Proceedings of CONF-MLA 2025 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2025.LD28245

Multi-Sensor Fusion and Collaborative Perception for Autonomous Robots in Complex Maze Environments

Shangze Kong

Department of Mechanical Engineering, University College London, London, The United Kingdom
shangze.kong.24@ucl.ac.uk

Abstract. Complex mazes are characterised by narrow passages, frequent obstructions, and mirrored or transparent boundaries. This paper reviews multi-sensor fusion and collaborative perception technologies for autonomous mobile robots. By comparing lidar, depth cameras, inertial measurement units (IMUs), wheel speed sensors, ultrasonic sensors, and infrared sensors, this paper highlights the complementary advantages of each sensor. It defines sensor pairing modes and deployment scenarios. A four-layer framework is adopted: data layer, feature layer, decision layer, and hybrid layer. The data layer fuses information at the pixel or point echo level to maximise information. The feature layer balances accuracy and latency. The decision layer adds fault-tolerant mechanisms. Hybrid or adaptive scheduling switches between layers based on the scenario and computational budget. All fusion algorithms in this paper are based on Bayesian inference. Kalman filter-type algorithms (KF/EKF/UKF/MSCKF/ESKF) achieve tightly coupled LIO/VIO. Particle filter-type algorithms (PF/AMCL/RBPF) perform global positioning. The deep learning fusion algorithm BEV achieves a unified cross-view and cross-modal representation. Under bandwidth and latency constraints, information sharing, map stitching, collaborative path planning, and task allocation among multiple robots achieve virtually wider-angle vision and significantly enhanced coverage capabilities. Overall, multi-sensor collaborative perception substantially improves the robustness and efficiency of maze exploration, though main challenges such as synchronisation, calibration, and domain adaptation still need to be addressed.

Keywords: Multi-Sensor Fusion, Collaborative Perception, Maze Exploration Robots, Autonomous Navigation.

1. Introduction

Against the backdrop of rapid advancements in artificial intelligence, sensor technology, and automated control systems, autonomous mobile robots are continually being enhanced in their ability to navigate complex environments. As a significant application in this field, maze-exploring robots are widely used for tasks including logistics transport, post-disaster search and rescue, underground pipeline inspection, and medical navigation assistance. These tasks impose high demands on a robot's capacity for autonomous environmental perception, path planning and obstacle avoidance within unfamiliar environments.

DOI: 10.54254/2755-2721/2025.LD28245

However, real-world maze environments typically have complex structures, dense obstacles, and incomplete information, posing significant challenges to a robot's environmental perception and cognition. Depending only on a single sensor (such as infrared, ultrasonic, or visual sensors) often results in a limited field of view, insufficient recognition accuracy, and poor environmental adaptability [1]. Furthermore, positioning methods based on odometry and inertial measurement units (IMUs) are prone to cumulative errors, so they struggle to achieve continuous and reliable navigation [2].

To overcome these challenges, numerous multi-sensor fusion and collaborative perception solutions have been proposed. Systems integrate diverse sensors, including lidar, cameras, IMUs, and encoders. Fusion algorithms such as Kalman filtering and deep learning models are employed [3,4]. Information from multiple sources complements each other, enhancing perception accuracy. The system's robustness and fault tolerance are enhanced. Furthermore, within multi-robot systems, collaborative perception mechanisms enable environmental information sharing, map stitching, and task coordination, significantly boosting overall exploration efficiency and intelligence levels [5].

In recent years, this field has undergone rapid evolution from low-level data fusion to high-level intelligent perception. Fusion algorithms have evolved from early heuristic rules to probabilistic graphical models, and now to current deep fusion and multi-robot distributed SLAM systems, with the technical framework becoming increasingly mature [6].

This paper introduces the concepts and system architecture of multi-sensor fusion and collaborative perception technologies, specifically targeting maze-exploring robots. It reviews the key technological pathways and research progress in this field from aspects such as perception system configuration, multi-source information fusion methods, multi-robot collaborative mechanisms, typical application cases, and future challenges. It also proposes the development trends and research prospects of multi-sensor collaborative perception in autonomous exploration of complex environments.

2. The perception system for maze exploration robots

The maze exploration task requires robots to perform obstacle avoidance stably, positioning and mapping, and target guidance in unknown environments. To balance real-time performance and robustness, the system usually uses multiple sensors working together.

Standard sensors and their functions: LiDAR (high-precision geometric modelling and obstacle recognition), depth cameras (high-precision depth perception), inertial measurement units (IMUs) (high-precision attitude measurement and short-range odometry), encoders (odometry constraints), and ultrasonic or infrared sensors (near-range blind spot compensation). These sensors are complementary in field of view, scale, and environmental adaptability, yet each possesses limitations. For instance, lidar is easily affected by glass and dust; visual systems suffer from illumination and occlusion interference; inertial measurement units exhibit drift issues; and ultrasonic sensors demonstrate low directional resolution [7].

Therefore, a single perception modality cannot sufficiently cover complex scenarios within the maze, such as frequent turns and narrow passages. Multi-sensor fusion enhances positioning accuracy and system robustness through spatio-temporal synchronisation, redundancy verification, and feature complementarity [8]. Typical combinations include LiDAR + IMU (Geometric Inertial Coupled Positioning), visual ray tracing = camera + IMU (Lightweight Navigation), and RGB depth sensor + IMU (Balancing Semantic and Kinematic Constraints) [9].

Maze exploration requires robots to stably complete tasks such as obstacle avoidance, positioning, mapping, and target guidance in unknown environments. To balance real-time

DOI: 10.54254/2755-2721/2025.LD28245

performance and robustness, the system typically uses multiple sensors working together.

3. Multi-sensor fusion methods

This chapter mainly explores the fusion mechanisms within single-body maze exploration robots. It first proposes a hierarchical fusion framework, followed by a systematic exposition of commonly used fusion algorithms. This includes their classification, core formulas, recent research developments, and respective advantages and disadvantages.

3.1. Integration level classification

3.1.1. Data layer fusion (early fusion)

The core idea of data layer fusion is to align and jointly utilise multi-source raw data at the lowest level [10]. This process contains three steps. Firstly, spatio-temporal alignment is performed; secondly, through external parameter calibration and coordinate unification, observation data from different sensors are projected onto a standard reference frame; finally, joint coding or resampling is performed at the pixel, echo, or point level, and the results are fed into downstream modules. This method maximises the retention of complementary information [11]. On complex road sections such as sharp bends, areas of heavy occlusion, or surfaces with mirror or glass reflections, this method effectively achieves channel identification, obstacle boundary extraction, and geometric detail reconstruction. To accomplish that, higher engineering demands are needed. The system requires greater synchronisation accuracy, more stable external parameter values, and enhanced bandwidth, storage, and computational capabilities. Any synchronisation deviation will cause subsequent modules to generate cumulative errors. In practical applications, the combined approach of 'radar-vision candidate region generation + image refinement (RRPN)' is widely adopted [12]. This strategy integrates three-dimensional candidate regions generated from the bird's-eye view of LiDAR with RGB image features, enhancing detection accuracy.

3.1.2. Feature layer fusion (mid fusion)

Feature layer fusion emphasises 'aligning effective information'. Each modality first completes feature encoding within its own channel (e.g., semantic information for images, geometric data for lidar, short-term trajectory information for inertial measurement units), followed by complementary information exchange to ultimately generate richer feature representations [10]. This approach typically achieves a better balance between accuracy and real-time performance, while exhibiting greater robustness to minor external parameter drifts and non-strict synchronisation. Consequently, it often serves as the primary path for maze exploration. It is particularly noteworthy that alignment quality determines the upper limit of fusion effectiveness: projection errors or temporal misalignments may lead to 'false consistency'. In practice, the MV3D [13] method significantly enhances detection accuracy by fusing three-dimensional candidate regions generated from the bird's-eye view of lidar with RGB image features.

3.1.3. Decision-layer fusion(late fusion)

Decision-layer integration emphasises the independence and fault isolation of processing chains. Each sensor or subsystem first independently completes a series of tasks, including detection, tracking, and segmentation. Ultimately, confidence normalisation, conflict detection, and robustness

DOI: 10.54254/2755-2721/2025.LD28245

verification occur at the result layer [10]. This approach supports incremental integration and degraded operation. Even if one chain degrades, the remaining results can sustain service. However, it relies more heavily on upstream data quality and may lose underlying detail. Overall, the decision layer excels in fault fallback and fault tolerance, serving as a cornerstone for system stability.

3.1.4. Hybrid-layer fusion

In practical robotic systems, particularly when tackling complex maze environments characterised by high dynamics and uncertainty, single-level fusion often fails to address all perception requirements. Consequently, an increasing number of studies are exploring hybrid fusion strategies based on task complexity. These approaches achieve greater system robustness and perception accuracy by flexibly integrating sensory information across multiple levels, from raw data processing to final decision output. By dynamically scheduling different fusion methods, this strategy demonstrates superior adaptability and performance in complex and dynamic environments, establishing itself as a significant development direction within the field of multi-sensor fusion [10].

3.2. Common fusion algorithms

Uniform symbols and objectives: state x_k , control input u_k , and observation z_k ; the process noise $w_k \sim \mathcal{N}(0, Q_k)$ and the measurement noise $v_k \sim \mathcal{N}(0, R_k)$. All fusion algorithms in this paper are based on Bayesian inference. Kalman filter-type algorithms (KF/EKF/UKF/MSCKF/ESKF) implement tightly coupled linear input-output (LIO) or visual input-output (VIO) fusion. Particle filter-type algorithms (PF/AMCL/RBPF) perform global localisation. The deep learning fusion algorithm BEV achieves a unified cross-view and cross-modal representation.

3.2.1. Bayesian estimation

The Bayesian paradigm employs a posteriori-driven fusion, unifying motion priors with multi-modal observations within a “prediction-update” framework. It adapts weighting according to scene quality, ensuring consistent probabilistic semantics and interpretability across different solvers.

$$\boldsymbol{p}\left(\boldsymbol{x_k z}_{1:\boldsymbol{k}-1}\right)=\int \boldsymbol{p}\left(\boldsymbol{x_k x}_{\boldsymbol{k}-1}\right)\boldsymbol{p}\left(\boldsymbol{x}_{\boldsymbol{k}-1}\boldsymbol{z}_{1:\boldsymbol{k}-1}\right)\boldsymbol{dx}_{\boldsymbol{k}-1} \tag{1}$$

$$\boldsymbol{p}\left(\boldsymbol{x_k z}_{1:\boldsymbol{k}}\right)=\frac{\boldsymbol{p}(\boldsymbol{z_k x_k})\,\boldsymbol{p}(\boldsymbol{x_k z}_{1:\boldsymbol{k}-1})}{\int \boldsymbol{p}(\boldsymbol{z_k x_k})\,\boldsymbol{p}(\boldsymbol{x_k z}_{1:\boldsymbol{k}-1})\,\boldsymbol{dx_k}} \tag{2}$$

From the above equations, in recursive Bayesian estimation, Eq. (3) is the prediction step: it propagates the previous posterior forward via the state-transition probability $\boldsymbol{p}\left(\boldsymbol{x_k x}_{\boldsymbol{k}-1}\right)$ under the first-order Markov assumption and marginalizes the unobserved $\boldsymbol{x}_{\boldsymbol{k}-1}$ by integration, yielding the prior $\boldsymbol{p}\left(\boldsymbol{x_k z}_{1:\boldsymbol{k}-1}\right)$ conditioned on past measurements. $\boldsymbol{p}\left(\boldsymbol{x_k x}_{\boldsymbol{k}-1}\right)$ is specified by the system dynamics and the process-noise covariance $\boldsymbol{Q_k}$, describing the evolution of the state from $\boldsymbol{k}-1$ to $\boldsymbol{k}$; $\boldsymbol{p}\left(\boldsymbol{x}_{\boldsymbol{k}-1}\boldsymbol{z}_{1:\boldsymbol{k}-1}\right)$ aggregates all information available up to the previous time step.Eq. (4) is the update step: applying Bayes’ rule fuses the new measurement $\boldsymbol{z_k}$ with the prior to obtain the current posterior $\boldsymbol{p}\left(\boldsymbol{x_k z}_{1:\boldsymbol{k}}\right)$, where $\boldsymbol{p}\left(\boldsymbol{z_k x_k}\right)$ is the measurement likelihood and the integral in the denominator is a normalizing constant ensuring the distribution integrates to one [14].

DOI: 10.54254/2755-2721/2025.LD28245

This method demonstrates excellent performance in semantic consistency and scalability, integrating smoothly with Kalman filtering, particle filtering, and graph optimisation techniques. Its primary shortcomings include sensitivity to model assumptions, susceptibility to inconsistencies under model mismatch, and significant reliance on prior knowledge. In maze-based applications, data-driven noise calibration and quality gating mechanisms can mitigate estimation biases. By consolidating noise's time-varying characteristics through replay calibration and cross-validation, these measures enable stable updates across different degraded segments while preserving result interpretability.

3.2.2. Kalman-Filter (KF / EKF / UKF / MSCKF / ESKF)

This type of algorithm provides online optimal estimation under Gaussian distribution and linear or quasi-linear assumptions, whilst ensuring numerical stability and robustness through error state and gating mechanisms [7].

System model

$$x_{k+1} = A_k x_k + B_k u_k + w_k \tag{3}$$

$$z_k = H_k x_k + v_k \tag{4}$$

Prediction steps

$$\widehat{x_k^-} = A_{k-1}\widehat{x_{k-1}} + B_{k-1}u_{k-1} \tag{5}$$

$$P_k^- = A_{k-1}P_{k-1}A_{k-1}^T + Q_{k-1} \tag{6}$$

Update Procedure

$$K_k = P_k^- H_k^T \left(H_k P_k^- H_k^T + R_k\right)^{-1} \tag{7}$$

$$\widehat{x_k} = \widehat{x_k^-} + K_k\left(z_k - H_k\widehat{x_k^-}\right) \tag{8}$$

$$P_k = (I - K_k H_k)P_k^- \tag{9}$$

In the discrete linear–Gaussian state-space setting adopted in this work, Eqs. (1)– (2) specify the state transition and measurement models, where x_k denotes the state, u_k the control input, z_k the measurement, and A_k , H_k , B_k are known model matrices. At each time step, the filter first performs the time update: from Eqs. (3)– (4) it obtains the prior estimate x_k^- and its uncertainty P_k^- , which are predictions based solely on the model. The subsequent measurement update computes the Kalman gain K_k via Eq. (5) to weight the innovation (residual) $r_k = z_k - H_k\widehat{x_k^-}$; the information is fused with the prior using Eq. (6) to produce the posterior estimate x_k^- , and Eq. (7) updates the posterior covariance P_k to reflect the reduction in uncertainty. The gain K_k adaptively balances “trust in the model prediction” against “trust in the measurement”: when the measurement noise R_k is small or the prior uncertainty P_k^- is large, the update relies more on the measurement, and vice versa. The choices of Q_k and R_k directly determine the convergence rate

DOI: 10.54254/2755-2721/2025.LD28245

and steady-state accuracy. Under this iterative predict–correct scheme, the Kalman filter yields a minimum mean-squared error (MMSE) estimate of the system state under the linear–Gaussian assumptions [15].

In recent studies, LIO-SAM achieves the elimination of point cloud skew and enhanced real-time performance by constructing a laser radar inertial odometer and pre-integrating the IMU on a factor graph [16]. LVI-SAM employs a factor graph to tightly couple visual-inertial (VIS) and laser-inertial (LIS) systems. VIS initialisation is accomplished using LIS estimates; LIS employs VIS estimates as initial values to support scan matching, with loops first detected by VIS and subsequently refined by LIS. Operation persists even if either subsystem fails, thereby enhancing robustness in textureless and featureless environments [17].

This approach offers low latency, interpretability, and seamless integration with backend optimisation. Its limitations include sensitivity to time synchronisation and external participation noise modelling, alongside the accumulation of linearisation drift. In labyrinthine environments, rolling shutter compensation is employed alongside hot-swapping of observation weights, with loop factor correction applied to linearisation errors. Attitude and pose updates are unified under an error state formulation. This approach maintains stable, continuous positioning even under time-varying degradation.

3.2.3. Particle Filtering (PF/AMCL/RBPF)

PF approximates the posterior distribution using a sample set, thereby directly handling non-linear, non-Gaussian and multi-modal distributions. It excels at global positioning, abduction recovery and re-localisation, and complements continuous filtering [7].

Particle filtering (PF), as a sequential Monte Carlo (SMC) method, fundamentally employs importance sampling and discrete stochastic measures to approximate the posterior distribution of states recursively. Initially applied primarily in polymer growth, this methodology later expanded into physics and engineering domains. Its application was historically constrained by computational complexity and processing power. Yet, it has seen a rapid resurgence in recent years due to advances in hardware and parallel computing, alongside its potential in signal processing. Key engineering challenges lie in weight decay and insufficient sample diversity. When the effective number of particles (ESS) diminishes, a few high-weight particles dominate the posterior distribution, leading to unstable estimation. Common countermeasures include (adaptive) resampling to suppress weight concentration, improved proposal distributions (coupling kinematic priors with observed likelihoods to form locally optimal proposals), low-variance resampling, and particle 'tempering' (random perturbations or MCMC moves) to maintain exploration capabilities [18].

In recent research, KLD-Sampling achieves an automatic trade-off between accuracy and real-time performance by adaptively scaling particle size based on KL error upper bounds. Indoor global positioning using semantic maps combined with PF accelerates re-localisation through semantic anchor-enhanced disambiguation [19]. Differentiable Active PF synchronises policy learning with particle emission to proactively gather information while reducing exploration costs, enabling PF to demonstrate superior convergence and recovery capabilities in complex topologies and large-scale environments.

Concurrently, this approach excels in maintaining self-localisation under high uncertainty and multimodal conditions whilst tolerating sensor failures. Its limitations lie in complexity scaling linearly with particle count, and likelihood mismatches impeding convergence. Consequently, for maze applications, it integrates hierarchical particle scattering, semantic priors, and adaptive particle

DOI: 10.54254/2755-2721/2025.LD28245

count control to manage computational expenditure. Furthermore, PF serves as an emergency bypass to the primary odometer, enabling rapid trajectory recovery during misalignment.

3.2.4. Deep Learning–based Fusion (BEV)

When employing bird's-eye view (BEV) as the unified representation, cross-view integration can be accomplished through the pipeline: calibration → rectification → alignment → stitching → fusion. Multiple fisheye cameras first utilise FOV distortion models and LM optimisation to estimate intrinsic parameters and perform distortion correction. Subsequently, all views are unified to a top-down coordinate system via a ground plane homothetically consistent HHH transformation. Sewing point selection relies on a quadratic error field derived from calibration residuals, employing greedy or dynamic programming to choose the lowest-residual seam within overlapping regions. Pixel sequences are then registered at seams using dynamic image time regularisation (DIW). Subsequently, Wendland's tight-support RBF smoothly propagates seam registration deformations across the entire image. Finally, global exposure is normalised using gain or bias adjustment. Weighted fusion is applied using a 'calibration residual × distance from boundary' weighting scheme to eliminate visible seams and brightness inconsistencies. Additionally, a history term is incorporated into DIW to suppress inter-frame jitter, ensuring stable BEV output under occlusion and minor pose fluctuations [20].

In recent research, BEVFusion unifies cameras and LiDAR within a shared BEV while significantly improving cross-modal pooling efficiency [21]. TransFusion mitigates pixel-point cloud mismatch through decoder soft alignment, demonstrating greater robustness under low-light conditions [22]. BEVDet4D introduced temporal BEVs to substantially reduce velocity estimation errors, while RCBEVDet integrated radar-camera data into BEVs to enhance penetration and velocity observation. FusionLoc employed a camera + 2D LiDAR with multi-head attention for end-to-end pose regression indoors, markedly improving robustness in maze corners, occlusion, and smoke-filled segments.

Concurrently, this approach excels in robust high-dimensional semantic understanding, streamlined end-to-end optimisation, and seamless multi-task coordination. Its limitations lie in substantial data and computational demands, coupled with risks of domain mismatch. Consequently, within maze applications, it establishes synthetic-to-real domain adaptation and testing alignment, coupled with robust kernel or loopback interaction at the backend. This achieves steady-state operation, balancing high-performance ceilings with deployability.

4. Cooperative perception and multi-robot collaboration

4.1. Concept of collaborative perception: information sharing and blind-spot compensation

In highly occluded, structurally repetitive and time-varying environments such as mazes, single robots are often constrained by field of view limitations, measurement degradation and computational budgets. Consequently, the core objective of collaborative perception is to share multiple robots' local observations, estimates and uncertainties within a unified spatio-temporal framework, thereby forming a broader 'virtual field of view' and more stable global cognition. Specifically, multiple robots can exchange or occupy semantic slices, keyframes, cross-robot feedback loops, relative poses and covariances, alongside observation quality labels (e.g., brightness, echo intensity, point cloud density). Thus, blind spots such as obstructed corners, low-light areas, dust-filled passages, and glass surfaces can be compensated for through the perspectives

DOI: 10.54254/2755-2721/2025.LD28245

of others. To prevent bandwidth and latency from becoming bottlenecks, information is typically organised hierarchically: raw data is only directly connected for short durations in local emergency scenarios, while more frequently, summarised exchanges occur at the feature, map or decision layers. These exchanges uniformly carry timestamps, coordinate systems, and external parameter version numbers to facilitate backend alignment. Through this approach of 'task-oriented sharing coupled with quality-oriented weighting,' the system achieves robustness and exploration efficiency surpassing individual units' limitations without significantly increasing computational or communication burdens.

4.2. Multi-robot perception architectures

Multi-robot "swarm architectures" provide the infrastructure for collective behaviour, determining a system's capabilities and limitations. Key characteristics include centralised or decentralised (decentralisation further subdivided into hierarchical and fully distributed), role differentiation (homogeneous or heterogeneous), communication, and modelling capabilities for others. Decentralisation is frequently touted for its fault tolerance, parallelism, and scalability, yet direct theoretical or empirical comparisons remain scarce. In practice, numerous systems adopt hybrid approaches: broadly decentralised while incorporating a 'leader' or central planner for high-level coordination. Heterogeneity increases task allocation complexity and challenges in modelling teammates; 'task coverage' can be used to measure an individual's capacity for independent task completion, with lower coverage indicating greater reliance on collaboration.

4.3. Collaborative path planning and task allocation strategies

Multi-robot path planning fundamentally constitutes a 'resource contention problem within finite spaces,' necessitating the coordination of multiple bodies' movements without intersection. Classical reviews categorise approaches into centralised (where a unified planner coordinates all robots) and distributed or online (where each entity plans independently and adjusts during operation), with hybrid variants blending centralised-decentralised or online-offline methodologies also existing. An alternative equivalent classification distinguishes between centralised approaches (considering all robots simultaneously) and decoupled approaches: the latter either plan sequentially based on global priorities (each robot avoiding only higher-priority entities) or treat 'path-time' as a schedulable resource for path coordination (i.e., sequencing conflicts within the configuration space-time). The literature also presents a distributed approach: agents initially attempt straight-line travel, switching to visible vertices upon encountering obstacles, and resolving conflicts through dynamic prioritisation and local negotiation or blackboard mechanisms. As 'pre-calculating all paths' is often impractical in real systems, implementation frequently devolves to designated routes + rules (modelled on traffic laws) to prevent collisions and deadlocks. Verified approaches encompass rules like 'keep right, stop at junctions or maintain distance', priority-based conflict resolution, mutual exclusion protocols controlling passage numbers, and distributed algorithms addressing multi-junction and deadlock detection issues [23].

5. Conclusion

This paper unfolds around the closed-loop framework of 'perception-estimation-coordination-planning': given that maze environments feature both high occlusion and multi-class degradation, a single sensor is prone to limitations imposed by field-of-view and noise models. Therefore, robust

DOI: 10.54254/2755-2721/2025.LD28245

prior knowledge must be established through a clearly defined multimodal combination. Given the inherent difficulty in maintaining long-term temporal and spatial consistency across modalities, hierarchical fusion serves as the organising principle: early fusion preserves detail when raw data permits high-precision alignment; feature-level fusion achieves optimal accuracy-latency trade-offs under alignment errors and computational constraints; result-level fusion isolates failures when service continuity is paramount. Adaptive scheduling enables seamless inter-layer switching. Algorithmically, error state filtering and factor graph smoothing support high-frequency, tightly coupled LIO or VIO. Particle filtering performs global localisation and kinematic recovery under nonlinear and multi-modal posterior distributions. At the same time, BEV unifies geometry and semantics to maintain stable output around corners, in low light, and during occlusions. Concurrently, to expand the 'effective field of view' and mitigate individual degradation, multi-robot coordination requires sharing quality-annotated features, maps or loopback cues. At the same time, path planning and task allocation mutually constrain improvements in coverage efficiency and safety. Furthermore, considering that long-term synchronisation and external parameter thermal drift may cause inconsistencies, asynchronous sampling and OOSM may violate filtering assumptions, and BEV may experience mismatches in glass or smoke or high-reflectivity scenarios, embedded platforms face computational or energy constraints with limited communication bandwidth, and dense traffic demands stringent safety and deadlock avoidance. The system must therefore concurrently implement online Q/R learning and external parameter self-calibration to maintain probabilistic consistency, introduce utility-latency-energy-driven hybrid scheduling for controlled degradation, and combine radar or lightweight laser and adaptive testing to mitigate domain drift. Conservative information fusion and lightweight loopback cues maintain global consistency, while edge-cloud event-driven mechanisms absorb OOSM. Furthermore, uncertainty-driven active exploration, viewpoint selection, and market-based allocation decouple perception from planning. This is complemented by constraints such as "keep right, stop at intersections or mutually exclusive passage" to reduce collisions and congestion. This enables long-term autonomous operation within real-world complex mazes, balancing accuracy, robustness, and deployability.

References

[1] Brenner, M., Reyes, N. H., Susnjak, T., Barczak, A. L. C. (2023) RGB–D and thermal sensor fusion: A systematic literature review. arXiv preprint arXiv: 2305.11427.
[2] Buchanan, R., Agrawal, V., Camurri, M., Dellaert, F., Fallon, M. (2022) Deep IMU bias inference for robust visual–inertial odometry with factor graphs. arXiv preprint arXiv: 2211.04517.
[3] Fan, Z., Zhang, L., Wang, X., Shen, Y., Deng, F. (2025) LiDAR, IMU, and camera fusion for simultaneous localization and mapping: A systematic review. Artificial Intelligence Review, 58, 174.
[4] Li, C., Wang, S., Zhuang, Y., Yan, F. (2021) Deep sensor fusion between 2D laser scanner and IMU for mobile robot localization. IEEE Sensors Journal, 21(6), 8501–8509.
[5] Lajoie, P. Y., Ramtoula, B., Wu, F., Beltrame, G. (2022) Towards collaborative simultaneous localization and mapping: A survey of the current research landscape. arXiv preprint arXiv: 2108.08325.
[6] Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I., Leonard, J. J. (2016) Past, present, and future of simultaneous localisation and mapping: Toward the robust–perception age. IEEE Transactions on Robotics, 32(6), 1309–1332.
[7] Liu, Y., Wang, S., Xie, Y., Xiong, T., Wu, M. (2024) A review of sensing technologies for indoor autonomous mobile robots. Sensors, 24(4), 1222.
[8] Tran, Q. K., Ryoo, Y. J. (2025) Multi-sensor fusion framework for reliable localization and trajectory tracking of mobile robot by integrating UWB, odometry, and AHRS. Biomimetics, 10(7), 478.
[9] Wei, C., Qin, Z., Zhang, Z., Wu, G., Barth, M. J. (2025) Integrating multi-modal sensors: A review of fusion techniques for intelligent vehicles. arXiv preprint arXiv: 2506.21885.

DOI: 10.54254/2755-2721/2025.LD28245

[10] Gadzicki, K., Khamsehshari, R., Zetzsche, C. (2020) Early vs late fusion in multimodal convolutional neural networks. In Proceedings of the IEEE International Conference on Information Fusion (FUSION), pp. 1–6. IEEE.
[11] Nabati, R., Qi, H. (2019) RRPN: Radar region proposal network for object detection in autonomous vehicles. In Proceedings of the IEEE International Conference on Image Processing (ICIP), pp. 3093–3097. IEEE.
[12] Chen, X., Ma, H., Wan, J., Li, B., Xia, T. (2017) Multi-view 3D object detection network for autonomous driving. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6526–6534. IEEE.
[13] Kim, S., Petrunin, I., Shin, H. S. (2025) A review of Bayes filters with machine learning techniques and their applications. Information Fusion, 114, 102707.
[14] Welch, G., Bishop, G. (1995) An introduction to the Kalman filter. University of North Carolina at Chapel Hill, Department of Computer Science.
[15] Shan, T., Englot, B., Meyers, D., Wang, W., Ratti, C., Rus, D. (2020) LIO-SAM: Tightly-coupled Lidar inertial odometry via smoothing and mapping. arXiv preprint arXiv: 2007.00258.
[16] Shan, T., Englot, B., Ratti, C., Rus, D. (2021) LVI-SAM: Tightly-coupled Lidar–visual–inertial odometry via smoothing and mapping. arXiv preprint arXiv: 2104.10831.
[17] Djuric, P. M., Kotecha, J. H., Zhang, J., Huang, Y., Ghirmai, T., Bugallo, M. F. (2003) Particle filtering. IEEE Signal Processing Magazine, 20(5), 19–38.
[18] Zimmerman, N., Guadagnino, T., Chen, X., Behley, J., Stachniss, C. (2022) Long-term localization using semantic cues in floor plan maps. arXiv preprint arXiv: 2210.01456.
[19] Liu, Y. C., Lin, K. Y., Chen, Y. S. (2008) Bird's-eye view vision system for vehicle surrounding monitoring. In Proceedings of the International Workshop on Robot Vision (RobVis), Lecture Notes in Computer Science, vol. 4931, pp. 207–218. Springer.
[20] Liu, Z., Tang, H., Amini, A., Yang, X., Mao, H., Rus, D., Han, S. (2022) BEVFusion: Multi-task multi-sensor fusion with unified bird's-eye view representation. arXiv preprint arXiv: 2205.13542.
[21] Bai, X., Hu, Z., Zhu, X., Huang, Q., Chen, Y., Fu, H., Tai, C. L. (2022) TransFusion: Robust LiDAR-camera fusion for 3D object detection with transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1090–1099. IEEE.
[22] Cao, Y. U., Fukunaga, A. S., Kahng, A. (1997) Cooperative mobile robotics: Antecedents and directions. Autonomous Robots, 4(1), 7–27.