemical sensors broaden the application scope of robots in specific fields. Temperature and humidity, atmospheric pressure, and gas sensors are commonly integrated into environmental monitoring and industrial inspection systems. In medical and service robots, biochemical sensors facilitate the detection of physiological indicators or environmental components, enhancing the practical value of robots in health care and special environmental operations[14]. Emerging technologies such as quantum sensors are increasingly being deployed in high-precision measurement scenarios[15].

3. Sensor Data Processing and Multimodal Fusion

With the growing diversity of sensor types and the escalating complexity of data processing, multimodal fusion technology is developing from traditional filtering methods to end-to-end fusion based on deep learning, exhibiting features of multi-level and cross-modal fusion.

3.1 Low-level Signal Preprocessing and Calibration

In the low-level signal preprocessing and calibration stage, the raw sensor data needs to undergo a series of fine processing to improve the data quality. Owing to the inherent characteristics of sensors and environmental interference, the collected data typically contains Gaussian white noise, quantization error, temperature drift and nonlinear error[2]. The preprocessing stage typically employs digital filtering techniques, including Kalman filtering for system state estimation and measurement noise reduction, median filtering for impulse interference elimination, and Wiener filtering for frequency-domain signal denoising. In addition, the influence of thermal drift needs to be eliminated by temperature compensation algorithm, and nonlinear error needs to be corrected by polynomial fitting or lookup table method to provide an accurate and reliable data foundation for subsequent fusion. The effectiveness of these preprocessing steps directly affects the performance of subsequent high-level fusion algorithms and is an indispensable basic link.

3.2 Sensor Calibration and Time Synchronization

Sensor calibration and time synchronization constitute the fundamental guarantee for multimodal fusion, and determine the consistency and comparability of data from different sensors. The calibration process comprises intrinsic parameter calibration and extrinsic parameter calibration: internal parameter calibration determines the characteristic parameters of the sensor itself[3]; extrinsic parameter calibration determines the relative pose relationship among different sensors, such as hand-eye calibration to determine the transformation matrix between the camera and the robotic arm. Time synchronization addresses the issue of temporal consistency in data acquisition. Hardware synchronization can achieve microsecond-level accuracy through a unified clock signal, while software synchronization relies on timestamp alignment and interpolation methods. In recent years, deep learning-based self-calibration methods have also gradually emerged, which can estimate and correct calibration parameters online and adapt to parameter drift caused by environmental changes[4]. The ability to process asynchronous data directly affects the performance of the fusion system, and dedicated time registration and interpolation algorithms require design to ensure temporal consistency of the data.

3.3 Classical Fusion Algorithms

Classical fusion algorithms provide a solid mathematical foundation and theoretical framework forthe integration of multi-source uncertain information. Kalman filtering (KF) and its extended variants form a comprehensive recursive state estimation theoretical framework: Extended Kalman filtering (EKF) processes nonlinear systems via first-order Taylor expansion, and unscented Kalman filtering (UKF) employs deterministic sampling to more accurately approximate nonlinear distributions[5]. Particle filtering (PF) utilizes the Monte Carlo method to approximate the posterior probability distribution of a state via a set of weighted sample particles, and is particularly suitable for highly nonlinear and non-Gaussian systems. These methods are all based on Bayesian filtering theory, providing a unified paradigm for the fusion of multi-source uncertain information, and have been widely applied and verified in robot localization, navigation and control.

3.4 Deep Learning and Transformer Fusion Framework

Deep learning and Transformer fusion frameworks have greatly promoted the development of multimodal fusion technology in recent years, establishing a novel data-driven fusion paradigm. Deep learning methods automatically learn feature representations and fusion rules via neural networks: Convolutional Neural Networks (CNN) are good at processing gridded data such as images and point clouds, Recurrent Neural Networks (RNN) and its variant LSTM are suitable for modeling time series data, and Graph Neural Networks (GNN) can handle unstructured sensor network data. Transformer models, leveraging their robust self-attention and cross-attention mechanisms, can adaptively capture long-range dependencies and complex inter-modal correlations, demonstrating outstanding performance in heterogeneous sensor fusion. The vision-language-action multimodal large model technology has further elevated the fusion level to the semantic level, enabling robots to achieve high-level understanding and decision-making[8]. These methods have remarkably improved the perceptual capabilities of fusion systems in complex scenarios, but have also imposed higher requirements on computing resources and labeled data.

3.5 Real-time Performance, Robustness and Security Considerations

Real-time performance, robustness, and security are core engineering considerations in the practical deployment of multimodal fusion systems. Real-time performance necessitates a delicate balance between algorithm complexity and computing resources, and requires the adoption of various technical approaches such as model simplification, computation acceleration and hardware optimization to satisfy the online operation requirements of robotic systems[16]. Robustness refers to the system's ability to maintain performance under sensor failure, data loss or extreme environmental interference, and needs to be guaranteed through redundant design, fault detection and isolation (FDI) mechanism and adaptive fusion strategy. Security involves data privacy protection, system anti-adversarial capabilities, and functional safety, which is especially important in collaborative robots and unmanned systems[17], and requires the establishment of a comprehensive security protection system from hardware root of trust to software protection mechanisms. Furthermore, the interpretability of fusion systems needs to be emphasized to ensure a transparent and credible decision-making process, which is critical for accountability and system certification in key application scenarios.

4. Challenges and Future Trends

Despite significant progress in multimodal sensor technology and fusion methods, many challenges remain, indicating future directions for technological development. Sensor miniaturization, low power consumption, and self-powering capabilities are currently the primary challenges. With the increasing portability and miniaturization of robots, stringent requirements are imposed on the size and weight of sensors. Power consumption limitations directly impact system endurance. Future trends include the utilization of MEMS processes for miniature sensor fabrication, the development of low-power circuits, and the advancement of energy harvesting technologies. Self-powering technologies, utilizing environmental vibrations, thermal gradients, or electromagnetic waves to power sensors, are becoming a research hotspot. The integration of flexible wearables with soft robots represents a novel form of sensor technology. Flexible electronics technology enables the fabrication of stretchable and bendable sensors, which can be perfectly integrated with soft robots to achieve distributed tactile and strain sensing. In the future, flexible sensors with self-healing capabilities, good biocompatibility, and the ability to be fabricated on a large scale can be developed. Heterogeneous multimodal hyper-fusion constitutes a key approach to enhancing sensing capabilities. Sensors of different modalities exhibit significant differences in data types, spatiotemporal characteristics, and information content, rendering deep information fusion a major challenge. Deep learning models based on attention mechanisms can automatically learn the importance weights of different modalities, achieving adaptive fusion. In the future, the direct deployment of

ISSN 2959-6157

AI algorithms at or near the sensor edge can significantly reduce latency and data transmission bandwidth, while enhancing privacy protection.

Future trends also include the development of privacy-preserving technologies such as federated learning and differential privacy, alongside the establishment of comprehensive robotic ethics guidelines and regulatory frameworks. Standardization, repeatability, and open datasets are crucial foundations for the healthy development of the field. Presently, there is a lack of unified sensor interface standards, data formats, and evaluation benchmarks. Future development necessitates the establishment of industry-wide standards, as well as the promotion of open-source tool development and shared dataset construction. Sustainability and green manufacturing are environmental factors that must be considered for the long-term development of sensor technology.

5. Conclusion

This review illustrates that multimodal sensors and their fusion technologies serve as the core drivers for the development of intelligent robotic perception capabilities. The technology review indicates that the diversified development of sensing technologies, ranging from traditional vision, force, and inertial sensors to emerging event cameras, electronic skin, and flexible sensors, endows robots with rich and complementary environmental perception capabilities. Multimodal information fusion technologies—encompassing low-level fusion based on classical filtering algorithms and high-level semantic fusion driven by deep learning and Transformer frameworks—effectively mitigate the limitations of single sensors, thereby significantly improving the accuracy, robustness, and environmental adaptability of perception systems. From traditional vision and force sensors to emerging flexible sensors, the technological diversity, coupled with fusion algorithms based on filtering and deep learning, collectively overcomes the limitations of single sensors, thereby significantly improving the perception robustness and task execution capabilities of robots in industrial, service, and extreme environments. Looking forward, future research should focus on key directions including device miniaturization, cross-modal algorithm understanding, system standardization, and safety ethics to propel robot perception to new heights.

References

[1] CHU Z, FU L, HUANG D, et al. Rigid-flexible synergy dual-gradient microwrinkled strain sensor for machine learning-assisted wearable technology[J]. Chemical Engineering Journal, 2025, 515: 163818.

[2] Hu Funing, Wang Xiaoyan. Sensor Technology and Application [M]. Beijing: People's Posts and Telecommunications Publishing House, 2024.

[3] Wang Liyong, Jia Ran, Chen Tao. Principle and application of modern sensors [M]. Beijing: Chemical Industry Press, 2023.

[4] Cheng Yangrui, Wang Bingkun, Wang Xiyuan, et al. Progress and prospect of key technologies of underwater robot synchronous positioning and mapping [J]. Journal of Ocean Technology, 2025,44 (3): 97–111.

[5] Pan Yundan, Lin Yixin. Design of teaching mobile robot with multi-sensor information fusion [J]. Internet of Things Technology, 2024,14 (7): 99–103+106.

[6] Huang Xiaojun. Application of sensor technology in robot system [J]. Application of integrated circuits, 2024,41 (7): 335-337.

[7] Miao Jingliang, Wang Guoqiang. Development course and future prospect of robot sensors [J]. Zhangjiang Science and Technology Review, 2024 (11): 4-7.

[8] BERA D, RAUT S, MAHATA C, et al. Early prediction of forest fire events integrating multi-sensor datasets and machine learning techniques[J]. Stochastic Environmental Research and Risk Assessment, 2025 (prepublish): 1–20.

[9] XU J, SONG A, CHEN Y, et al. Research on a six-axis force/torque sensor based on thin film sputtering technology and its decoupling for space robot[J]. Measurement, 2025, 256: 117978.

[10] YADAV N, ZHANG Q, QI D, et al. AI-integrated wearable strain sensors: advances in e-skin, robotics, and personalized health monitoring[J]. Nanoscale advances, 2025.

[11] JAMADARKHANI M, RAPHAEL R, RAMPRASAD P H S, et al. IoT enabled smart water metering using multi sensor data and machine learning techniques[J]. Frontiers in Water, 2025, 7: 1586916.

[12] Lv Qiuhui, Zhu Lixue, Zhang Shiang, et al. Application status and trend of sensor technology in fruit picking robots [J]. Journal of Intelligent Agricultural Equipment (in Chinese and English), 2025,6 (2): 58–68.

[13] DOLUI C, ROY D. Advancement of bridge health monitoring using magnetostrictive sensor with machine learning techniques[J]. Nondestructive Testing and Evaluation, 2025, 40(9): 4219–4240.

[14] NASSABPOUR M, SHUQAIR M, ROSENFELD A T P

D, et al. Integrating Wearable Sensor Technology and Machine Learning for Objective m-CTSIB Balance Score Estimation[C]// Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2024: 1–4.
[15] Jiang Guanqun, Zhao Yi, Xu Zizhao, et al. Application of quantum sensor technology in navigation robot system for percutaneous puncture surgery [J]. China Medical Device Information, 2025,31 (13): 51–53+63.
[16] Xu Lan, Shang Li, Xu Zixin, et al. Application and development trend of robots in manufacturing industry [J]. Science and Technology Information, 2023,21 (24): 44-48.
[17] JAHANSHAHI H, ZHU H Z. A comprehensive review of tactile sensing technologies in space robotics[J]. Chinese Journal of Aeronautics, 2025, 38(7): 103423.