Recent Advances in Visual SLAM: Taxonomy, Comparative Analysis, and Open Challenges

Aidos Ibrayev
International Engineering Technological University, Kazakhstan | Al-Farabi Kazakh National University, Kazakhstan
aydos.ybraev@kaznu.edu.kz

Batyrkhan Omarov
International Information Technology University, Kazakhstan | Al-Farabi Kazakh National University, Kazakhstan | Kh. Dosmukhamedov Atyrau State University, Kazakhstan
batyahan@gmail.com (corresponding author)

Received: 2 July 2025 | Revised: 23 July 2025, 2 August 2025, and 16 August 2025 | Accepted: 20 August 2025

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | *DOI: https://doi.org/10.48084/etasr.13116*

ABSTRACT

Visual Simultaneous Localization and Mapping (SLAM) has become a cornerstone in the development of intelligent systems capable of perceiving and interacting with their environment in real time. This survey presents a comprehensive review of recent advances in visual SLAM algorithms, with a focus on their classification, performance characteristics, and application domains. This study categorizes existing methods into monocular, stereo, RGB-D, and multi-sensor/hybrid approaches, analyzing key contributions such as ORB-SLAM, DSO, ElasticFusion, and VINS-Mono. Each class is evaluated in terms of accuracy, robustness, and computational efficiency while highlighting the trade-offs associated with different sensor modalities. Additionally, this study explores cross-modal and deep learning-based hybrid SLAM systems, which incorporate semantic understanding, motion segmentation, and sensor fusion to enhance performance in complex and dynamic environments. Application areas, including robotics, augmented/virtual reality, 3D mapping, and wearable technologies, are discussed to underscore the practical relevance of visual SLAM. Finally, the survey outlines the main challenges and future directions, including lifelong mapping, real-time performance on edge devices, semantic integration, and the emergence of SLAM 2.0 systems. This work aims to serve as a resource for researchers and practitioners seeking to understand the state of the art and guide future innovation in the field of visual SLAM.

Keywords-Visual SLAM (VSLAM); monocular SLAM; stereo vision; RGB-D mapping; sensor fusion; deep learning; semantic SLAM; real-time localization; 3D reconstruction; autonomous navigation

I. INTRODUCTION

Simultaneous Localization and Mapping (SLAM) has emerged as a foundational technology for autonomous navigation, allowing systems to concurrently estimate their position while constructing a map of an unknown environment. Among its various modalities, Visual SLAM (VSLAM) has gained significant attention due to its cost-effectiveness, availability of consumer-grade cameras, and compatibility with human perception systems [1]. The ability to perform real-time visual understanding of environments has facilitated its widespread use in robotics, Augmented Reality (AR), Virtual Reality (VR), and autonomous driving [2]. VSLAM systems have evolved from geometric and feature-based techniques to deep learning-enhanced pipelines, leading to substantial improvements in accuracy, robustness, and scalability [3].

VSLAM algorithms are categorized into monocular, stereo, and RGB-D-based systems, with each modality having its own trade-offs. Monocular SLAM systems are lightweight and suitable for embedded platforms but suffer from scale ambiguity and drift [4]. Stereo SLAM leverages depth from disparity to overcome scale estimation issues, improving robustness and localization precision. RGB-D SLAM, which uses depth sensors, provides dense 3D information and enhances mapping quality, particularly in textureless or poorly lit environments [5]. In parallel, the incorporation of inertial measurements and semantic understanding has advanced the performance of vSLAM systems in dynamic and cluttered scenarios [6].

Recent research trends reveal a shift toward learning-based SLAM frameworks. These approaches integrate deep neural networks for tasks such as depth estimation, loop closure detection, and pose regression, yielding more resilient and adaptable systems [7]. Nevertheless, challenges such as real-time processing, dynamic object handling, and long-term autonomy remain unsolved. Moreover, the diversity of datasets and evaluation metrics complicates fair benchmarking and reproducibility [8]. Therefore, a comprehensive review and

analysis of VSLAM algorithms is crucial to understand current limitations, identify research gaps, and guide future developments in this rapidly evolving field.

II. FUNDAMENTALS OF VSLAM

VSLAM is the process by which an autonomous agent estimates its position while incrementally constructing a map of the environment using visual inputs. The SLAM problem can be probabilistically formulated as the joint posterior estimation of the robot's trajectory $x_{1:t}$ and the environment map m, given the sequence of sensor observations $z_{1:t}$ and control inputs $u_{1:t}$:

$$p(x_{1:t}, m | z_{1:t}, u_{1;t}) \quad (1)$$

This recursive Bayesian formulation is typically decomposed using factorization methods to enable real-time estimation in practical systems.

Pose estimation refers to the recovery of the camera's position and orientation in a global frame, often expressed as a rigid body transformation in the Special Euclidean (SE) group $SE(3)$ [9]. A pose $T \in SE(3)$ is defined by:

$$T = \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \quad (2)$$

where $R \in SE(3)$ is a rotation matrix and $t \in R^3$ is a translation vector. The transformation maps 3D world points X_w into camera coordinates x_c as:

$$x_c = RX_w + t \quad (3)$$

State Representation in SLAM typically includes both the agent's pose and the 3D map landmarks:

$$x = \{x_t, l_1, l_2, \ldots, l_N\} \quad (4)$$

where $l_i \in R^3$ are 3D landmark coordinates. These variables are estimated jointly for accurate mapping and localization.

Optimization techniques include Bundle Adjustment (BA), which minimizes the reprojection error across all keyframes and landmarks:

$$\min_{x,l} \sum_{i,j} \left\| z_{ij} - \pi(T_i, l_j) \right\|^2 \quad (5)$$

where $\pi(\cdot)$ is the camera projection function and z_{ij} is the observed image location of landmark j in frame i. In graph-based SLAM, a pose graph is constructed with vertices representing poses and edges encoding spatial constraints, leading to the following minimization problem:

$$\min_{x} \sum_{i,j} \left\| e_{ij}(x_i, x_j) \right\|^2_{\sum_{ij}} \quad (6)$$

where e_{ij} is the error function between poses and $\sum_{ij}$ is the covariance.

The VSLAM pipeline typically follows these stages:

1. Feature extraction using detectors such as ORB, FAST, or SIFT.
2. Feature matching across frames using descriptors and nearest-neighbor search.
3. Motion estimation, often achieved by solving the Perspective-n-Point (PnP) problem or through direct methods.
4. Loop closure detection using Bag-of-Words (BoW) or neural encodings to reduce drift.
5. Map management, involving keyframe insertion, culling, and landmark triangulation.

Sensor modalities play a critical role in SLAM accuracy and robustness.

- Monocular SLAM uses a single camera and requires scale estimation from motion.
- Stereo SLAM uses two synchronized cameras, leveraging disparity $d = x_L - x_R$ to infer depth:

$$Z = \frac{f \cdot B}{d} \quad (7)$$

where f is the focal length and B is the baseline.

RGB-D SLAM combines an RGB image and depth map, where each pixel (u, v) directly provides a 3D point $P = (X, Y, Z)$ via camera intrinsics.

Each of these components is fundamental to the success and applicability of VSLAM in real-world scenarios.

III. TAXONOMY OF VSLAM ALGORITHMS

VSLAM systems can be broadly categorized based on the type of visual input utilized for localization and mapping. As shown in Figure 1, the taxonomy is generally divided into three primary categories: Monocular SLAM, Stereo SLAM, and RGB-D SLAM. This classification reflects the underlying sensor modalities and the nature of the spatial information they provide. Each category has given rise to a range of algorithms with distinct characteristics, advantages, and limitations, enabling their deployment across various application domains and environments.

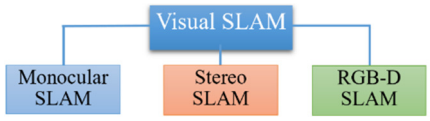

Fig. 1. Taxonomy of VSLAM algorithms based on sensor modalities.

Monocular SLAM relies on a single camera to estimate both motion and 3D structure, leveraging geometric cues and camera motion to infer depth information. Although it offers a compact and cost-effective solution, it inherently suffers from scale ambiguity and sensitivity to initialization [10]. Despite these challenges, it has evolved significantly with the development of high-performance algorithms such as ORB-SLAM and DSO.

In contrast, Stereo SLAM employs two spatially separated cameras to triangulate depth from image disparity, thereby resolving the scale ambiguity present in monocular systems [11]. The additional depth cue improves robustness and localization accuracy, particularly in low-texture or dynamic scenes.

RGB-D SLAM integrates color (RGB) data with depth information obtained from structured-light or time-of-flight sensors. This modality provides dense depth maps directly, significantly simplifying 3D reconstruction and enhancing performance in poorly textured or dark environments [12]. Algorithms in this class, such as ElasticFusion and DynaSLAM, have demonstrated strong capabilities in dense mapping and real-time performance.

This section provides a detailed overview of each category, examining key algorithms, methodological innovations, and performance trade-offs. Through this taxonomy, we aim to offer a comprehensive understanding of the landscape of VSLAM algorithms and highlight the technical trajectories that define their evolution.

A. Monocular VSLAM

Monocular vSLAM systems use a single camera to estimate both the motion of the sensor and the structure of the environment. As shown in Figure 2, the development of monocular SLAM has evolved through several generations of algorithms, beginning with early approaches such as MonoSLAM [13] and PTAM [14], which introduced key concepts of recursive filtering and parallel tracking and mapping, respectively. Later methods, such as LSD-SLAM [15] and ORB-SLAM [16], significantly improved robustness and scalability by adopting semi-dense and feature-based pipelines, respectively. ORB-SLAM, in particular, became a widely adopted benchmark due to its real-time performance and loop closure capability.

Direct methods such as DSO [17] and its successors optimize the photometric error on selected pixel sets, achieving higher accuracy in texture-rich environments. Recent contributions such as edgeSLAM [18], Struct-PL-SVO [19], and DVDS [20] continue to push the boundaries of monocular SLAM, incorporating structural regularities and deep features for improved perception in challenging scenes.

The main advantage of monocular SLAM lies in its simplicity, low power requirements, and minimal hardware cost. However, it inherently suffers from scale ambiguity, making the estimation of metric scale challenging without additional information. Furthermore, monocular systems are vulnerable to rapid motion, low-texture environments, and initialization sensitivity [21, 22]. Despite these limitations, monocular SLAM remains a fundamental building block in vSLAM research and a viable solution for applications constrained by size, weight, and power.

B. Stereo VSLAM

Stereo VSLAM leverages a pair of synchronized cameras to estimate depth through triangulation, thereby overcoming the inherent scale ambiguity found in monocular SLAM. By computing disparity between left and right images, stereo SLAM systems can obtain dense or semi-dense depth information directly, enabling more accurate pose estimation and robust mapping. As shown in Figure 3, the evolution of stereo SLAM includes a range of systems designed to enhance precision, efficiency, and robustness in various operational contexts.

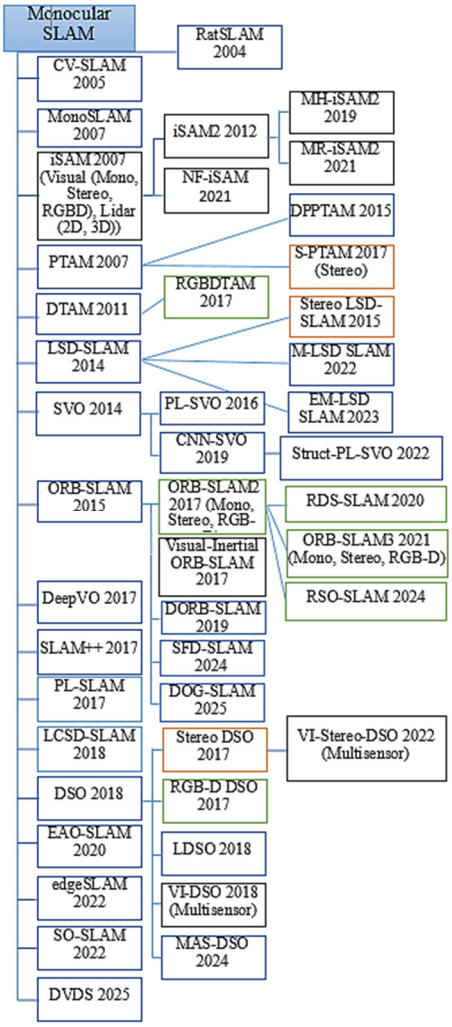

Fig. 2. Evolution of monocular VSLAM algorithms.

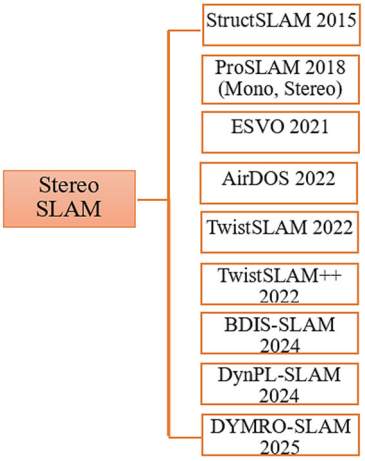

Fig. 3. Representative algorithms in stereo VSLAM.

Early systems such as StructSLAM [23] focused on structural constraints to improve mapping quality in man-made environments. ProSLAM [24], known for its lightweight implementation, supports both monocular and stereo

configurations and demonstrates reliable real-time performance. More recent methods, including AirDOS [25], TwistSLAM [26], and its extension TwistSLAM++ [27], integrate dynamic scene understanding and improved optimization techniques for enhanced motion robustness. Approaches such as DynPL-SLAM [28] and DYMRO-SLAM [29] address challenges in dynamic environments by incorporating motion segmentation and selective mapping strategies.

Stereo SLAM is generally more robust than monocular approaches in textureless scenes and under challenging lighting, due to its explicit depth sensing capability. However, stereo systems require careful calibration and are typically more computationally intensive. Despite this, their ability to deliver scale-consistent 3D maps makes them particularly valuable in autonomous vehicles, robotics, and AR/VR applications [30, 31].

C. RGB-D VSLAM

RGB-D VSLAM systems utilize color images (RGB) in conjunction with depth information provided by sensors such as Microsoft Kinect, Intel RealSense, or ASUS Xtion. These sensors capture per-pixel depth through structured light or time-of-flight technologies, allowing for direct 3D reconstruction of the environment. Unlike monocular or stereo SLAM, RGB-D SLAM circumvents the need for depth estimation via triangulation, significantly simplifying the SLAM pipeline and enhancing robustness in low-texture or low-light environments. As shown in Figure 4, RGB-D SLAM has seen rapid advances, with numerous algorithms developed to exploit the full potential of rich geometric and photometric data. Notable early contributions include KinectFusion [32] and ElasticFusion [33], which introduced dense surface reconstruction and surfel-based map representations. These methods enabled real-time, globally consistent 3D reconstruction and inspired a range of follow-up systems. DynaSLAM [34] extended ORB-SLAM2 with dynamic object detection and removal, improving performance in environments with moving entities. BAD SLAM [35], on the other hand, introduced a Bayesian approach for joint pose and dense geometry optimization, achieving high-precision mapping in challenging conditions. Fusion strategies play a central role in RGB-D SLAM. Systems like Co-Fusion [36] and MID-Fusion [37] perform joint segmentation and mapping to handle dynamic scenes effectively. BundleFusion [38], CodeSLAM [39], and FlowFusion [40] employ volumetric or learned representations for map construction. More recent frameworks, e.g., DeepLabv3+SLAM [41], Edge-SLAM [42], and DynNetSLAM [43], leverage deep learning for semantic understanding, motion segmentation, and robust feature extraction.

RGB-D SLAM is widely applied in robotic navigation, 3D scanning, augmented reality, and indoor mapping due to its capability to generate dense, metrically accurate maps in real time [9-11, 44-46]. However, its reliance on active sensors limits its usability in outdoor or high-illumination environments, and depth sensor range constraints can reduce performance in large-scale scenes.

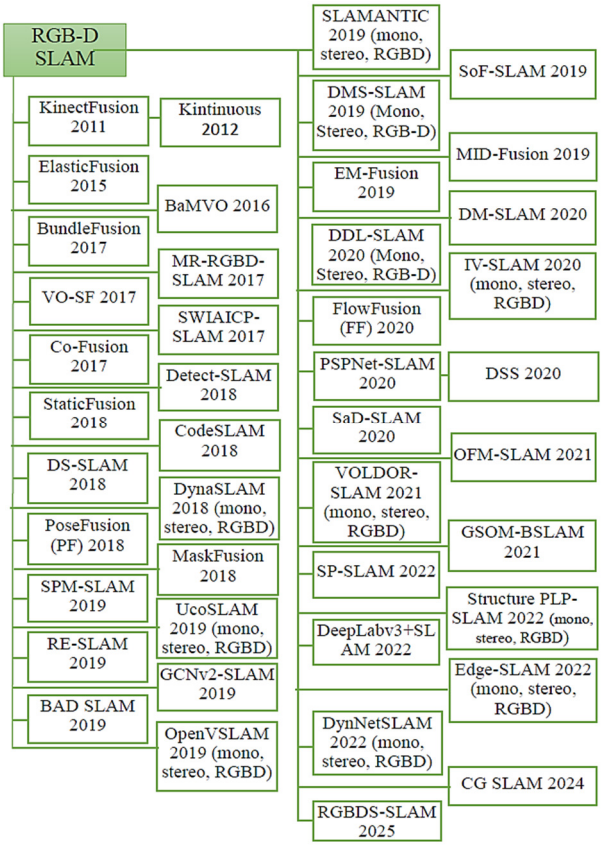

Fig. 4. Representative RGB-D SLAM algorithms.

D. Multi-Sensor and Hybrid VSLAM

Multi-sensor and hybrid VSLAM approaches have become essential in overcoming the limitations of traditional VSLAM, particularly in dynamic, large-scale, or perceptually challenging environments. These methods integrate data from multiple sources such as Inertial Measurement Units (IMUs), LiDAR, GNSS, ultrasonic sensors, or even semantic priors to enhance the robustness, precision, and versatility of SLAM systems. The fusion of complementary modalities provides more reliable localization and mapping by compensating for the weaknesses of individual sensors.

A prominent class within this paradigm is Visual-Inertial SLAM (VI-SLAM), which tightly couples visual odometry with inertial data through filters such as the Extended Kalman Filter (EKF) or non-linear optimization frameworks. Systems such as VINS-Mono and OKVIS have demonstrated superior robustness and accuracy in high-dynamic environments by leveraging inertial data to maintain motion estimation even during visual degradation [32-35]. Furthermore, LiDAR-Visual-Inertial fusion (e.g., LIO-SAM) introduces 3D geometric structure into SLAM pipelines, enabling more accurate and scalable mapping across varying terrain and illumination conditions [36-38].

Hybrid SLAM systems enhanced with deep learning integrate semantic scene understanding, object-level mapping, and robust feature extraction. Approaches such as JD-SLAM, VPS-SLAM, and YOLO-integrated SLAM employ CNNs for

dynamic object removal, real-time detection, and improved loop closure [39-43]. These methods boost performance in dynamic or cluttered environments, enabling applications in indoor navigation, urban robotics, UAV exploration, and assistive technologies. The growing adoption of sensor fusion and hybrid architectures reflects a shift toward adaptive, semantically enriched, and resilient SLAM frameworks for real-world deployment. Table I summarizes representative recent works, outlining their methods, environments, and hardware configurations.

TABLE I. SUMMARY OF RECENT MULTI-SENSOR AND HYBRID VSLAM APPROACHES

Title	Methods	Robustness	Hardware and datasets	Environment
Indoor Localization Using LiDAR and WSN [3]	SLAM + RSSI Fusion	Enhanced accuracy in complex layouts	WSN + Laser SLAM	Indoor complex
JD-SLAM: Joint Pose and Object Segmentation in Dynamic Scenes [5]	ORB-SLAM2 + object segmentation	Robust in dynamic scenes	Real-time CPU-based	Indoor/outdoor dynamic
Multi-Objective Mapping via Deep Learning and Visual SLAM [7]	ORB-SLAM2 + Mask R-CNN	Improved semantic mapping	NVIDIA Quadro M2000, Xeon CPU	Indoor/complex
NGLSFusion: Non-GPU Lightweight Semantic SLAM [9]	ORB-SLAM2/3, NGLSFusion	Efficient in low-resource settings	RGB-D + lightweight CPUs	Indoor, edge devices
3D Distance Measurement from a Camera [11]	Monocular, CNN, R-CNN, Triangulation	High; error < 0.9%	2 webcams	Indoor/outdoor
GNSS/INS/LiDAR Fusion for UAVs [13]	EKF, LOAM, Pix4D	Reduced RMSE (13–78%)	DJI UAV, IMU, GPS, LiDAR	GNSS-challenging outdoor
PLI-VINS: Visual-Inertial SLAM with Point-Line Fusion [13]	VINS-Mono, LSD, KLT Flow	Accurate in structured interiors	Low-cost IMU + Camera	Indoor
Visual Localization with Prior Semantic Maps [15]	ORB-SLAM2, SegNet, RandLA-Net	High accuracy in large indoor areas	Kinect, ZED, ORB-SLAM2	Indoor large-scale
Monocular SLAM with Semantic and Optical-Flow Fusion [15]	ORB-SLAM2, Mask R-CNN, LK Flow	10% higher accuracy than DynaSLAM	TUM RGB-D, Bonn datasets	Indoor dynamic
Visual SLAM for UAV-Based Obstacle Detection [15]	ORB-SLAM3, YOLOv3	Improved safety and localization	Pixhawk, UAV, Nvidia GPU	Indoor UAV
VPS-SLAM: Dynamic Object Removal via Object Detection + Optical Flow [17]	ORB-SLAM3 + YOLOR + LK Flow	89% object detection accuracy	Ricoh Theta Z1, MacBook Pro	Indoor/outdoor dynamic
Multi-Sensory Guidance System for Visually Impaired [17]	YOLO, ORB-SLAM, A-star, OCA	Effective, real-time assistance	Raspberry Pi, D435, ROS2	Indoor navigation
Real-Time Pose Estimation via Camera–LiDAR Fusion [17]	LIC-Fusion, PnP, PointNet	Robust pose estimation	KITTI dataset, LiDAR + Camera	Indoor/outdoor
Modular and Portable VSLAM for Multi-Environment Mapping [19]	ORB-SLAM2, GuPho	Modular, low-cost, multi-environment	Raspberry Pi, twin cameras	Indoor/outdoor/underwater
3D LiDAR-Based SLAM with Low-Cost RTK GPS [19]	HDL Graph SLAM, RTK-GPS	High accuracy (2–10 cm)	LiDAR, RTK-GPS, IMU	Outdoor
Dynamic Scene Vision SLAM using Object Detection [21]	ORB-SLAM3, YOLOv5, RANSAC	91% improved accuracy over ORB-SLAM3	TUM RGB-D	Indoor dynamic
Mobile Mapping Platform with LiDAR + Visual SLAM [21]	ORB-SLAM2, Cartographer, Open3D	Repeatable across domains	LiDAR, RealSense, Jetson Xavier	Indoor and outdoor
Mirror-Aware Visual SLAM Benchmark [23]	ORB-SLAM2/3, BundleFusion, RA-vSLAM	Improved in mirrored scenes	RealSense D435i, RTX 3070 GPU	Indoor mirrored environments

IV. APPLICATION DOMAINS

VSLAM has emerged as a cornerstone technology across diverse domains, enabling machines to perceive, localize, and interact with their surroundings in real time. In robotics and autonomous navigation, SLAM systems are essential for path planning, obstacle avoidance, and sustained autonomy. Whether in structured indoor spaces or unstructured outdoor environments, robots equipped with VSLAM can construct precise maps and estimate their trajectories without GPS. Well-established systems such as ORB-SLAM2, VINS-Mono, and LIO-SAM have demonstrated robust performance in service robots, UAVs, and self-driving vehicles [9, 10, 44, 45]. In AR and VR, VSLAM ensures accurate spatial tracking, enabling seamless alignment of virtual objects with the real world. Platforms such as Microsoft HoloLens and ARKit-based applications employ monocular or RGB-D SLAM to deliver immersive, low-latency, high-fidelity experiences [11]. 3D reconstruction and mapping benefit from RGB-D SLAM frameworks such as KinectFusion and ElasticFusion, which enable dense, real-time surface modeling for architecture, cultural heritage preservation, and industrial inspection [46]. Integrating deep learning into SLAM pipelines has advanced semantic reconstruction and contextual scene understanding, supporting autonomous exploration and smart city applications [12]. Wearable and mobile solutions—from navigation aids for the visually impaired to AR headsets—leverage energy-efficient visual-inertial SLAM, further expanded by edge computing, enabling deployment on resource-constrained consumer devices [13].

V. BENCHMARK DATASETS

Benchmark datasets play a pivotal role in the evaluation, comparison, and development of VSLAM algorithms, providing standardized environments, ground-truth trajectories, and sensor data that enable reproducible experimentation and quantitative assessment. In the past decade, several datasets emerged, each catering to specific sensor configurations, environmental conditions, and SLAM modalities.

One of the most widely used datasets is the TUM RGB-D dataset, which offers RGB-D sequences captured in indoor environments with ground-truth trajectories from motion capture systems [47]. It supports the evaluation of RGB-D SLAM and visual-inertial SLAM, particularly under dynamic and low-texture conditions. The KITTI Vision Benchmark Suite is another foundational dataset, designed for autonomous driving scenarios. It provides stereo images, LiDAR scans, and GPS/IMU measurements across urban and highway scenes, making it ideal for evaluating large-scale outdoor SLAM performance [48]. For micro aerial vehicles and high-precision localization tasks, the EuRoC MAV dataset includes synchronized stereo and IMU data from a drone operating in industrial and cluttered indoor environments [14]. This dataset is frequently used to benchmark visual-inertial SLAM systems. The ICL-NUIM dataset offers synthetic RGB-D data with precise ground-truth, which is beneficial for evaluating dense reconstruction methods without real-world sensor noise [15]. Datasets such as New College [16] and the MIT Stata Center [17] were among the earliest benchmarks for monocular and stereo SLAM in real-world robotics environments. More recent additions include the UZH-FPV Drone Racing dataset [18], which introduces high-speed drone flight data for testing SLAM under aggressive motion, and the ScanNet dataset, which offers RGB-D videos of indoor scenes with semantic annotations for training and evaluating semantic SLAM systems [23]. Datasets tailored to dynamic environments, such as TUM Dynamic Objects [49] and Bonnet RGB-D [24], allow testing the robustness of SLAM against moving objects and occlusions, supporting recent advances in dynamic scene understanding. In general, the availability of diverse and challenging datasets has been instrumental in accelerating SLAM research. However, there remains a need for more unified benchmarks that include dynamic, long-term, multi-sensor, and semantic-rich sequences to fully support the evaluation of emerging SLAM 2.0 systems.

VI. OPEN CHALLENGES AND FUTURE DIRECTIONS

Despite significant advances in VSLAM, several challenges hinder its scalability and real-world deployment. Lifelong SLAM remains a critical issue, requiring continuous mapping and localization without drift or excessive memory use, necessitating strategies such as map summarization, submap fusion, and place re-recognition [25, 26]. Robustness in dynamic environments is another concern, as traditional methods assume static scenes, while moving objects can cause failures, and solutions such as motion segmentation and instance-aware tracking remain computationally expensive [27]. Semantic SLAM, which integrates object-level mapping and affordance reasoning, improves scene understanding but introduces latency and dependence on large datasets [50]. Achieving real-time performance on edge devices demands network compression, computational optimization, and hardware accelerators [51, 52]. Additionally, benchmarking inconsistencies and limited reproducibility hinder progress, leading to calls for standardized datasets and metrics [28]. Future SLAM 2.0 systems are expected to be learning-based, semantic-aware, and cloud-integrated, enabling collaborative mapping, adaptive operation, and integration with external knowledge sources.

VII. CONCLUSION

This survey presented an in-depth examination of recent developments in VSLAM, following the progression from traditional feature-based and direct methods to hybrid and learning-based approaches that improve accuracy, robustness, and semantic awareness. By categorizing monocular, stereo, RGB-D, and multi-sensor SLAM, this review outlined the distinct strengths and trade-offs associated with each modality. VSLAM has become integral to robotics, AR/VR, 3D reconstruction, and mobile systems, yet challenges such as long-term scalability, dynamic environment adaptation, real-time edge deployment, and reproducibility persist. The future direction, often referred to as SLAM 2.0, envisions semantically enriched, cloud-integrated, and AI-driven frameworks capable of collaborative mapping and context-aware perception. Advances in deep learning, sensor fusion, and edge–cloud architectures are poised to push SLAM toward more autonomous, adaptive, and intelligent capabilities. Realizing this vision will require sustained interdisciplinary research, standardized benchmarks, and open-source collaboration, ensuring that SLAM technology continues to evolve as a cornerstone of next-generation autonomous, interactive, and intelligent systems.

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," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 1418–1425, Jul. 2018, https://doi.org/10.1109/LRA.2018.2800113.

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

[3] B. Kulambayev, B. Gleb, N. Katayev, I. Menglibay, and Z. Momynkulov, "Real-Time Road Damage Detection System on Deep Learning Based Image Analysis," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 9, 2024, https://doi.org/10.14569/IJACSA.2024.01509107.

[4] T. Shan, B. Englot, C. Ratti, and D. Rus, "LVI-SAM: Tightly-coupled Lidar-Visual-Inertial Odometry via Smoothing and Mapping," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, Xi'an, China, May 2021, pp. 5692–5698, https://doi.org/10.1109/ICRA48506.2021.9561996.

[5] V. H. Le, H. S. Do, V. N. Phan, and T. H. Te, "TQU-SLAM Benchmark Feature-based Dataset for Building Monocular VO," *Engineering, Technology & Applied Science Research*, vol. 14, no. 4, pp. 15330–15337, Aug. 2024, 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, Mar. 2022, Art. no. 169, https://doi.org/10.3390/machines10030169.

[7] B. Omarov, M. Baikuvekov, D. Sultan, N. Mukazhanov, M. Suleimenova, and M. Zhekambayeva, "Ensemble Approach Combining Deep Residual Networks and BiGRU with Attention Mechanism for

Classification of Heart Arrhythmias," *Computers, Materials & Continua*, vol. 80, no. 1, pp. 341–359, 2024, https://doi.org/10.32604/cmc.2024.052437.

[8] B. Omarov *et al.*, "Electronic Stethoscope for Heartbeat Abnormality Detection," in *Smart Computing and Communication*, vol. 12608, M. Qiu, Ed. Springer International Publishing, 2021, pp. 248–258.

[9] S. Hensel, M. B. Marinov, and M. Obert, "3D LiDAR Based SLAM System Evaluation with Low-Cost Real-Time Kinematics GPS Solution," *Computation*, vol. 10, no. 9, Sep. 2022, Art. no. 154, https://doi.org/10.3390/computation10090154.

[10] 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, https://doi.org/10.1007/s41095-022-0329-x.

[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, Jan. 2023, Art. no. 3529, https://doi.org/10.3390/s23073529.

[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," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLVIII-2-W1-2022, pp. 153–162, Dec. 2022, https://doi.org/10.5194/isprs-archives-XLVIII-2-W1-2022-153-2022.

[13] I. Kalisperakis *et al.*, "A Modular Mobile Mapping Platform for Complex Indoor and Outdoor Environments," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLIII-B1-2020, pp. 243–250, Aug. 2020, https://doi.org/10.5194/isprs-archives-XLIII-B1-2020-243-2020.

[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, Jul. 2022, Art. no. 343, https://doi.org/10.3390/info13070343.

[15] C. Bonfanti, G. Patrucco, S. Perri, G. Sammartano, and A. Spanò, "A New Indoor LiDAR-Based MMS Challenging Complex Architectural Environments," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLVI-M-1–2021, pp. 79–86, Aug. 2021, https://doi.org/10.5194/isprs-archives-XLVI-M-1-2021-79-2021.

[16] T. Alhmiedat *et al.*, "A SLAM-Based Localization and Navigation System for Social Robots: The Pepper Robot Case," *Machines*, vol. 11, no. 2, Feb. 2023, Art. no. 158, https://doi.org/10.3390/machines11020158.

[17] C. Mai *et al.*, "A Smart Cane Based on 2D LiDAR and RGB-D Camera Sensor-Realizing Navigation and Obstacle Recognition," *Sensors*, vol. 24, no. 3, Jan. 2024, Art. no. 870, https://doi.org/10.3390/s24030870.

[18] G. Patrucco, G. Sammartano, C. Bonfanti, and A. Spanò, "Assessing Terrestrial MMS 3D Data for Outdoor Multi-Scale Modelling," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLVIII-1-W1-2023, pp. 371–378, May 2023, https://doi.org/10.5194/isprs-archives-XLVIII-1-W1-2023-371-2023.

[19] D. Merkle and A. Reiterer, "Automated Method for SLAM Evaluation in GNSS-Denied Areas," *Remote Sensing*, vol. 15, no. 21, Jan. 2023, Art. no. 5141, https://doi.org/10.3390/rs15215141.

[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, Jan. 2023, Art. no. 1455, https://doi.org/10.3390/math11061455.

[21] Y. Wang, S. Zhang, and J. Wang, "Ceiling-View Semi-Direct Monocular Visual Odometry with Planar Constraint," *Remote Sensing*, vol. 14, no. 21, Jan. 2022, Art. no. 5447, https://doi.org/10.3390/rs14215447.

[22] V. Di Pietra, N. Grasso, M. Piras, and P. Dabove, "Characterization of a Mobile Mapping System for Seamless Navigation," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLIII-B1-2020, pp. 227–234, Aug. 2020, https://doi.org/10.5194/isprs-archives-XLIII-B1-2020-227-2020.

[23] 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, Jan. 2023, Art. no. 18, https://doi.org/10.3390/s23010018.

[24] D. Bolkas, M. O'Banion, and C. J. Belleman, "Combination of TLS and SLAM Lidar for Levee Monitoring", *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. V-3–2022, pp. 641–647, May 2022, https://doi.org/10.5194/isprs-annals-V-3-2022-641-2022.

[25] L. He, Z. Jin, and Z. Gao, "De-Skewing LiDAR Scan for Refinement of Local Mapping," *Sensors*, vol. 20, no. 7, Jan. 2020, Art. no. 1846, https://doi.org/10.3390/s20071846.

[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 the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. X-4-W1-2022, pp. 523–530, Jan. 2023, https://doi.org/10.5194/isprs-annals-X-4-W1-2022-523-2023.

[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, Jan. 2022, Art. no. 6941, https://doi.org/10.3390/s22186941.

[28] T. Bodrumlu and F. Caliskan, "Indoor Position Estimation Using Ultrasonic Beacon Sensors and Extended Kalman Filter," *Engineering Proceedings*, vol. 27, no. 1, 2022, Art. no. 16, https://doi.org/10.3390/ecsa-9-13353.

[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, Jan. 2022, Art. no. 939, https://doi.org/10.3390/app12030939.

[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, Jun. 2023, Art. no. 404, https://doi.org/10.3390/drones7060404.

[31] G. Ge, Z. Qin, and X. Chen, "Integrating WSN and Laser SLAM for Mobile Robot Indoor Localization," *Computers, Materials & Continua*, vol. 74, no. 3, pp. 6351–6369, 2023, https://doi.org/10.32604/cmc.2023.035832.

[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, Jan. 2021, Art. no. 1729881421994447, https://doi.org/10.1177/1729881421994447.

[33] Y. Sun *et al.*, "Multi-Objective Location and Mapping Based on Deep Learning and Visual Slam," *Sensors*, vol. 22, no. 19, Jan. 2022, Art. no. 7576, https://doi.org/10.3390/s22197576.

[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, Jan. 2022, Art. no. 4258, https://doi.org/10.3390/rs14174258.

[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, Jan. 2023, Art. no. 15, https://doi.org/10.3390/rs15010015.

[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, Jan. 2022, Art. no. 5457, https://doi.org/10.3390/s22145457.

[37] G. Chen and L. Hong, "Research on Environment Perception System of Quadruped Robots Based on LiDAR and Vision," *Drones*, vol. 7, no. 5, May 2023, Art. no. 329, https://doi.org/10.3390/drones7050329.

[38] J. Geng *et al.*, "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, Feb. 2023, Art. no. 012027, https://doi.org/10.1088/1742-6596/2467/1/012027.

[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, Jan. 2021, Art. no. 772, https://doi.org/10.3390/rs13040772.

[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, Feb. 2023, Art. no. 085202, https://doi.org/10.1088/1361-6501/acd1a4.

[41] Md. A. A. Noman *et al.*, "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 (IJECE)*, vol. 13, no. 1, pp. 347-357, Feb. 2023, https://doi.org/10.11591/ijece.v13i1.pp347-357.

[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, Jan. 2021, Art. no. 409, https://doi.org/10.3390/s21020409.

[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, Jan. 2022, Art. no. 7553, https://doi.org/10.3390/s22197553.

[44] 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, Jan. 2023, Art. no. 3048, https://doi.org/10.3390/su15043048.

[45] 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, Jan. 2022, Art. no. 9908, https://doi.org/10.3390/s22249908.

[46] 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, Jan. 2023, Art. no. 167, https://doi.org/10.18494/SAM4225.

[47] 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, Jan. 2022, Art. no. 3396, https://doi.org/10.3390/electronics11203396.

[48] W. Chen *et al.*, "A Monocular-Visual SLAM System with Semantic and Optical-Flow Fusion for Indoor Dynamic Environments," *Micromachines*, vol. 13, no. 11, Nov. 2022, Art. no. 2006, https://doi.org/10.3390/mi13112006.

[49] L. Morelli, F. Ioli, R. Beber, F. Menna, F. Remondino, and A. Vitti, "COLMAP-SLAM: A Framework for Visual Odometry," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLVIII-1-W1-2023, pp. 317–324, May 2023, https://doi.org/10.5194/isprs-archives-XLVIII-1-W1-2023-317-2023.

[50] 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, Jan. 2023, Art. no. 2908, https://doi.org/10.3390/app13052908.

[51] 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, no. 1, Nov. 2022, Art. no. 18815, https://doi.org/10.1038/s41598-022-22938-y.

[52] N. Li *et al.*, "Indoor and Outdoor Low-Cost Seamless Integrated Navigation System Based on the Integration of INS/GNSS/LIDAR System," *Remote Sensing*, vol. 12, no. 19, Jan. 2020, Art. no. 3271, https://doi.org/10.3390/rs12193271.