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. Maze-exploring robots are a large contender in this application, and have numerous applications such as transporting logs, post-disaster search and rescue, underground pipeline inspection, and medical navigation guidance. The work of these tasks puts significant pressure on a robot in terms of its autonomy in perceiving the surrounding environment, plan its path and avoids obstacles in new environments.

Nevertheless, the maze environments in the real world are usually complex, cluttered with obstacles and lack complete information, which is very demanding to the environmental perception

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.LD28245

and cognition of a robot. Trusting in one sensor only (either in infrared, ultrasonic, or visual sensor) tends to come out with a small field of view, low recognition accuracy and low environmental adaptability [1]. Moreover, odometry and inertial measurement unit (IMU) based positioning methods have a tendency to accumulate errors, hence they would experience difficulty in providing continuous and stable navigation [2].

Many multi- sensor fusion and collaborative perception systems have been suggested to overcome these challenges. Various sensors are incorporated into systems such as the lidar, cameras, IMUs and encoders. Kalman filtering and deep learning models of fusion algorithms are used [3,4]. The data provided by various sources pass complements on each other to improve precision of perceptions. The fault tolerance and robustness of the system is improved. Moreover, collaborative perception system allows distribution of environmental knowledge, map stitching and task coordination in multi-robot systems which contributes to high levels of overall exploration and levels of intelligence considerably [5].

Over the last several years, this sphere has been evolving on a accelerated pace of low-level data fusion into high-level intelligent perception. Earlier heuristic rules have been developed into probabilistic graphical models and now developed into the current deep fusion and multi-robot distributed SLAM systems, the technical framework of which is also more developed today [6].

This paper presents the ideas and architecture of the multi-sensor fusion and collaborative perception technologies, which are specifically applied in maze-exploring robots. It examines the current technological directions and research advances in this area based on the perspectives of perception system structure, multi-source information fusion technology, on-robot collaboration technology, common application examples, and research challenges in the future. It further suggests development tendencies and study opportunities of the multi-sensor collaborative perception in autonomous search of challenging environments.

2. The perception system for maze exploration robots

The maze exploration involves the use of robots to carry out obstacle avoidance under steady velocity, positioning and mapping, as well as guidance to targets in unfamiliar settings. The system commonly employs many sensors in conjunction to win a trade off in performance during real time and strength.

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) are all standard sensors and functions. These sensors are the complementary ones in terms of the field of view, scale, and environmental adaptability, but both have limitations. As an example, lidar is easily influenced by glass and dust; visual mechanisms are prone to interference by illumination and occlusion; ultrasonic sensors are characterized by low directional resolution; and inertial measurement devices exhibit problems of drift [7].

Consequently, one modality of perception is not enough to address more sophisticated situations in the maze as in frequent turns and narrow walks. Multi-sensor fusion improves positioning accuracy and robustness of the system by means of spatio-temporal synchronisation, redundancy checking and feature supplementality [8]. Common subsets are 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 involves the robots to accomplish tasks, including obstacle avoidance, positioning, mapping, and target guidance, in unknown environments in a stable manner. The system usually employs multiple sensors, taking and addressing this issue to achieve a balance between the real-time performance and a robust system.

DOI: 10.54254/2755-2721/2025.LD28245

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 fundamental concept of the data layer fusion is to normalise and simultaneously exploit the raw data, at the most bottom-level [10]. This is a process that has three steps. First, there is spatio-temporal alignment; secondly, external parameter calibration with the unification of the coordinates, observation data of various sensors are projected onto a standard reference frame, and then the joint coding or resampling, which can be at the pixel, echo, or point level, is carried out and passed through downstream modules. The approach optimises the storage of supplementary information [11]. In complicated regions of the road like sharp curves, high density of occupancy, or mirror or glass reflection surfaces, the technique is efficient in producing channel identification, boundary edges of obstacles, and the geometric details. In order to do that, one needs more engineering requirements. The system needs to be more precisely synchronised, have more constant values of the external parameters and better bandwidth, storage and computer capabilities. Any deviation of synchronisation will result in cumulative errors of later modules. In real practice, the joint method of radar vision candidate region generation + image refinement (RRPN) is a common practice [12]. This method combines three-dimensional candidate areas based on the bird-eye view of LiDAR and RGB image features which increase the rate of detection.

3.1.2. Feature layer fusion (mid fusion)

The focus on alignment of effective information is featured in feature layer fusion. Individually each of the modalities will encode features in its respective channel (e.g., semantic feature of images, geometric feature of lidar, short-term features of trajectory of inertial measurement units), with complementary exchange of information finally leading to the formation of richer feature representations [10]. This method generally balances accuracy and real-time performance better, and is more resilient to small-scale non-specific external parameter drifts as well as non-strict synchronisation. It is, therefore, the major avenue of exploring the maze. It is more specifically interesting to note that the quality of alignment can establish the upper bound to fusion effectiveness: projection error or time misalignment can produce so-called false consistency. Practically, MV3D [13] approach can improve significantly the detection accuracy by combining three dimensional candidate areas that are obtained through the birds-eye view of the lidar and features of the RGB image.

3.1.3. Decision-layer fusion(late fusion)

The independence of processing chains and fault isolation used in decision-layer integration. A series of tasks, such as detection, tracking, and segmentation are then performed separately, by each sensor or subsystem. Finally, normalisation of confidence, conflict identification, and strength check are done at the result layer [10]. This is a technique that encourages gradual integration and impoverished operation. Even in the case of a chain degradation, the other chains can continue to provide service. It is more dependent upon upstream data quality, however, and can be lost in obsolete detail. All in

DOI: 10.54254/2755-2721/2025.LD28245

all, the decision layer is an outstanding establishment on fault fallback and fault tolerance, which has become a foundation on system stability.

3.1.4. Hybrid-layer fusion

Single-level fusion is sometimes not sufficient to meet all the perception demands in practical robotic systems, especially when complex maze environments are to be addressed (high dynamics and uncertainty). Therefore, more studies are looking into hybrid fusion direction according to the complexity of the tasks. These methods have a higher system strength and image precision through adaptive incorporation of sensory data in a variety of levels including raw data processing to the ultimate decision output. This active time scheduling of various fusion approaches proves to be more adaptable and perform better in dynamic and complicated situation and as such it is a big step forward developing this as a main direction in a 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.

$$\boldsymbol{p(x_k|z_{1:k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|z_{1:k-1})dx_{k-1}} \tag{1}$$

$$\boldsymbol{p(x_k \mid z_{1:k}) = \frac{p(z_k|x_k)\,p(x_k|z_{1:k-1})}{\int p(z_k|x_k)\,p(x_k|z_{1:k-1})\,dx_k}} \tag{2}$$

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 $\boldsymbol{p(x_k|x_{k-1})}$ under the first-order Markov assumption and marginalizes the unobserved $\boldsymbol{x_{k-1}}$by integration, yielding the prior $\boldsymbol{p(x_k|z_{1:k-1})}$ conditioned on past measurements. $\boldsymbol{p(x_k|x_{k-1})}$ 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 $\boldsymbol{k}$; $\boldsymbol{p(x_{k-1}|z_{1:k-1})}$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 $\boldsymbol{p(x_k \mid z_{1:k})}$, where $\boldsymbol{p(z_k \mid x_k)}$ is the measurement likelihood and the integral in the denominator is a normalizing constant ensuring the distribution integrates to one [14].

This algorithm shows good consistency with semantics and scalability, which can be harmonized with Kalman filtering, particle filtering and graph optimisation algorithms. Its key weaknesses are that its sensitivity to model assumptions, which is likely to be found in the wrong set of model assumptions, and its high reliance on prior knowledge. Data-driven noise calibration and data quality gate mechanisms can be used in the maze-based applications to reduce estimation biases. These measures allow stable updates across various degraded segments without losing the interpretation of the results by factors of noise that vary over time, as they consolidate noise time-varying properties using replay calibration and cross-validation.

DOI: 10.54254/2755-2721/2025.LD28245

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

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

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

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

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 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].

Recent works show LIO-SAM is capable of skeletonizing point cloud to remove skew and increasing real-time operational efficiency by building a laser radar inertial odometer and integrating the IMU in a factor graph priori [16]. LVI-SAM uses an inter-visual-inertial-laser-initiated (LIS) coupled factor graph. The estimation of VIS initialisation is done by the LIS estimation, the LIS also uses VIS estimation as initial values to aid the scan matching, and the loops are first identified by the VIS and then schemes are refined by the LIS. It can also survive backed by the failure of either of the sub systems and hence maintains robustness both in textureless and featureless environments [17].

This technique is of low latency, interpretable and well integrated with optimisation of the backends. It has weaknesses such as an sensitivity to time synchronisation and noise modelling of external participation as well as linearisation drift accretion. Rolling shutter compensation is used in labyrinthine environments, and the observation weights of these situations are hot-swapped, and loop factor correction is used to remove errors in linearisation. There is a single formulation of attitude and pose updates basis on an error state. Such strategy is used to ensure continuous and stable position despite time-varying degradation.

DOI: 10.54254/2755-2721/2025.LD28245

3.2.3. Particle filtering (PF/AMCL/RBPF)

PF uses a set of samples to approximate the posterior distribution and thus, it directly treats non-linear, non-Gaussian and multi-modal distributions. It best at global positioning, abduction recovery and re-localisation, and supplements continuous filtering [7].

The sequential Monte Carlo (SMC) technique of particle filtering (PF) is an algorithmically simpler algorithm that uses importance sampling and discrete stochastic measures to provide an approximation of the posterior distribution of the state in a recursive manner. This methodology was originally used in the field of polymer growth, but eventually was extended to physics and engineering. The computational complexity and processing power currently limited its application in the past. However, in recent years it has been reemerging at a very fast pace with the development of hardware and parallel computing, and its possibilities in signal processing. Major engineering issues are degradation in weight and lack of sample diversity. As the effective number of particles (ESS) becomes smaller, few high-weight particles will dominate the posterior distribution and the estimation will become unstable. Widely used countermeasures are (adaptive) resampling to discourage weight concentration, better proposal distributions (knifing kinematic priors with observed likelihoods to create locally optimal proposals), low-variance resampling, and particle tempering (random perturbation or MCMC moves) to retain exploration canopies [18].

Most recently, KLD-Sampling has managed to attain a trade-off between accuracy and real-time performance through an adaptive scaling of the particle size, which is driven by upper bounds of KL errors. Semantic maps with PF are used as indoor global positioning to enhance re-localisation with semantic anchor-based disambiguation [19]. To balance the use of policy learning and particle emission, Differentiable Active PF engages in active information gathering and minimizes exploration costs, allowing PF to exhibit better convergence and recovery properties on challenging topologies and at scale.

Simultaneously, this methodology is exceptional regarding the preservation of self-localisation during high uncertainty and multiform situations in addition to its absence of capability to withstand sensor failures. It has weaknesses of complexity dependence on the number of particles linearly with these particles and a mismatch of likelihoods hindering convergence. As a result, on the case of maze applications, it combines hierarchical particle scattering, semantic priors, as well as adaptive particle count control as a strategy to control computational expenditure. Moreover, PF can be used as an emergency bypass to the main odometer and allow controlling a quick recovery of the trajectory in case of misalignment.

3.2.4. Deep learning–based Fusion (BEV)

Using cross-view integration to generate bird-eye-view (BEV) as the single point of representation requires the following sequence of operations: calibration, rectification, alignment, stitching, and fusion. The raw models are also referred to as the fiskeye models of FOV distortion and LM optimisation, which uses these models to estimate intrinsic parameters and carry out distortion correction of multiple fiskeye cameras. Any of the views are then transformed to a top-down coordinate system through the homothetically consistent HHH transformation through a ground plane. The selection of sewing points is based on quadratic error field which uses calibration residual, greedy or dynamic programming is used to select the smallest-residual seam when parts are overlapping. The pixel sequences are then recorded at seams by applying dynamic image time regularisation (DIW). The tight-support RBF of Wendland then spreads seam registration deformations throughout the whole picture. Lastly, the global exposure is brought to normal with normalisation via gain adjustment or bias adjustments. A weighted fusion is performed on a calibration residual × distance-to-boundary weight operation to remove apparent cracks and brightness discontinuities. Also, a

DOI: 10.54254/2755-2721/2025.LD28245

history value is added to DIW to avoid inter-frame jitter so that the BEV output remains constant during occlusion and temporal variations under a small pose changes [20].

Recently, BEVFusion integrates cameras and LiDAR into a common BEV along with drastically enhancing the cross-modal pooling effectiveness [21]. Transfusion eliminates pixel point cloud misalignment by soft decoder alignment which is more robust in low-light situations [22]. BEVDet4D proposed use of temporal BEVs to minimise the error in velocity estimation by significant factors whereas RCBEVDet incorporates radar-camera data into BEVs to improve penetration and speed tracking. The FusionLoc used a camera + 2D LiDAR that used multi-head attention to do end-to-end pose regression indoors which significantly enhanced robustness at corners of mazes, occlusion and smoke-filled areas.

At the same time, the method performs best at in-depth high-dimensional semantic insight, optimised end-to-end, and coordinated multi-task. Its limitations are its high data and computation requirements as well as the probability of a mismatch of domains. Therefore in the maze applications, it provides synthetic to real domain adaptation and testing alignment together with strong kernel or loopback interaction in the backend. This attains steady state operation striking a balance between high-performance ceilings and deployability.

4. Cooperative perception and multi-robot collaboration

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

Field of view constraints, degradation of measurement and computational constraints often limit single robots in highly occluded, structurally repeating, and time-varying worlds, like in mazes. As a result, the essence of collaborative perception is to exchange the local observations, estimates, and uncertainties of several robots in a single spatio-temporal map and thus create a larger virtual field of view and more consistent global thinking. In particular, several robots might share or share semantic slices, keyframes, across-robot feedback loops, relative pose / covariances, observation quality indicators (e.g. brightness, echo intensity, point cloud density). Therefore, the obstructions of corners, low light spots, passages full of dust, glass surfaces, etc. can be compensated by the view of other people. In order to avoid bottlenecks caused by bandwidth and latency, information is generally structured hierarchically, the only direct linkage being directly between raw data on short durations in local emergencies, but much more often between summarised exchanges in feature, map or decision layer. Such interactions are consistently timed, and coordinate-systemed with external parameter version numbers, to make more efficient aligning the backend with them. By the means of such a method of sharing tasks coupled with a weighting in terms of quality, it is possible to establish a system which is robust and most effective in terms of exploration in a way which is not overly constrained by computational or communication costs.

4.2. Multi-robot perception architectures

Swarm architectures of multi-robots offer the support of collective behaviour and decide on capabilities and constraints of a system. The main points to note are; centralised or decentralised (further subdivided into hierarchical and fully distributed), role differentiation (homogeneous or heterogeneous), communication, and the ability to model others. It has many times been claimed that decentralisation has the attractive characteristics of fault tolerance, parallelism, and scalability but comparisons between these methods in theory or experiment have rarely been made. Practically, many systems transition to take a hybrid solution: decentralised on average, but there will be a leader or a central planner to coordinate on the high level. Heterogeneity complicates the dispensing of tasks and other difficulties in modelling taskmates; the extent to which an individual can complete a task independently can be built in terms of task coverage (where low values, it depends on group work).

DOI: 10.54254/2755-2721/2025.LD28245

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 .

5. Conclusion

The paper is developed on the closed loop system of perception estimation coordination planning: since maze environments are characterized by both high occlusion and multi-class degradation, the single sensor is likely to be limited by field-of-view and noise models. Hence, these strong background knowledge needs to be formulated in a well-articulated multimodal synthesis. Since a long-term temporal and spatial consistency across modalities may be difficult to attain, the organising principle is hierarchical fusion: early fusion captures detail where raw data allows high-precision alignment; feature-level fusion has the best accuracy-latency trade-offs on alignment errors and computational costs; result-level fusion is where service continuity is critical. Inter-layer switching is made easy with adaptive scheduling. Factor graph smoothing and error state filtering is alcoholically feasible to support high frequency, tightly coupled LIO or VIO. Global localisation and kinematic recovery are done using particle filtering with nonlinear and multi-modal posterior distributions. Meanwhile, BEV also combines geometry and semantics to ensure consistent corner performance, low-light and occlusion performance. At the same time, to increase the effective field of view and to reduce personal deterioration, multi-robot coordination is necessary; that is, sharing quality-labeled features, maps or loopback cues. Meanwhile, the mutually constraining improvements in the coverage efficiency and safety are achieved by path planning and task allocation. In addition, based on the assumption that long-term thermal drift of external parameters and prototypical sampling impose an asynchronous sampling constraint and OOSM a filtering hypothesis, and BEV may have mismatches in visibility to glass or smoke or high-reflective objects, embedded platforms are constrained by computational or energy factors with small communication bandwidth, and dense traffic requires high safety and deadlock avoidance. The system should also have the concomitant realization of online Q/R learning as well as external parameter self-calibration to achieve probabilistic consistency, introduce utility-latency-energy-driven hybrid scheduling to controllable degradation and a combination of radar or lightweight laser and adaptive testing to correct domain drift. Lightweight loopback cues and conservative information fusion ensure global consistency and edge-cloud event-driven mechanisms consume OOSM. Also, perceived, and planned are decoupled through uncertainty-based active exploration, selection of viewpoints, and an environment where decisions

DOI: 10.54254/2755-2721/2025.LD28245

depend on the market. It is supplemented by limitations like keep right, stop at intersections or mutually exclusive passage, etc. to minimize the number of collisions and congestion. This facilitates long-term independent performance in complex mazes of the real world by balancing accuracy, strength 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.

[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.