A Comprehensive Survey of Visual SLAM Technology: Methods, Challenges, and Perspectives

Aidos Ibrayev, Amanzhol Bektemessov
International Engineering Technology University, Almaty, Kazakhstan

***Abstract*—Visual Simultaneous Localization and Mapping (Visual SLAM) has become a cornerstone of autonomous navigation and spatial understanding in robotics, augmented reality, and computer vision. This review presents a comprehensive examination of algorithmic progress in Visual SLAM, focusing on the three principal paradigms: monocular, stereo, and RGB-D SLAM. Monocular SLAM, known for its minimal hardware requirements, has evolved from feature-based methods to deep learning-enhanced systems, addressing challenges like scale ambiguity and drift. Stereo SLAM leverages depth through triangulation, improving scale accuracy and robustness, particularly in dynamic and low-texture environments. RGB-D SLAM, utilizing depth-sensing technology, has enabled dense and semantically enriched mapping, finding significant application in indoor and real-time scenarios. Through a chronological and technical exploration of representative methods including RatSLAM, ORB-SLAM, DSO, ProSLAM, ElasticFusion, DynaSLAM, and recent hybrid and learning-based frameworks. This review identifies major milestones and architectural innovations across paradigms. A cross-paradigm analysis highlights the trade-offs in accuracy, computational efficiency, and adaptability, while also discussing emerging trends such as semantic integration, multimodal fusion, and neural implicit representations. Furthermore, the paper outlines future directions that include lifelong learning, real-time deployment on edge devices, dynamic environment adaptation, and the convergence of geometry and learning-based pipelines. Supported by a detailed taxonomy and historical evolution illustrated in visual summaries, this review serves as a foundational reference for researchers and developers aiming to understand and contribute to the advancement of Visual SLAM technologies in both academic and real-world contexts.**

***Keywords*—Visual SLAM; monocular SLAM; Stereo SLAM; RGB-D SLAM; 3D mapping; pose estimation; loop closure; semantic SLAM; deep learning; sensor fusion**

I. INTRODUCTION

Simultaneous Localization and Mapping (SLAM) is a fundamental problem in robotics and computer vision, enabling an agent to build a map of an unknown environment while estimating its own pose within that space. Among the various modalities of SLAM, Visual SLAM (V-SLAM) has gained increasing attention due to its cost-effectiveness, wide applicability, and ability to capture rich scene semantics using standard cameras [1]. Unlike traditional SLAM approaches that rely on expensive range sensors such as LiDAR, visual SLAM uses images as the primary sensory input, offering a lightweight and scalable solution for many real-world applications including autonomous vehicles, augmented reality (AR), unmanned aerial vehicles (UAVs), and mobile robotics [2].

Visual SLAM systems are typically classified into three main categories based on the type of visual input: monocular, stereo, and RGB-D SLAM. Monocular SLAM uses a single camera and offers a compact and inexpensive setup but suffers from scale ambiguity and sensitivity to initialization [3]. Stereo SLAM employs a pair of cameras to estimate depth via triangulation, thereby overcoming scale issues and providing more stable estimates in textured environments [4]. RGB-D SLAM leverages cameras that deliver both color and depth information simultaneously, enhancing accuracy in indoor and structured scenes while introducing hardware constraints and limitations in outdoor or long-range settings [5].

Over the past two decades, the field has witnessed a remarkable evolution in algorithmic design, ranging from sparse, feature-based methods to dense and semi-dense approaches, and more recently, learning-based frameworks that integrate deep neural networks with traditional geometric pipelines [6]. Despite significant progress, many challenges persist, such as dealing with dynamic environments, achieving real-time performance on resource-constrained devices, and ensuring robustness under varying lighting or texture conditions [7]. Furthermore, achieving generalizable SLAM that can adapt across domains and scales remains an open research problem [8].

This review paper provides a structured and comprehensive overview of Visual SLAM methods across the three primary categories. By examining algorithmic trajectories, technical innovations, and representative benchmarks, this survey aims to map the evolving landscape of SLAM research and identify promising directions for future exploration.

II. BACKGROUND AND FUNDAMENTALS

Visual SLAM combines techniques from geometry, signal processing, and optimization to simultaneously estimate a camera's pose and reconstruct its environment. The core objective is to recover the trajectory $T_t = SE(3)$ and a map M from a sequence of image observations $\{I_t\}$ over time. Mathematically, the SLAM problem can be formulated as a maximum a posteriori (MAP) estimation:

$$\arg\max_{X,M} p(X, M \mid Z) \tag{1}$$

where, X is the set of poses, M is the map, and Z denotes the sensor measurements [9]. Factor graph models are commonly used to express the probabilistic dependencies between variables in SLAM, allowing the estimation to be represented as a graph optimization problem [10].

A key task in visual SLAM is camera pose estimation, typically achieved by minimizing the reprojection error:

$$\min_T \sum_{i=1}^{N} \left\| u_i - \pi\left(T \cdot X_i\right) \right\|^2 \tag{2}$$

where, u_i are image features, π is the projection function, and X_i are 3D landmarks [11]. Map representation varies from sparse point clouds to dense voxel grids, depending on the method [12].

Loop closure detection corrects accumulated drift by identifying previously visited locations and enforcing pose graph consistency [13]. Pose graph optimization is often solved using nonlinear least squares:

$$\min_X \sum_{(i,j)\in\varepsilon}^{N} \left\| f_{ij}\left(X_i, X_j\right) - z_{ij} \right\|^2 \tag{3}$$

where, f_{ij} models relative motion and z_{ij} is the measured constraint [14]. Feature extraction, tracking, and data association form the visual front-end, while back-end optimization refines the trajectory and map [15].

Recent works also incorporate deep learning for depth estimation and semantic mapping, improving robustness in unstructured environments [16]. Benchmarks such as KITTI and TUM RGB-D provide standardized datasets for evaluation and comparison [17].

III. Algorithmic Paradigms for Slam

This section synthesizes the principal algorithmic paradigms that have shaped SLAM development for mobile robots, providing a structured overview of classical and contemporary frameworks. We first examine filter based methods such as EKF-SLAM and UKF-SLAM that leverage recursive Bayesian estimation to jointly infer robot trajectory and map features. Next, particle filter approaches and their Rao-Blackwellized variants (FastSLAM) are discussed for their ability to represent multimodal posteriors in highly nonlinear settings. We then turn to graph based optimization techniques, which cast SLAM as a sparse nonlinear least squares problem over pose and landmark variables, and explore incremental solvers such as iSAM. Finally, hybrid and emerging paradigms that integrate multiple estimation strategies, learning based components, and submapping schemes are reviewed to highlight the evolving landscape of SLAM algorithms.

A. Filter-Based Approaches

Filter based SLAM methods formulate the simultaneous estimation of robot pose and map features as a recursive Bayesian filtering problem [16]. At each time step, the joint state vector.

$$s_t = \begin{bmatrix} x_t \\ m \end{bmatrix} \tag{4}$$

where, x_t denotes the robot pose and m concatenates all landmark coordinates, is represented by a Gaussian belief $N\left(\hat{s}_t, P_t\right)$. The filter alternates between a prediction step

$$\begin{aligned} \hat{s}_{t|t-1} &= f\left(\hat{s}_{t-1}, u_t\right), \\ P_{t|t-1} &= F_t P_{t-1} F_t^T + Q_t \end{aligned} \tag{5}$$

and a correction step upon receiving a measurement z_t:

$$\begin{aligned} K_t &= P_{t|t-1} H_t^T \left(H_t P_{t|t-1} H_t^T + R_t\right)^{-1} \\ \hat{s}_t &= \hat{s}_{t|t-1} + K_t\left(z_t - h\left(\hat{s}_{t|t-1}\right)\right) \\ P_t &= \left(I - K_t H_t\right) P_{t|t-1} \end{aligned} \tag{6}$$

Here, f and h are the process and measurement models, F_t and H_t their Jacobians, and Q_t, R_t covariance matrices. Within this framework, the Extended Kalman Filter SLAM (EKF-SLAM) linearizes f and h about the current estimate, resulting in first-order approximations of non-linear dynamics and sensor models [17]. EKF-SLAM maintains a full covariance matrix across robot and landmark states, ensuring consistent uncertainty propagation but incurring $O\left(n^2\right)$ computational complexity in the number of landmarks n. To mitigate linearization errors, the Unscented Kalman Filter SLAM (UKF-SLAM) employs the unscented transform: a set of deterministically chosen sigma points $\left\{\chi_{t-1}^i\right\}$ is propagated through the true non-linear functions, yielding more accurate posterior mean and covariance estimates without explicit Jacobian computation [18]. UKF-SLAM typically achieves higher accuracy in highly non-linear regimes at the cost of increased per-step computation proportional to the number of sigma points ($2L+1$ for a state of dimension L) [19]. Both EKF-SLAM and UKF-SLAM serve as foundational benchmarks against which more advanced, optimization-based, and hybrid SLAM paradigms are compared.

B. Particle Filter SLAM

Particle-filter SLAM methods approximate the full SLAM posterior $p\left(x_{0:t}, m \mid z_{1:t}, u_{1:t}\right)$ by a finite set of N weighted particles $\left\{x_{0:t}^i, m^i, w_t^i\right\}_{i=1}^N$. At each time step, particles are propagated according to the motion model $x_t^i \approx p\left(x_t \mid x_{t-1}^i, u_t\right)$ and importance weights are updated via:

$$w_t^i \propto w_{t-1}^i, p\left(z_t \mid x_t^i, m^i\right) \tag{7}$$

Followed by resampling to focus computational resources on high-likelihood hypotheses. Unlike Gaussian-based filters, particle filters can represent arbitrary, multi-modal distributions, making them well-suited to highly non-linear and ambiguous observation models. However, naive particle-filter SLAM suffers from the "curse of dimensionality," as the joint state space grows with the number of landmarks, leading to excessive computational demand and rapid sample impoverishment when N is insufficient.

FastSLAM Variants. FastSLAM addresses these challenges by factorizing the posterior into a product of a trajectory posterior and independent landmark posteriors:

Trajectory estimation is performed via a particle filter, while each landmark's position is maintained by a separate, low-dimensional Kalman filter (or alternative Gaussian estimator) [20]. FastSLAM 1.0 introduced this Rao-Blackwellized approach, reducing computational complexity from $O(NM^2)$ to $O(NM)$. FastSLAM 2.0 refined the proposal distribution by incorporating the current measurement into the sampling step, thereby improving sample efficiency and reducing particle deprivation in highly informative environments [21]. Subsequent variants have further enhanced robustness through adaptive resampling thresholds, hierarchical landmark clustering, and incorporation of grid-based mapping techniques for environments with dense feature distributions [22]. These innovations have established FastSLAM as a versatile, real-time SLAM framework on mobile-robot platforms with moderate computational resources.

C. Graph-Based Optimization

Graph-based SLAM reformulates the joint estimation problem as a nonlinear least-squares optimization over a sparse graph, in which nodes represent robot poses (and optionally landmark states) and edges encode spatial constraints derived from odometry and observations. In the pose-graph formulation, only robot poses $\{x_i\}$ are included as variables, and measurements between poses (e.g., odometry or loop closures) define binary factors. The optimization seeks

$$x^* = \arg\min_{x} \sum_{(i,j)\in\varepsilon} \| z_{ij} \Theta \left(x_i^{-1} \oplus x_j \right) \|_{\Omega_{ij}}^2 \quad (8)$$

where, z_{ij} is the measured relative transform between poses i and j, Ω_{ij} the information matrix, and $\oplus$, $\ominus$ manifold-specific composition and inverse-composition operators [23]. Factor-graph SLAM extends this concept by including landmark nodes $\{m_k\}$ and multiary factors, yielding a unified representation amenable to general-purpose solvers. The factor-graph cost function becomes.

$$\{x^*, m^*\} = \arg\min_{x,m} \sum_{f\in F} \| \phi_f \left(x_f, m_f \right) - z_f \|_{\Omega_{ij}}^2 \quad (9)$$

where, each factor f relates a subset of pose and landmark variables through the measurement function ϕ_f and information matrix Ω_f [24]. Sparsity in the graph structure enables efficient exploitation of sparse linear algebra techniques, yielding favorable scaling to large-scale environments.

Incremental smoothing and mapping (iSAM) algorithms tackle the computational demands of online graph optimization by updating the solution and factorization incrementally as new measurements arrive. The original iSAM algorithm incrementally constructs and maintains a square-root information matrix via QR or Cholesky factorization, avoiding the need to re-solve the entire system at each step [25]. iSAM2 further introduces a Bayes tree data structure that encapsulates the conditional dependency structure of variables, allowing selective relinearization and re-factorization only in affected subtrees when new factors are added or existing factors change [26]. This strategy achieves real-time performance on resource-limited platforms while preserving consistency and accuracy, making iSAM variants a de facto choice for optimization-based SLAM in mobile-robot applications.

IV. VISUAL SLAM

Visual SLAM (Simultaneous Localization and Mapping) is a technique that leverages visual input, typically from one or more cameras, to perform real-time mapping and localization in unknown environments. As shown in Fig. 1, Visual SLAM is broadly categorized into three main types: Monocular SLAM, Stereo SLAM, and RGB-D SLAM, each characterized by its unique camera configuration and depth perception capabilities. This classification allows for flexibility in application domains, ranging from lightweight mobile robotics to dense indoor reconstruction [18].

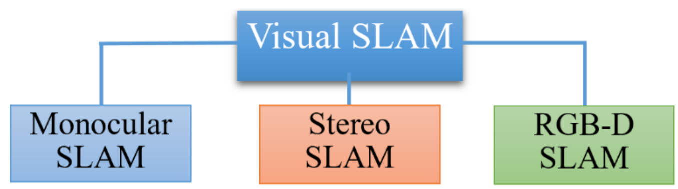

Fig. 1. Taxonomy of Visual SLAM: Classification into Monocular, Stereo, and RGB-D SLAM approaches.

Monocular SLAM employs a single camera and is attractive for its simplicity and low hardware cost. However, it faces challenges such as scale ambiguity and higher drift in long trajectories [19]. Stereo SLAM, by using two synchronized cameras, can triangulate depth directly, resulting in improved robustness and scale consistency [20]. In contrast, RGB-D SLAM integrates color and depth data from structured light or time-of-flight sensors, enabling dense and accurate reconstructions, particularly in indoor scenes [21].

Recent advancements in visual SLAM have focused on combining geometric methods with learning-based models to enhance robustness in dynamic or poorly textured environments [22]. Additionally, visual-inertial fusion and multi-sensor integration have further extended the capabilities of SLAM systems under challenging conditions [23]. These developments highlight the rapid evolution of visual SLAM technologies across all three categories.

V. MONOCULAR SLAM: EVOLUTION AND METHODS

Monocular SLAM has emerged as one of the most thoroughly investigated areas in the broader SLAM research landscape, primarily due to its minimal hardware requirements and its adaptability across diverse platforms and applications.

By using a single RGB camera, monocular SLAM systems aim to simultaneously estimate camera motion and reconstruct the surrounding environment. However, the absence of direct depth measurements introduces inherent challenges such as scale ambiguity, motion degeneracy in planar scenes, and increased sensitivity to lighting and texture variations [24]. As depicted in Fig. 2, the evolution of monocular SLAM has progressed through a series of methodological innovations in both the visual front-end and the back-end optimization strategies.

Initial efforts began with RatSLAM (2004), which was biologically inspired and modeled after rodent navigation systems. While simplistic, it laid foundational principles in appearance-based mapping. This was succeeded by MonoSLAM (2007), a pioneering approach that applied an Extended Kalman Filter (EKF) to jointly estimate camera pose and sparse 3D map points in real time [25]. The advent of PTAM in the same year introduced a game-changing architectural shift by decoupling tracking and mapping into parallel threads, drastically improving the performance of real-time monocular SLAM systems [26].

Subsequent methods diverged into two main categories: feature-based and direct methods. Feature-based approaches rely on the detection and matching of keypoints, while direct methods operate on image intensities without extracting features. DTAM and LSD-SLAM demonstrated that it is possible to achieve dense and semi-dense reconstructions using photometric information alone, resulting in more detailed environmental maps [27][28]. These techniques offered higher resolution outputs but required careful handling of photometric calibration and motion assumptions.

A major leap occurred with the release of ORB-SLAM in 2015, which combined ORB keypoint descriptors with a robust framework for tracking, mapping, loop closure, and relocalization [29]. Its success led to further extensions, including ORB-SLAM2 and ORB-SLAM3, which supported multiple sensor configurations such as stereo and RGB-D, and included inertial fusion for improved robustness in challenging motion conditions [30]. These improvements allowed for more accurate scale estimation and reduced drift during long-term operations [31].

Simultaneously, direct methods like DSO (Direct Sparse Odometry) and its variants, including Stereo DSO, LDSO, and V-LDSO, refined photometric error minimization and introduced keyframe-based filtering to enhance both accuracy and computational efficiency [32][33]. These methods, while sensitive to illumination changes, offered high precision in texture-rich environments and demonstrated robustness in frame-to-frame tracking without reliance on feature extraction.

Hybrid approaches have also gained attention, blending the strengths of direct and feature-based methods. Algorithms like SVO (Semi-Direct Visual Odometry), PL-SVO (Point and Line SVO), and CNN-SVO achieved a balance between robustness and efficiency by integrating direct photometric tracking with geometric landmarks [34][35]. These approaches reduced the dependency on handcrafted features and improved performance in a wider range of scene geometries.

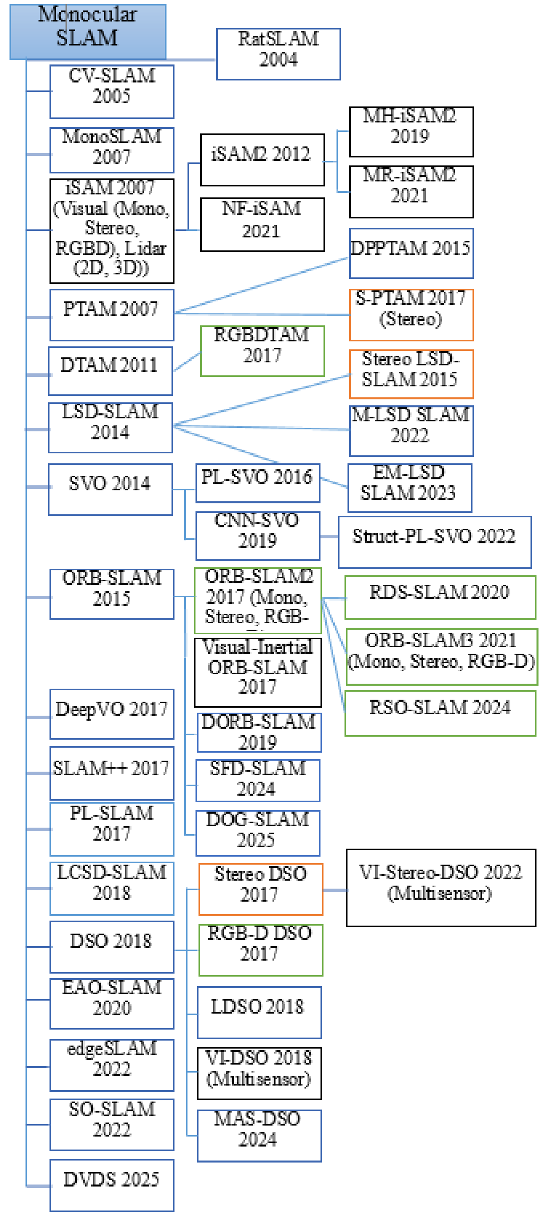

Fig. 2. Chronological development of monocular SLAM methods from 2004 to 2025.

In recent years, the incorporation of deep learning has significantly influenced the monocular SLAM domain. Frameworks such as DeepVO, edgeSLAM, and CNN-SLAM utilized neural networks for depth prediction, feature learning, and semantic scene understanding [36][37]. These models not only enhanced robustness in low-texture or dynamic environments but also enabled the inference of scene geometry from a single image. Moreover, methods such as SLAM++, SO-SLAM, and Struct-PL-SVO integrated object-level and structural cues into the mapping process, moving toward semantic and structural awareness in localization [38][39].

Recent state-of-the-art systems like DVDS (2025) and MAS-DSO demonstrate the growing emphasis on multi-sensor integration, semantic enrichment, and real-time adaptability [40][41]. These systems incorporate inertial data, semantic segmentation, and learned depth priors to improve resilience in diverse and unpredictable environments.

Despite these advancements, monocular SLAM still faces unresolved issues, particularly in handling motion blur, rapid camera movements, dynamic object interference, and texture-poor scenes. The pursuit of robustness, generalization, and computational scalability continues to drive innovation in this domain, with future directions likely to focus on hybrid geometric-learning systems, unsupervised adaptation, and cloud-enhanced mapping frameworks.

VI. Stereo SLAM: Advances and Applications

Stereo SLAM systems utilize a pair of synchronized cameras to estimate depth via triangulation, enabling accurate 3D reconstruction and metric-scale motion estimation. This configuration eliminates the scale ambiguity inherent in monocular systems and enhances robustness in texture-poor or low-light environments [42]. As illustrated in Fig. 3, the development of stereo SLAM has progressed steadily with contributions that focus on improving accuracy, structural awareness, and adaptability in dynamic scenarios.

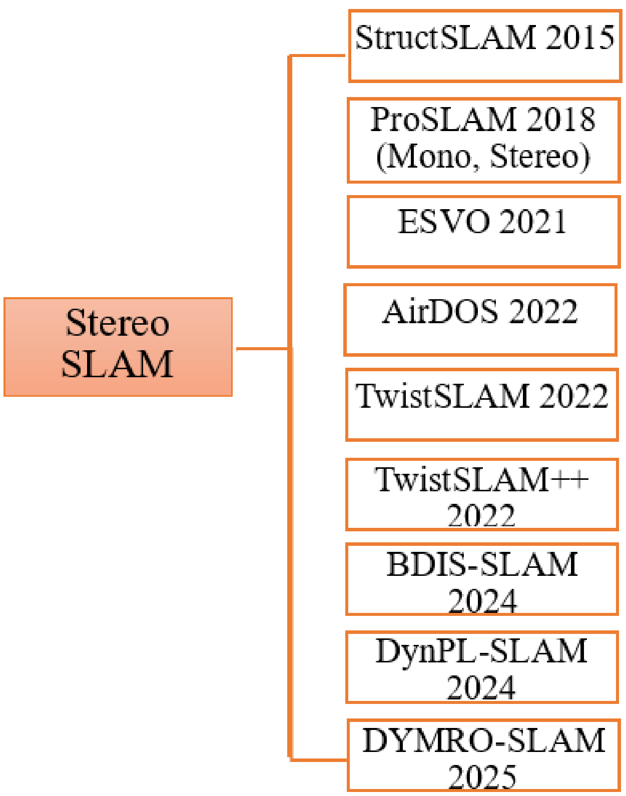

Fig. 3. Evolution of stereo SLAM methods and their applications.

StructSLAM (2015) laid the foundation by integrating stereo imagery with structural constraints to achieve semantically meaningful mapping [43]. The introduction of ProSLAM (2018) provided a lightweight and modular stereo SLAM framework capable of real-time performance, supporting both mono and stereo inputs for flexibility across platforms [44]. More recent methods, such as ESVO (2021) and AirDOS (2022), have extended stereo SLAM to address aerial applications and dynamic obstacle detection using stereo image alignment and depth filtering techniques [45][46].

The emergence of TwistSLAM and TwistSLAM++ (2022) represents a shift toward incorporating twist-based motion models for improved tracking under fast camera movements [47]. These systems enhance pose estimation stability while reducing drift in challenging trajectories. Building on these ideas, BDIS-SLAM and DynPL-SLAM (2024) have pushed stereo SLAM into the realm of dynamic environments, where the presence of moving objects demands adaptive strategies for scene segmentation and motion compensation [48].

Finally, DYMRO-SLAM (2025) exemplifies the future direction of stereo SLAM by integrating motion robustness, real-time segmentation, and adaptive feature association in highly dynamic and cluttered settings [49]. In addition to robotics, stereo SLAM has found applications in autonomous driving, drone navigation, and indoor mapping where depth accuracy and consistency are critical [50].

Collectively, the advancements in stereo SLAM reflect a balance between computational efficiency, structural integrity, and semantic awareness, making it a viable solution for diverse real-world deployment scenarios. However, the challenges of real-time performance in high-speed settings and long-term consistency in large-scale environments continue to motivate ongoing research.

VII. RGB-D SLAM: Depth-Enhanced Approaches

RGB-D SLAM systems integrate color (RGB) images with depth data, typically captured using structured light or time-of-flight sensors, to produce accurate and dense 3D maps in real-time. This additional depth channel enhances scene understanding, especially in indoor environments where scale accuracy and dense reconstruction are critical. As illustrated in Fig. 4, the evolution of RGB-D SLAM has been marked by diverse approaches that balance geometric precision, semantic richness, and computational efficiency.

Early RGB-D SLAM efforts began with KinectFusion (2011), which demonstrated the potential of consumer-grade depth sensors for dense real-time 3D reconstruction [51]. Shortly thereafter, Kintinuous (2012) extended KinectFusion by enabling continuous tracking and mapping across larger spaces [52]. Later methods such as ElasticFusion (2015) introduced surfel-based representations to maintain loop-closure consistency and surface flexibility during global optimization [53]. Similarly, BundleFusion (2017) incorporated global pose graph optimization with real-time feedback loops to minimize drift in long-term operations [54].

The year 2017 marked a proliferation of RGB-D SLAM algorithms. Systems such as VO-SF, Co-Fusion, and StaticFusion focused on improving tracking robustness and map representation in the presence of dynamic objects [55][56][57]. Others like MR-RGBD-SLAM and SWAIACP-SLAM added semantic cues and adaptive filtering mechanisms to improve depth consistency [58][59]. Simultaneously, methods like Detect-SLAM and CodeSLAM explored object-level SLAM and compact depth representations, respectively [60][61].

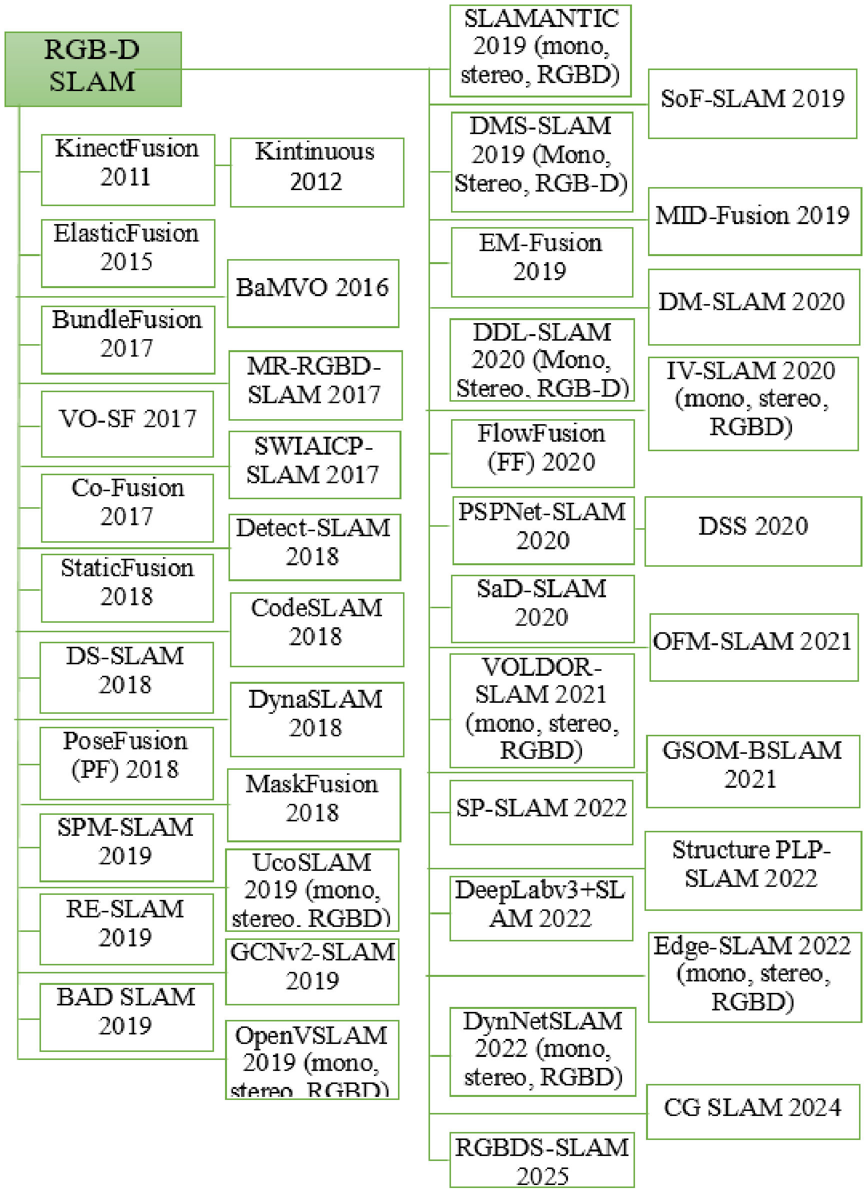

Fig. 4. Overview of RGB-D SLAM methods and their evolution.

In 2018, efforts like DS-SLAM, PoseFusion, and DynaSLAM focused on dynamic scene handling by introducing segmentation pipelines and learning-based object recognition to remove moving elements from mapping [62][63][64]. The inclusion of neural networks for motion segmentation and visual odometry marked a turning point, bridging traditional geometry-based techniques with deep learning paradigms [65]. Additionally, MaskFusion combined semantic instance segmentation with dense mapping, enabling object-level awareness in SLAM systems [66].

The year 2019 introduced generalized SLAM systems like UcoSLAM, OpenVSLAM, and SLAMANTIC, which supported multiple input modalities including mono, stereo, and RGB-D [67][68][69]. These frameworks emphasized modularity, scalability, and reusability, paving the way for flexible deployments in multi-sensor configurations. Simultaneously, RE-SLAM, SPM-SLAM, and BAD SLAM explored better loop closure techniques, keyframe management, and robustness under challenging illumination and texture conditions [70][71][72].

A wave of fusion-based systems emerged in 2019 and 2020, including EM-Fusion, DMS-SLAM, and DDL-SLAM, which aimed to fuse depth and color cues through probabilistic and optimization-driven pipelines [74][75]. FlowFusion, PSPNet-SLAM, and SaD-SLAM utilized optical flow and deep learning-based segmentation to enhance temporal consistency in fast-moving scenes [76][77][78]. Moreover, IV-SLAM and DM-SLAM extended these capabilities with multi-view consistency and dense depth propagation [79][80].

In 2021 and 2022, the emphasis shifted toward dynamic SLAM environments, as demonstrated by systems like OFM-SLAM, GSOM-BSLAM, and Structure PLP-SLAM, which integrated real-time semantic mapping and predictive modeling of dynamic elements [81][82][83]. VOLDOR-SLAM and DynNetSLAM incorporated multi-branch neural networks for end-to-end depth estimation and pose prediction, effectively reducing dependency on handcrafted pipelines [84][85]. Additionally, Edge-SLAM brought attention to computational constraints, adapting RGB-D SLAM pipelines for edge devices by leveraging lightweight networks and compressed representations [86].

Recent approaches such as DeepLabv3+SLAM and SP-SLAM integrated semantic segmentation directly into the mapping process, enabling category-level understanding of environments [87][88]. This trend is furthered by hybrid methods like CG-SLAM and RGBDS-SLAM (2024–2025), which fuse cognitive perception models with conventional SLAM to support task-specific intelligence, such as scene classification and goal-directed navigation [89][90].

Notably, these RGB-D SLAM systems rely on benchmark datasets and real-time performance evaluation. Datasets like TUM RGB-D, ICL-NUIM, and ScanNet have been instrumental in validating performance across different benchmarks, offering standardized trajectories, dynamic scenes, and semantic annotations [91]. While these advancements have brought SLAM closer to robust deployment, issues related to sensor noise, occlusion, generalization in unseen environments, and energy efficiency still remain open challenges [92-94].

Overall, RGB-D SLAM has grown from simple depth-assisted mapping to rich, semantically-aware systems that integrate deep learning, object recognition, and multimodal fusion. This trajectory underscores the increasing convergence of geometric modeling and machine learning in the pursuit of highly adaptable and context-aware visual SLAM solutions.

VIII. Cross-Paradigm Analysis and Discussion

The evolution of Visual SLAM has produced a diverse array of systems across monocular, stereo, and RGB-D paradigms, each with specific strengths and limitations. A comparative analysis of these approaches reveals fundamental trade-offs in scalability, accuracy, computational complexity, and adaptability to dynamic environments.

Monocular SLAM remains the most lightweight and cost-effective option, requiring only a single camera for localization and mapping. However, it suffers from inherent scale ambiguity and limited depth perception, which restricts its application in scenarios demanding metric-scale accuracy [24]. Despite advancements like ORB-SLAM and DSO, monocular methods still rely heavily on loop closure and motion modeling to mitigate accumulated drift [29][32].

In contrast, Stereo SLAM inherently resolves the scale issue by estimating depth via triangulation. It provides a reliable middle ground between simplicity and depth accuracy, making it suitable for mobile robotics and drone navigation [42]. Systems such as ProSLAM and TwistSLAM++ demonstrate improved performance in fast-motion and low-texture environments [44][47]. Nevertheless, stereo systems require careful camera calibration and synchronization, which can introduce additional hardware constraints.

RGB-D SLAM, leveraging dense depth data from sensors like structured-light or time-of-flight cameras, excels in indoor mapping and 3D reconstruction [51][54]. These systems are particularly effective in dynamic and low-texture scenes due to the richness of depth information. However, they are limited by the operating range and sensitivity to lighting conditions, especially in outdoor settings [62]. Moreover, the high data throughput from RGB-D sensors increases computational demands, necessitating efficient data fusion and real-time optimization strategies [73][78].

Hybrid approaches integrating visual-inertial odometry, semantic segmentation, and learning-based modules are emerging across all three paradigms. For instance, visual-inertial extensions to ORB-SLAM and deep semantic mapping in RGB-D systems such as DynNetSLAM demonstrate that cross-domain fusion enhances robustness and scene understanding [30][85]. These trends reflect a shift toward generalizable SLAM systems capable of adapting to diverse environmental conditions and sensory inputs.

Despite these advances, unresolved challenges persist. Dynamic object handling, long-term consistency, real-time performance on edge devices, and scene generalization remain key bottlenecks. The convergence of geometry-based and data-driven methods is likely to play a pivotal role in addressing these issues and shaping the next generation of visual SLAM technologies.

IX. Future Directions

The field of Visual SLAM continues to experience rapid progress, yet numerous challenges remain unsolved, paving the way for future innovations. As the demand for intelligent autonomous systems rises, SLAM must evolve beyond static, structured environments to handle complex, unstructured, and dynamic real-world conditions. Several key directions are anticipated to define the next generation of SLAM systems.

One of the most pressing challenges is the robust handling of dynamic environments, where moving objects introduce inconsistencies in localization and mapping. While state-of-the-art methods such as DynaSLAM and DynNetSLAM have introduced dynamic object filtering using motion segmentation and learning-based modules, real-time robustness in highly dynamic scenarios remains limited [64][85]. Future systems must incorporate adaptive scene understanding and prediction mechanisms to account for object trajectories, occlusions, and interactions, thereby enhancing robustness under motion and clutter.

Another critical direction involves lifelong and continual SLAM, which allows systems to incrementally learn from their environments over extended periods. Current methods tend to operate in episodic modes, with limited memory and generalization across time or domains. Incorporating memory-augmented learning, self-supervision, and lifelong adaptation strategies would enable SLAM systems to refine their internal models continually and recover from localization failures or environmental changes. Techniques applied in DeepVO and other learning-based systems offer a foundation for such progress [36].

Semantic SLAM is also emerging as a transformative direction, where the integration of object-level understanding and scene semantics augments traditional geometric mapping. Systems such as MaskFusion, DeepLabv3+SLAM, and CG-SLAM demonstrate that embedding semantics enables higher-level reasoning, object manipulation, and contextual awareness [66][87][90]. Future research is expected to integrate multi-task learning pipelines that jointly optimize for pose estimation, semantic segmentation, and instance recognition. Such integration would be vital in service robotics, AR/VR, and autonomous navigation tasks where understanding scene content is as critical as knowing spatial geometry.

The development of resource-aware and edge-compatible SLAM algorithms is gaining significance as SLAM expands into mobile platforms, drones, and embedded robotics. High-performance SLAM often requires intensive computation and memory, making deployment on constrained devices difficult. Solutions like Edge-SLAM and optimized variants of SLAMANTIC and ORB-SLAM address these issues by incorporating lightweight architectures, neural compression, and on-device learning [86][69][30]. Future directions may involve further advances in model pruning, quantization, and hardware-aware design for real-time, low-power SLAM inference.

Multimodal SLAM represents another growing frontier. The fusion of multiple sensor inputs—including visual, inertial, depth, thermal, and even auditory data—enables robust mapping under diverse environmental conditions. Systems like ORB-SLAM3 and VOLDOR-SLAM illustrate how integrating inertial measurements and depth cues improves accuracy and consistency, especially in low-texture or fast-motion scenarios [30][84]. Expanding such systems to include semantic and audio signals could further enhance context-awareness, particularly in human-robot interaction and search-and-rescue missions.

In addition, learning-based SLAM continues to evolve. While geometry-based methods have dominated due to their interpretability and reliability, deep learning models are increasingly demonstrating their potential in monocular depth prediction, motion estimation, and loop closure detection. Hybrid pipelines that combine data-driven features with geometric consistency are likely to define future SLAM architectures. DMS-SLAM, CodeSLAM, and PSPNet-SLAM exemplify this convergence [74][61][77].

Finally, neural implicit representations, such as neural radiance fields (NeRF), are expected to redefine how SLAM systems model and store environments. These methods replace discrete maps with continuous, differentiable scene representations that offer high-fidelity reconstructions and efficient storage. Their integration with SLAM could lead to novel systems that unify localization, mapping, and scene rendering into a single framework, unlocking new possibilities in mixed reality and digital twin technologies.

In summary, the future of Visual SLAM lies in its ability to become more adaptive, intelligent, semantic, and efficient, leveraging cross-disciplinary advances in machine learning, sensor fusion, and cognitive modeling. These advancements will shape the next generation of autonomous systems capable of understanding, interacting with, and navigating the real world in an increasingly human-like manner.

X. Conclusion

This review has presented a structured and comprehensive analysis of the evolution, methodologies, and challenges in Visual SLAM across three primary paradigms: monocular, stereo, and RGB-D systems. Monocular SLAM remains valuable for its simplicity and affordability, yet it continues to face challenges related to scale ambiguity and robustness under complex motion. Stereo SLAM offers a balance between hardware complexity and depth accuracy, making it suitable for a wide range of robotic and autonomous applications. RGB-D SLAM systems have significantly advanced scene understanding and dense mapping capabilities, particularly in indoor and structured environments, though they remain constrained by sensor limitations and computational demands. Across all paradigms, recent trends reveal a growing convergence of traditional geometric methods with learning-based models, enabling SLAM systems to operate more robustly in dynamic and perceptually challenging scenarios. The incorporation of semantic information, multi-modal sensor fusion, and real-time optimization strategies has further extended the operational scope of SLAM technologies. Despite these advances, persistent issues such as adaptability in dynamic environments, efficient computation on edge devices, and generalization across unseen domains continue to hinder full autonomy. The future of Visual SLAM lies in the development of lifelong, context-aware systems capable of learning from experience, reasoning semantically, and operating reliably under uncertainty. By integrating insights from computer vision, robotics, and machine learning, the next generation of SLAM algorithms will move beyond spatial localization to support intelligent interaction and decision-making in real-world environments. This review serves as a foundation for understanding current capabilities and identifying key research directions in the evolving landscape of Visual SLAM.

Acknowledgment

This work was supported by the Science Committee of the Ministry of Higher Education and Science of the Republic of Kazakhstan within the framework of grant AP19679910 “Development of an experimental model of an autonomous unmanned mobile robot with off-road capability”, International Engineering Technological University, Almaty.

References

[1] T. Schneider et al., "Maplab: An Open Framework for Research in Visual-Inertial Mapping and Localization," in IEEE Robotics and Automation Letters, vol. 3, no. 3, pp. 1418-1425, July 2018, doi: 10.1109/LRA.2018.2800113.

[2] X. Zuo, P. Geneva, W. Lee, Y. Liu and G. Huang, "LIC-Fusion: LiDAR-Inertial-Camera Odometry," 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 2019, pp. 5848-5854, doi: 10.1109/IROS40897.2019.8967746.

[3] Y. -S. Shin and A. Kim, "Sparse Depth Enhanced Direct Thermal-Infrared SLAM Beyond the Visible Spectrum," in IEEE Robotics and Automation Letters, vol. 4, no. 3, pp. 2918-2925, July 2019, doi: 10.1109/LRA.2019.2923381

[4] T. Shan, B. Englot, C. Ratti and D. Rus, "LVI-SAM: Tightly-coupled Lidar-Visual-Inertial Odometry via Smoothing and Mapping," 2021

IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 2021, pp. 5692-5698, doi: 10.1109/ICRA48506.2021.9561996.

[5] Le, V.-H., Do, H.-S., Phan, V.-N., & Te, T.-H. (2024). TQU-SLAM Benchmark Feature-based Dataset for Building Monocular VO. Engineering, Technology & Applied Science Research, 14(4), 15330–15337. https://doi.org/10.48084/etasr.7611

[6] G. Ge, Y. Zhang, W. Wang, Q. Jiang, L. Hu, and Y. Wang, ''Text-mcl: autonomous mobile robot localization in similar environment using text-level semantic information,'' Machines, vol. 10, no. 3, Feb. 2022, 10.3390/machines10030169G. O. Young, "Synthetic structure of industrial plastics," in Plastics, 2nd ed., vol. 3, J. Peters, Ed. New York, NY, USA: McGraw-Hill, 1964, pp.15–64.

[7] Altayeva, A., Omarov, B., Suleimenov, Z., & Im Cho, Y. (2017, June). Application of multi-agent control systems in energy-efficient intelligent building. In 2017 Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems (IFSA-SCIS) (pp. 1-5). IEEE.

[8] S. Martínez-Díaz, "3D Distance Measurement from a Camera to a Mobile Vehicle Using Monocular Vision," Journal of Sensors, vol. 2021, 5526931, 2021.

[9] S. Hensel, M. B. Marinov, M. Obert, "3D LiDAR Based SLAM System Evaluation with Low-Cost Real-Time Kinematics GPS Solution" Remote Sensing, Computation 10, no. 9, 154, 2022.

[10] Igor V. Malyk, Yevhen Kyrychenko, Mykola Gorbatenko, Taras Lukashiv, "Data Optimization through Compression Methods Using Information Technology", International Journal of Information Technology and Computer Science(IJITCS), Vol.17, No.5, pp.84-99, 2025. DOI:10.5815/ijitcs.2025.05.07

[11] Q. Lu, Y. Pan, L. Hu, and J. He, "A Method for Reconstructing Background from RGB-D SLAM in Indoor Dynamic Environments," Sensors, vol. 23, no. 7, pp. 3529, Mar. 2023.

[12] F. Menna, A. Torresani, R. Battisti, E. Nocerino, and F. Remondino, "A Modular and Low-Cost Portable Vslam System for Real-Time 3d Mapping: From Indoor and Outdoor Spaces to Underwater Environments," International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLVIII-2/W1-2022, pp. 153-162, 2022.

[13] I. Kalisperakis, T. Mandilaras, A. El Saer, P. Stamatopoulou, C. Stentoumis, S. Bourou, and L. Grammatikopoulos, "A Modular Mobile Mapping Platform for Complex Indoor and Outdoor Environments," International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLIII-B1-2020, pp. 243-250, 2020.

[14] Z. Xie, Z. Li, Y. Zhang, J. Zhang, F. Liu, and W. Chen, "A Multi-Sensory Guidance System for the Visually Impaired Using YOLO and ORB-SLAM," Information, vol. 13, no. 7, pp. 343, Jul. 2022.

[15] Omarov, B., Suliman, A., Kushibar, K. Face recognition using artificial neural networks in parallel architecture. Journal of Theoretical and Applied Information Technology 91 (2), pp. 238-248. Open Access.

[16] Alexander Pavlov, Kateryna Lishchuk, Oleg Melnikov, Mykyta Kyselov, Cennuo Hu, "Mathematics and Software for Coordinated Planning Using Aggregated Linear Volume-time Models of Discrete Manufacturing Systems", International Journal of Information Technology and Computer Science(IJITCS), Vol.17, No.4, pp.1-15, 2025. DOI:10.5815/ijitcs.2025.04.01

[17] C. Mai, H. Chen, L. Zeng, Z. Li, G. Liu, Z. Qiao, Y. Qu, L. Li, and L. Li, "A Smart Cane Based on 2D LiDAR and RGB-D Camera Sensor-Realizing Navigation and Obstacle Recognition," Sensors, vol. 24, no. 3, pp. 870, Jan. 2024.

[18] A. Altayeva, B. Omarov, H.C. Jeong, Y.I. Cho. Multi-step face recognition for improving face detection and recognition rate. Far East Journal of Electronics and Communications 16(3), pp. 471-491.

[19] D. Merkle and A. Reiterer, "Automated Method for SLAM Evaluation in GNSS-Denied Areas" Remote Sensing, vol. 15, no. 21, pp. 5141, 2023.

[20] J. Sun, J. Zhao, X. Hu, H. Gao, and J. Yu, "Autonomous Navigation System of Indoor Mobile Robots Using 2D Lidar", Mathematics, vol. 11, no. 6, pp. 1455, 2023.

[21] Y. Wang, S. Zhang, and J. Wang, "Ceiling-View Semi-Direct Monocular Visual Odometry with Planar Constraint", Remote Sens., vol. 14, no. 21, pp. 5447, 2022.

[22] V. Di Pietra, N. Grasso, M. Piras, and P. Dabove, "Characterization of a Mobile Mapping System for Seamless Navigation", International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLIII-B1-2020, pp. 227–234, 2020.

[23] Omarov, B., & Altayeva, A. (2018, January). Towards intelligent IoT smart city platform based on OneM2M guideline: smart grid case study. In 2018 IEEE International Conference on Big Data and Smart Computing (BigComp) (pp. 701-704). IEEE.

[24] D. Bolkas, M. O'Banion, and C. J. Belleman, "Combination of TLS and SLAM Lidar for Levee Monitoring", ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. V-3-2022, pp. 641–647, 2022.

[25] L. He, Z. Jin, and Z. Gao, "De-Skewing LiDAR Scan for Refinement of Local Mapping", Sensors, vol. 20, no. 7, pp. 1846, 2020.

[26] L. Moradi, M. Saadatseresht, and P. Shokrzadeh, "Development of a Voxel-Based Local Plane Fitting for Multi-Scale Registration of Sequential MLS Point Clouds," ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. X-4/W1-2022, pp. 523–530, 2023.

[27] L. Chen, Z. Chen, and Z. Ji, "Expectation–Maximization-Based Simultaneous Localization and Mapping for Millimeter-Wave Communication Systems," Sensors, vol. 22, no. 18, 6941, 2022.

[28] T. Bodrumlu and F. Caliskan, "Indoor Position Estimation Using Ultrasonic Beacon Sensors and Extended Kalman Filter", Engineering Proceedings, vol. 27, no. 1, pp. 16, 2022. Presented at the 9th International Electronic Conference on Sensors and Applications, 1–15 November 2022.

[29] K. Xiao, W. Yu, W. Liu, F. Qu, and Z. Ma, "High-Precision SLAM Based on the Tight Coupling of Dual-Lidar Inertial Odometry for Multi-Scene Applications", Applied Sciences, vol. 12, no. 3, 939, 2022.

[30] A. Basiri, V. Mariani, and L. Glielmo, "Improving Visual SLAM by Combining SVO and ORB-SLAM2 with a Complementary Filter to Enhance Indoor Mini-Drone Localization under Varying Conditions", Drones, vol. 7, no. 6, 404, 2022.

[31] Al Noman, M. A., Zhai, L., Almukhtar, F. H., Rahaman, M. F., Omarov, B., Ray, S., ... & Wang, C. (2023). A computer vision-based lane detection technique using gradient threshold and hue-lightness-saturation value for an autonomous vehicle. International Journal of Electrical and Computer Engineering, 13(1), 347. DOI: 10.11591/ijece.v13i1.pp347-357.

[32] Y. Zhai, B. Lu, W. Li, J. Xu, and S. Ma, "JD-SLAM: Joint camera pose estimation and moving object segmentation for simultaneous localization and mapping in dynamic scenes", International Journal of Advanced Robotic Systems, vol. 18, no. 1, pp. 1–12, 2021.

[33] Y. Sun, J. Hu, J. Yun, Y. Liu, D. Bai, X. Liu, G. Zhao, G. Jiang, J. Kong, and B. Chen, "Multi-Objective Location and Mapping Based on Deep Learning and Visual SLAM," Sensors, vol. 22, no. 19, 7576, 2022.

[34] Z. Chen, A. Xu, X. Sui, Y. Hao, C. Zhang, and Z. Shi, "NLOS Identification- and Correction-Focused Fusion of UWB and LiDAR-SLAM Based on Factor Graph Optimization for High-Precision Positioning with Reduced Drift", Remote Sensing, vol. 14, no. 17, 4258, 2022.

[35] H. Wu, W. Wu, X. Qi, C. Wu, L. An, and R. Zhong, "Planar Constraint-Assisted LiDAR SLAM Algorithm Based on Manhattan World Assumption", Remote Sensing, vol. 15, no. 1, 15, 2023.

[36] Z. Zhao, T. Song, B. Xing, Y. Lei, and Z. Wang, "PLI-VINS: Visual-Inertial SLAM Based on Point-Line Feature Fusion in Indoor Environment", Sensors, vol. 22, no. 14, 5457, 2022.

[37] G. Chen and L. Hong, "Research on Environment Perception System of Quadruped Robots Based on LiDAR and Vision", Drones, vol. 7, no. 5, 329, 2023.

[38] J. Geng, A. A. Mokhtarzadeh, H. Fei, H. Gao, N. Ding, R. Kong, and J. Zhou, "Robot positioning and navigation technology is based on Integration of the Global Navigation Satellite System and real-time kinematics", Journal of Physics: Conference Series, vol. 2467, no. 1, 012027, 2023.

[39] C. Xu, Z. Liu, and Z. Li, "Robust Visual-Inertial Navigation System for Low Precision Sensors under Indoor and Outdoor Environments", Remote Sensing, vol. 13, no. 4, 772, 2021.

[40] Q. Zhang and C. Li, "Semantic SLAM for mobile robots in dynamic environments based on visual camera sensors", Measurement Science and Technology, vol. 34, no. 8, 085202, 2023.

[41] Y. Xia, J. Cheng, X. Cai, S. Zhang, J. Zhu, and L. Zhu, "SLAM Back-End Optimization Algorithm Based on Vision Fusion IPS", Sensors, vol. 22, no. 23, 9362, 2022.

[42] O. F. Ince and J.-S. Kim, "TIMA SLAM: Tracking Independently and Mapping Altogether for an Uncalibrated Multi-Camera System", Sensors, vol. 21, no. 2, 409, 2021.

[43] C. Theodorou, V. Velisavljevic, and V. Dyo, "Visual SLAM for Dynamic Environments Based on Object Detection and Optical Flow for Dynamic Object Removal", Sensors, vol. 22, no. 19, 7553, Oct. 2022.

[44] S. Hensel, M. B. Marinov, M. Obert, "3D LiDAR Based SLAM System Evaluation with Low-Cost Real-Time Kinematics GPS Solution" Remote Sensing, Computation 10, no. 9, 154, 2022.

[45] H. Guan, C. Qian, T. Wu, X. Hu, F. Duan, and X. Ye, "A Dynamic Scene Vision SLAM Method Incorporating Object Detection and Object Characterization," Sustainability, vol. 15, no. 4, pp. 3048, Feb. 2023.

[46] P. Herbert, J. Wu, Z. Ji, and Y.-K. Lai, "Benchmarking Visual SLAM Methods in Mirror Environments," Computational Visual Media, vol. 10, no. 2, pp. 215-241, Apr. 2024.

[47] A. Elamin, N. Abdelaziz, and A. El-Rabbany, "A GNSS/INS/LiDAR Integration Scheme for UAV-Based Navigation in GNSS-Challenging Environments," Sensors, vol. 22, no. 24, 9908, 2022.

[48] Q. Lu, Y. Pan, L. Hu, and J. He, "A Method for Reconstructing Background from RGB-D SLAM in Indoor Dynamic Environments," Sensors, vol. 23, no. 7, pp. 3529, Mar. 2023.

[49] D. Chen, Q. Yan, Z. Zeng, J. Kang, and J. Zhou, "A Model of Real-Time Pose Estimation Fusing Camera and LiDAR in Simultaneous Localization and Mapping by a Geometric Method," Sensors and Materials, vol. 35, no. 1, pp. 167-181, 2023.

[50] F. Menna, A. Torresani, R. Battisti, E. Nocerino, and F. Remondino, "A Modular and Low-Cost Portable Vslam System for Real-Time 3d Mapping: From Indoor and Outdoor Spaces to Underwater Environments," International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLVIII-2/W1-2022, pp. 153-162, 2022.

[51] I. Kalisperakis, T. Mandilaras, A. El Saer, P. Stamatopoulou, C. Stentoumis, S. Bourou, and L. Grammatikopoulos, "A Modular Mobile Mapping Platform for Complex Indoor and Outdoor Environments," International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLIII-B1-2020, pp. 243-250, 2020.

[52] T. Lu, Y. Liu, Y. Yang, H. Wang, and X. Zhang, "A Monocular Visual Localization Algorithm for Large-Scale Indoor Environments through Matching a Prior Semantic Map" Electronics, vol. 11, no. 20, pp. 3396, 2022.

[53] W. Chen, G. Shang, K. Hu, C. Zhou, X. Wang, G. Fang, and A. Ji, "A Monocular-Visual SLAM System with Semantic and Optical-Flow Fusion for Indoor Dynamic Environments," Micromachines, vol. 13, no. 11, pp. 2006, Nov. 2022.

[54] Z. Xie, Z. Li, Y. Zhang, J. Zhang, F. Liu, and W. Chen, "A Multi-Sensory Guidance System for the Visually Impaired Using YOLO and ORB-SLAM," Information, vol. 13, no. 7, pp. 343, Jul. 2022.

[55] C. Bonfanti, G. Patrucco, S. Perri, G. Sammartano, and A. Spanò, "A New Indoor LiDAR-Based MMS Challenging Complex Architectural Environments," International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLVI-M-1-2021, pp. 79-86, 2021.

[56] T. Alhmiedat, A. M. Marei, W. Messoudi, S. Albelwi, A. Bushnag, Z. Bassfar, F. Alnajjar, and A. O. Elfaki, "A SLAM-Based Localization and Navigation System for Social Robots: The Pepper Robot Case," Machines, vol. 11, no. 2, pp. 158, Jan. 2023.

[57] C. Mai, H. Chen, L. Zeng, Z. Li, G. Liu, Z. Qiao, Y. Qu, L. Li, and L. Li, "A Smart Cane Based on 2D LiDAR and RGB-D Camera Sensor-Realizing Navigation and Obstacle Recognition," Sensors, vol. 24, no. 3, pp. 870, Jan. 2024.

[58] G. Patrucco, G. Sammartano, C. Bonfanti, and A. Spanò, "Assessing Terrestrial MMS 3D Data for Outdoor Multi-Scale Modelling," International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLVIII-1/W1-2023, pp. 371-378, 2023.

[59] W. Hu, K. Zhang, L. Shao, Q. Lin, Y. Hua, and J. Qin, "Clustering Denoising of 2D LiDAR Scanning in Indoor Environment Based on Keyframe Extraction", Sensors, vol. 23, no. 1, pp. 18, 2023.

[60] L. Morelli, F. Ioli, R. Beber, F. Menna, F. Remondino, and A. Vitti, "Colmap-Slam: A Framework for Visual Odometry", International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLVIII-1/W1-2023, pp. 317–324, 2023.

[61] D. Bolkas, M. O'Banion, and C. J. Belleman, "Combination of TLS and SLAM Lidar for Levee Monitoring", ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. V-3-2022, pp. 641–647, 2022.

[62] L. He, Z. Jin, and Z. Gao, "De-Skewing LiDAR Scan for Refinement of Local Mapping", Sensors, vol. 20, no. 7, pp. 1846, 2020.

[63] K. Jegadeeswari, R. Rathipriya, "Green AI Practices in Multi-objective Hyperparameter Optimization for Sustainable Machine Learning", International Journal of Information Technology and Computer Science(IJITCS), Vol.17, No.2, pp.1-9, 2025. DOI:10.5815/ijitcs.2025.02.01

[64] L. Chen, Z. Chen, and Z. Ji, "Expectation–Maximization-Based Simultaneous Localization and Mapping for Millimeter-Wave Communication Systems," Sensors, vol. 22, no. 18, 6941, 2022.

[65] D. Damodaran, S. Mozaffari, S. Alirezaee, and M. J. Ahamed, "Experimental Analysis of the Behavior of Mirror-like Objects in LiDAR-Based Robot Navigation", Applied Sciences, vol. 13, no. 5, pp. 2908, 2023.

[66] Ł. Sobczak, K. Filus, J. Domańska, and A. Domański, "Finding the best hardware configuration for 2D SLAM in indoor environments via simulation based on Google Cartographer", Scientific Reports, vol. 12, pp. 18815, 2022.

[67] N. Li, L. Guan, Y. Gao, S. Du, M. Wu, X. Guang, and X. Cong, "Indoor and Outdoor Low-Cost Seamless Integrated Navigation System Based on the Integration of INS/GNSS/LIDAR System", Remote Sensing, vol. 12, no. 19, pp. 3271, 2020.

[68] T. Bodrumlu and F. Caliskan, "Indoor Position Estimation Using Ultrasonic Beacon Sensors and Extended Kalman Filter", Engineering Proceedings, vol. 27, no. 1, pp. 16, 2022. Presented at the 9th International Electronic Conference on Sensors and Applications, 1–15 November 2022.

[69] Wang, K., Guo, J., Chen, K., & Lu, J. (2025). An in-depth examination of SLAM methods: Challenges, advancements, and applications in complex scenes for autonomous driving. IEEE Transactions on Intelligent Transportation Systems.

[70] Wang, K., Zhao, G., & Lu, J. (2024). A deep analysis of visual SLAM methods for highly automated and autonomous vehicles in complex urban environment. IEEE Transactions on Intelligent Transportation Systems, 25(9), 10524-10541.

[71] Yan, C., Qu, D., Xu, D., Zhao, B., Wang, Z., Wang, D., & Li, X. (2024). Gs-slam: Dense visual slam with 3d gaussian splatting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 19595-19604).

[72] Zhu, S., Wang, G., Blum, H., Liu, J., Song, L., Pollefeys, M., & Wang, H. (2024). Sni-slam: Semantic neural implicit slam. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 21167-21177).

[73] Altayeva, A. B., Omarov, B. S., Aitmagambetov, A. Z., Kendzhaeva, B. B., & Burkitbayeva, M. A. (2014). Modeling and exploring base station characteristics of LTE mobile networks. Life Science Journal, 11(6), 227-233.

[74] Lin, Z., Zhang, Q., Tian, Z., Yu, P., & Lan, J. (2024). DPL-SLAM: enhancing dynamic point-line SLAM through dense semantic methods. IEEE Sensors Journal, 24(9), 14596-14607.

[75] Zhang, H., Peng, J., & Yang, Q. (2024). PR-SLAM: parallel real-time dynamic SLAM method based on semantic segmentation. IEEE Access, 12, 36498-36514.

[76] He, L., Li, S., Qiu, J., & Zhang, C. (2024). DIO-SLAM: a dynamic RGB-D slam method combining instance segmentation and optical flow. Sensors (Basel, Switzerland), 24(18), 5929.

[77] Pan, H., Liu, D., Ren, J., Huang, T., & Yang, H. (2024). LiDAR-imu tightly-coupled SLAM method based on IEKF and loop closure detection. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17, 6986-7001.

[78] Islam, Q. U., Ibrahim, H., Chin, P. K., Lim, K., Abdullah, M. Z., & Khozaei, F. (2024). ARD-SLAM: Accurate and robust dynamic SLAM using dynamic object identification and improved multi-view geometrical approaches. Displays, 82, 102654.

[79] Rastorgueva-Foi, E., Kaltiokallio, O., Ge, Y., Turunen, M., Talvitie, J., Tan, B., ... & Valkama, M. (2024). Millimeter-wave radio SLAM: End-to-end processing methods and experimental validation. IEEE Journal on Selected Areas in Communications, 42(9), 2550-2567.

[80] Li, Y., Shen, H., Fu, Y., & Wang, K. (2024). A method of dense point cloud SLAM based on improved YOLOV8 and fused with ORB-SLAM3 to cope with dynamic environments. Expert Systems with Applications, 255, 124918.

[81] Jiang, J., Zhang, T., & Li, K. (2025). LiDAR-based 3D SLAM for autonomous navigation in stacked cage farming houses: An evaluation. Computers and Electronics in Agriculture, 230, 109885.

[82] Wang, Y., Ren, B., Zhang, X., Wang, P., Wang, C., Song, R., ... & Meng, M. Q. H. (2025). ROLO‐SLAM: rotation‐optimized LiDAR‐only SLAM in uneven terrain with ground vehicle. Journal of Field Robotics, 42(3), 880-902.

[83] Liang, M., Leitinger, E., & Meyer, F. (2025). Direct multipath-based SLAM. IEEE Transactions on Signal Processing.

[84] Tapu Biswas, Farhan Sadik Ferdous, Zinniya Taffannum Pritee, Akinul Islam Jony, "ScrumSpiral: An Improved Hybrid Software Development Model", International Journal of Information Technology and Computer Science(IJITCS), Vol.16, No.2, pp.57-65, 2024. DOI:10.5815/ijitcs.2024.02.05.

[85] Jiang, H., Xu, Y., Li, K., Feng, J., & Zhang, L. (2024). Rodyn-slam: Robust dynamic dense rgb-d slam with neural radiance fields. IEEE Robotics and Automation Letters.

[86] Huang, H., Li, L., Cheng, H., & Yeung, S. K. (2024). Photo-slam: Real-time simultaneous localization and photorealistic mapping for monocular stereo and rgb-d cameras. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 21584-21593).

[87] Zheng, J., Zhu, Z., Bieri, V., Pollefeys, M., Peng, S., & Armeni, I. (2025). Wildgs-slam: Monocular gaussian splatting slam in dynamic environments. In Proceedings of the Computer Vision and Pattern Recognition Conference (pp. 11461-11471).

[88] Liu, X., Wen, S., Zhao, J., Qiu, T. Z., & Zhang, H. (2024). Edge-assisted multi-robot visual-inertial SLAM with efficient communication. IEEE Transactions on Automation Science and Engineering, 22, 2186-2198.

[89] Yang, Y., Pan, M., Tang, D., Wang, T., Yue, Y., Liu, T., & Fu, M. (2024). Mcov-slam: A multicamera omnidirectional visual slam system. IEEE/ASME Transactions on Mechatronics, 29(5), 3556-3567.

[90] Xin, Z., Yue, Y., Zhang, L., & Wu, C. (2024, May). Hero-slam: Hybrid enhanced robust optimization of neural slam. In 2024 IEEE International Conference on Robotics and Automation (ICRA) (pp. 8610-8616). IEEE.

[91] Yang, L., Ye, J., Zhang, Y., Wang, L., & Qiu, C. (2024). A semantic SLAM-based method for navigation and landing of UAVs in indoor environments. Knowledge-Based Systems, 293, 111693.

[92] Wen, S., Tao, S., Liu, X., Babiarz, A., & Yu, F. R. (2024). CD-SLAM: A real-time stereo visual–inertial SLAM for complex dynamic environments with semantic and geometric information. IEEE Transactions on Instrumentation and Measurement, 73, 1-8.

[93] Muhammad Amirul Asyraaf Roslan, Haryani Haron, "Designing the Smart Shopping Cart Mobile Application (SmartCart) Using Mobile Application Development Life Cycle", International Journal of Information Technology and Computer Science(IJITCS), Vol.16, No.4, pp.66-81, 2024. DOI:10.5815/ijitcs.2024.04.05.

[94] Wang, W., Wang, C., Liu, J., Su, X., Luo, B., & Zhang, C. (2024). HVL-SLAM: Hybrid vision and LiDAR fusion for SLAM. IEEE Transactions on Geoscience and Remote Sensing.