Recent Developments in SLAM and Drone Autonomy: A Five-Year Perspective

Prashant Raj Bista Toshika Ojha

August 2025

1 Abstract

This paper presents an overview of significant advancements in SLAM and drone autonomy from 2020 to 2025. Some of the key innovations include resource-constrained SLAM solutions like NanoSLAM for nano-UAVs, dynamic environment adaptations such as DynaVINS and DynaGSLAM, and improvements in depth image quality for enhanced 3D reconstruction. Recent developments in AI-driven drone control, including foundation model-guided path planning and LLM-generated command systems which showcase the integration of machine learning for autonomous navigation. The survey highlights breakthroughs in global consistency and drift-free localization through digital twin alignment and efficient sparse mapping methods addressing computational constraints. Future research directions emphasize improving real-time processing, processing at lower resources, robustness in dynamic settings, and ensuring the safety and reliability of AI-enabled drone systems. This review provides a reference point for learners who are seeking to know the landscape of drone autonomy and SLAM technologies.

2 Introduction

This paper provides a review of the advancements in drone autonomy and Simultaneous Localization and Mapping (SLAM) from 2020 to 2025, drawing insights from various literature. The increasing availability of compact, cost-effective drones, coupled with significant progress in artificial intelligence and sensor technologies, has propelled UAVs into diverse applications, including surveillance, remote sensing, delivery services, disaster management, and even space exploration. The main aspect of the utility of drones is the ability to operate autonomously in complex, dynamic, and GPS-denied environments, which requires robust environment perception, precise self-localization, and intelligent navigation capabilities. SLAM enables UAVs to construct a map of an unknown environment while simultaneously estimating its position within it. This becomes particularly vital in an environment where GPS cannot be accurate and where reliance on visual and inertial data is crucial, such as indoors. SLAM, a fundamental algorithm enabling robots to concurrently estimate their own pose and construct a map of their surroundings, is at the core of achieving autonomy. This review synthesizes the progress in both SLAM and drone autonomy, highlighting how these fields have evolved to address limitations and expand operational capabilities over the period of last 5 years.

3 Scope of the Review

The main objective of this review paper is to provide researchers with an overview of the significant progress, key innovations, and emerging trends in the fields of drone autonomy and SLAM during the 2020-2025 period. By synthesizing information from the provided sources, this paper aims to offer insights that will enhance understanding of these rapidly evolving domains and guide future research directions, particularly concerning the interplay between AI/ML, sensor technology, and computational efficiency in autonomous drone systems. So the scope of the paper can be summarized in the following.

- Review of autonomy of drones from 2020 to 2025.
- Review of progress in the field of SLAM from 2020 to 2025.

4 Literature Review

4.1 Development of Drone Autonomy

For low-cost toy drones, Tukan et al. (2023) developed an autonomous indoor exploration system that uses a monocular RGB camera and ORB-SLAM3. Their work tackles key problems such as the lack of accurate positioning in GPS-denied indoor environments, the presence of noisy outlier data generated by ORB-SLAM3's sparse point clouds, and the challenge of adapting path planners to such sparse data representations. Their primary contributions include a novel outlier removal algorithm with provable guarantees and a method to transform sparse point clouds into a suitable format for path planning, employing techniques like K-means clustering and convex hulls. The system showed effectiveness in real-time experiments for path generation and navigating unexplored areas. Future work suggests adapting the system for even more resource-limited platforms like the Raspberry Pi and integrating deep learning models with compression techniques to further enhance performance under severe hardware constraints.

For nano-drones, Sartori, Singhal, Roy, Brunelli, and Gross (2025) specifically addressed the challenge of safe autonomous navigation and high-level tasks like exploration and surveillance, considering their inherent resource limitations. Their main contribution is an AI-aided, vision-based reactive planning method for obstacle avoidance, employing a split computing strategy where computationally intensive deep learning-based object detection runs on an external edge device, while the planning algorithm executes onboard. This innovative approach enabled them to command the drone at approximately 8 frames per second (FPS). Their object detection model achieved a mean-average-precision (mAP) of 60.8%. Field experiments validated the solution's feasibility, with the drone flying at a top speed of 1 m/s while successfully avoiding obstacles; specific success rates were 76% for short obstacles and 64% for larger ones. Future work for this system includes exploring multi-point navigation, trajectory optimization, and further analyzing the trade-offs in offloading AI model execution to edge devices versus onboard processing.

Xiao, Tsao, Zhang, and Feroskhan (2025) introduced FM-Planner, a framework for autonomous drone path planning that leverages foundation models (Large Language Models (LLMs) and Vision-Language Models (VLMs)). This work addresses the limitations of traditional path planning algorithms (e.g., A*, RRT) and the generalization issues of earlier learning-based approaches (e.g., imitation learning, deep reinforcement learning) in complex or unforeseen environments. The main contribution is a systematic benchmarking study of various LLM and VLM approaches, leading to the development of an integrated LLM-Vision planner for real-time navigation. Their evaluation demonstrated that the fine-tuned Llama-3.1-8B-Instruct consistently outperformed other models, maintaining a success rate of $\geq$90% even in scenarios with five obstacles, where other models failed entirely. The LLM-Vision-guided planner had an average reasoning time of approximately 9 seconds per query, which is considered acceptable for global path planning that does not require high-frequency updates. Future research will delve into more advanced multimodal foundation models for highly dynamic environments and aim to optimize real-time computational performance for broader operational scenarios.

Finally, in a pioneering effort, Burke (2025) demonstrated a fully AI-generated drone command and control system (WebGCS), showcasing the ability of a robot to build another robot's brain. This directly addresses the extensive human effort that is typically required to develop complex drone control software, which traditionally involves years of manual coding. The system uniquely enables real-time mapping, flight telemetry, and autonomous mission planning through a web interface hosted directly on the drone itself. This AI-authored codebase consists of approximately 10k LOC. The program was developed in just about two weeks of human time, indicating an increase in development speed compared to traditional methods. Flight tests confirmed that the basic autonomous actions were executed flawlessly, and the drone's Wi-Fi connection remained stable up to 100 m. The study notes that current AI models are near their practical limits in handling such complex codebases (around 10k lines), especially concerning intricate information flow and transient events. Future directions include exploring AI agent swarms for controlling real drone swarms and critical research into the safety and reliability of AI-generated code, particularly for higher-risk operations.

4.2 Development of SLAM

DynaVINS Song, Lim, Lee, and Myung (2022) focused on enhancing the robustness of visual-inertial SLAM (VI-SLAM) in dynamic environments. It aimed to overcome pose estimation degradation and false positive loop closures caused by dynamic and temporarily static objects, which is one of the common limitations of SLAM algorithms that assume all the landmarks to be static. DynaVINS introduced a robust bundle adjustment method that leverages IMU pre-integration to effectively reject features from dynamic objects. It also implemented keyframe grouping and multi-hypothesis-based constraints grouping to minimize the adverse effects of temporarily static objects during loop closure. Experimental results on the VIODE dataset demonstrated promising performance with less degeneration in Absolute Trajectory Error (ATE) compared to other SOTA methods like VINS-Fusion-S-I and ORB-SLAM3-S-I. Future work includes improving the system's speed and overall performance, and extending the DynaVINS concept to the LiDAR-Visual-Inertial (LVI) SLAM framework.

Addressing the escalating computational and memory demands of visual SLAM, Park and Bae Park and Bae (2022) presented an efficient graph optimization method for map point sparsification. This technique directly tackles the problem of quadratically increasing memory size and computation costs as the map grows in SLAM systems. Their approach formulates the problem of maximizing pose-visibility and spatial diversity as a minimum-cost maximum-flow graph optimization. This can be integrated as an additional step into existing feature-based visual SLAM pipelines, such as ORB-SLAM2. The method achieved significantly more accurate camera poses while reducing map points by approximately 1/3 and computation by 1/2. Their sparsified map, when compared to the original ORB-SLAM2, consisted of 39% fewer points while marginally improving pose accuracy. Future research was directed towards the involvement of marginal graph optimization for even faster processing and considering the spatial density of 3D points in addition to 2D feature diversity.

For highly resource-constrained platforms, NanoSLAM Niculescu, Polonelli, Magno, and Benini (2024) made a breakthrough by enabling fully onboard SLAM for tiny robots like nano-UAVs. This was previously challenging, often requiring high-wattage processors or external computation offloading due to the computational load and memory demands of SLAM. NanoSLAM integrates low-power multi-zone Time of Flight (ToF) sensors with a parallel RISC-V processor (GAP9 SoC) to overcome limitations in payload, power budget, and memory. The system demonstrated a mapping accuracy of 4.5 cm and an end-to-end execution time of less than 250ms, leading to a trajectory error reduction of up to 67%. Its onboard Iterative Closest Point (ICP) scan-matching operates efficiently in just 55ms, and the entire mapping pipeline consumes less than 500 KB of RAM and only 87.9mW of power for loop closure. Future work plans to integrate deep learning models for object identification and leverage the generated maps for optimal path planning.
Chakrabarty and Sarangi (2024) introduced VoxDepth, a solution designed for the rectification of depth images on edge devices. They addressed the prevalent issue of noise (including flickering pixels and algorithmic holes resulting from stereoscopic matching failures) in depth images, which significantly degrades the performance

of 3D reconstruction and visual SLAM. Conventional ML-based methods were deemed too slow for edge devices (operating at 2-3 FPS, far below the required rate), while non-ML methods lacked the necessary accuracy for complex algorithmic holes. VoxDepth offers a fast and accurate non-ML approach that utilizes 3D point cloud construction and fusion to create a robust template for correcting incorrect depth images. It demonstrated a 31% improvement in quality (PSNR) and maintained a competitive framerate of 27 FPS on an NVIDIA Jetson Nano board, proving to be 58% faster than its closest competing proposals. The study underscored the critical importance of precise depth images, revealing a super-linear increase in drone swarming collision rates as noise levels rise.

Davletshin et al. (2024) focused on enhancing 3D reconstruction and segmentation within object-oriented SLAM (DSP-SLAM) by improving the quality of RGB images, specifically mitigating motion blur caused by the agile movements of drones. This system incorporates a GAN architecture to deblur low-quality frames, which in turn improves localization, mapping, and object reconstruction for highly dynamic UAV systems. Experiments showed that applying deblurring led to a 38.46% increase in detected points on objects, an improvement in Intersection over Union (IoU) from 74.51% to 75.67%, and a reduction in the RMSE of the trajectory's signed distance function from 17.2 cm to 15.4 cm. Future work aims to optimize the reconstruction pipeline for better initial position estimation and object correction, and to explore multi-agent SLAM systems for enhanced area coverage and reconstruction quality.

Merat, Cioffi, Bauersfeld, and Scaramuzza (2024) introduced a novel approach for drift-free Visual SLAM leveraging Digital Twins to counteract the long-term drift in VIO/VSLAM, particularly in urban environments where GPS signals are unreliable or unavailable. Their core methodology involves aligning the sparse 3D point cloud generated by the VIO/VSLAM system with an existing 3D digital twin of the environment using point-to-plane matching. This technique removes the reliance on visual data association and provides a 6-DoF global measurement tightly integrated into the VIO/VSLAM system. The method exhibited superior robustness to viewpoint changes compared to traditional visual SLAM techniques and improved absolute trajectory errors in real-world drone flights by 32% in position and 28% in rotation when compared to state-of-the-art VIO-GPS systems. However, the approach can be susceptible to 3D aliasing and inaccuracies in either the digital twin or the local point cloud.

DynaGSLAM Li et al. (2025) introduced the first online SLAM system based on 3D Gaussian Splatting that renders, tracks, and predicts dynamic object motions alongside ego-motion estimation in real-time. The core problem addressed was the failure of existing GS-SLAM methods in dynamic scenes where moving objects negatively impact mapping quality. DynaGSLAM resolves this by distinguishing between static and dynamic objects through a world-centric graph optimization strategy. Evaluations on real-world datasets like OMD, TUM, and BONN demonstrated that DynaGSLAM significantly outperformed baselines such as RTGSLAM and SplatTAM, particularly in accurately mapping areas around moving objects. Quantitatively, DynaGSLAM achieved faster mapping speeds (347 ms/frame) and localization speeds (95 ms/frame), with efficient resource usage (22K GS numbers, 2.6 GB memory). A key gap identified for future work is the need to explore more complex motion models and functions while ensuring the system's efficiency, as the current version prioritizes real-time performance over the sophistication of the motion model.

Finally, Veg Tyagi and Gaur (2025) developed a comprehensive autonomous surveillance quadcopter featuring integrated SLAM, onboard fault management, and embedded vision, specifically engineered for operation in GPS-denied environments and with robust hardware fault tolerance. It utilizes ORB-SLAM3 in visual-inertial mode to provide real-time 6-DoF pose estimation and build sparse maps, thereby eliminating the need for external positioning systems. The system also incorporates an LQR-based flight control system and a real-time Fault Detection and Identification (FDI) pipeline. ORB-SLAM3 operates at approximately 10 Hz, while onboard object detection functions at around 2 FPS with a high classification accuracy of over 94%. The total system power consumption on a Raspberry Pi 4 is approximately 4.3 W, with the SLAM and vision modules accounting for the majority of this power draw. Simulations validated Veg's performance in trajectory tracking and rotor loss scenarios, demonstrating minimal drift during maze navigation with successful SLAM loop closure. Future work for Veg includes integrating stereo vision for obstacle detection, enhancing vision modules with hardware acceleration, and exploring adaptive control strategies and reinforcement learning for dynamic mission planning.

5 Methodology

This review was conducted using a literature review methodology based on papers from the last five years. The papers are all open-access papers only. The selection criteria focused on studies published within or relevant to the 2020-2025 timeframe and that are related to SLAM and drone autonomy. Keywords such as Drone, SLAM, and Drone Autonomy were used for finding the papers. Key information regarding drone control, navigation strategies, AI integration (e.g., GAN, object detection, LLMs), and SLAM advancements (e.g., mapping techniques, dynamic scene handling, efficiency improvements, hardware integration) was extracted and critically analyzed. The identified advancements were then categorized and synthesized to address the defined research questions, highlighting both successes and persistent challenges.

6 Critical Analysis

The period from 2020 to 2025 represents a transformative phase for drone autonomy and SLAM, characterized by a concerted effort to push beyond theoretical boundaries and address practical deployment challenges.

6.1 Progress in Drone Autonomy

Progress in drone autonomy has significantly advanced, moving beyond basic remote control to sophisticated AI and vision-based systems capable of complex tasks. The integration of machine learning (ML) algorithms is considered important for future autonomous drone operation, enhancing efficiency and adaptability across various sectors such as search and rescue, agricultural oversight, and environmental monitoring. A revolutionary development is the emergence of fully AI-generated drone control systems, presented by "WebGCS," where an AI model authors all the code for a real-time, self-hosted drone command and control platform. This system, which enables real-time mapping, flight telemetry, autonomous mission planning, and safety protocols accessible via a web interface directly from the drone, represents a significant improvement in robotics engineering by accelerating development cycles and multiplying the range of possible applications. For complex global path planning tasks, foundation models, including large language models (LLMs) and vision-language models (VLMs), are being explored, with frameworks like FM-Planner leveraging LLMs for high-level spatial reasoning and integrating visual perception to plan safe and efficient trajectories in both simulated and real-world environments. Furthermore, to overcome the severe hardware limitations (weight, power, memory) of miniaturized platforms, split computing strategies are crucial, offloading computationally intensive tasks like deep learning-based object detection to external edge devices while core planning algorithms run onboard the nano-drone. This allows for safe autonomous flight and obstacle avoidance for tiny platforms, demonstrating the feasibility of real-time navigation despite significant resource constraints. Such advancements are foundational for enabling more sophisticated functionalities, including optimal path planning and multi-agent collaboration in highly constrained robotic systems

6.2 Progress in SLAM

SLAM is a critical algorithm for environment perception and navigation in autonomous systems like drones, with continuous progress in addressing complex real-world challenges. Key areas of advancement include:

6.2.1 Handling Dynamic Environments

Traditional Gaussian Splatting SLAM (GS-SLAM) methods often fail in scenes with moving objects, corrupting static map regions. DynaGSLAM addresses this by explicitly managing dynamic Gaussian Splats, segmenting them from static ones and estimating their 3D motion, leading to superior mapping quality, especially around dynamic objects like people or balloons. Similarly, DynaVINS is a robust visual-inertial SLAM (VI-SLAM) framework designed to handle both dynamic and temporarily static objects, improving pose estimation accuracy by adaptively adjusting weights for dynamic features and rejecting false positive loop closures.

6.2.2 Mitigating Image Quality Issues

Visual SLAM systems on UAVs are susceptible to motion blur and sensor noise, which can degrade accuracy. SharpSLAM utilizes Generative Adversarial Networks (GANs) to deblur low-quality images, enhancing interest point detection and significantly improving the precision of 3D reconstruction and segmentation for agile drones. For depth images, which often suffer from flickering pixels and algorithmic holes, VoxDepth proposes a fast and accurate rectification scheme. It achieves this by constructing and fusing 3D point clouds to create a template, demonstrating a 31% improvement in quality compared to state-of-the-art methods on real-world datasets.

6.2.3 Overcoming Resource Constraints

Running SLAM onboard miniaturized drones faces significant computational and memory hurdles. NanoSLAM is a pioneering framework enabling fully onboard SLAM for nano-UAVs, operating with a power budget of only 87.9mW and achieving a mapping accuracy of 4.5 cm by exploiting low-power depth sensors and parallel processing. For larger systems, point sparsification algorithms efficiently reduce the number of map points and computation cost (e.g., by approximately 1/3 of map points and 1/2 of computation when building on ORB-SLAM2) while maintaining or improving pose accuracy, which is crucial for real-time performance on embedded systems. Furthermore, specific implementations like the Tello Drone navigation system using ORB-SLAM3 incorporate outlier removal algorithms (e.g., using minimax jointly-submodular function optimization) to enhance map quality by distinguishing noisy features from actual obstacles.

6.2.4 Achieving Global Consistency

To address long-term drift, particularly where GPS is unreliable (e.g., urban environments), a novel approach integrates VSLAM systems with geometric digital twins. This method localizes the sparse 3D point cloud generated by VSLAM to a city digital twin via point-to-plane matching, providing globally consistent, low-drift pose estimates without relying on visual feature matching. This approach has shown superiority over VIO-GPS systems and greater robustness to viewpoint changes.

7 Research Gap and Future Directions

7.1 Research Gaps

Current SLAM methods, particularly those using Gaussian splats, often struggle with dynamic scenes, either by failing to account for moving objects or by removing them, thus rendering only the static background. Traditional Visual-Inertial Odometry (VIO) and Visual SLAM (VSLAM) systems are prone to long-term drift and rely heavily on local sensor data, with GPS being unreliable or unavailable in many environments. There's a persistent challenge in handling noisy and inaccurate depth images, especially errors due to occlusion and camera movement, and ML-based rectification methods are often too computationally intensive for edge devices. For resource-constrained platforms like nano-drones, achieving fully onboard mapping is difficult due to significant computational load and memory demands, often necessitating offloading computation, which introduces latency and limits operational range. Furthermore, adjusting path planners to sparse point clouds generated by SLAM (e.g., ORB-SLAM3) remains a challenge, as it lacks clear distinctions between obstacles and noisy points, potentially leading to collisions. The practical application of bio-inspired UAVs is still experimental and costly, with significant technical hurdles. There is also a notable gap in comprehensive risk assessments regarding privacy, ethical, and environmental impacts (like carbon emissions and noise pollution) of widespread drone usage. From an AI perspective, current learning-based approaches struggle with generalization to unseen scenarios and require vast amounts of domain-specific data. While foundation models (LLMs, VLMs) show promise, their practical applicability and effectiveness in global path planning for drones are largely unexplored, often lacking robust spatial reasoning and integration with real-time perception. Moreover, AI models are currently unable to write complex robot operating systems code from scratch, and the rigorous testing and safety validation of AI-generated code remain open problems.

7.2 Future Direction

A significant gap lies in advancing SLAM for dynamic environments, as current methods, including Gaussian splatting-based SLAM, primarily operate on static scenes or remove dynamic objects, necessitating future work in real-time rendering, tracking, and motion prediction of dynamic objects. This also extends to VIO and VSLAM systems, which currently suffer from long-term drift and struggle with temporarily static objects, requiring more robust global optimization methods. Enhancing mapping and reconstruction quality is another critical area, specifically addressing noisy and inaccurate depth images, especially errors due to occlusion and camera movement, and the need for computationally efficient rectification methods suitable for edge devices. Future work should focus on improving depth image rectification, generating more coherent 3D maps, and discarding dynamic object measurements. For resource-constrained platforms, such as nano-drones, achieving fully onboard mapping remains a challenge due to high computational and memory demands, often requiring computation offloading that introduces latency. Future efforts aim for optimized onboard computation, efficient map decimation, and leveraging deep learning models with compression for limited hardware like Raspberry Pi. In path planning and autonomous navigation, a key gap is adapting planners to sparse point clouds generated by SLAM systems (e.g., ORB-SLAM3), which often lack clear distinctions between obstacles and noisy data, potentially leading to collisions. This requires transforming sparse data into denser representations and exploring multi-point navigation and trajectory optimization. The practical applicability of bio-inspired UAVs is still largely experimental and costly, facing significant technical hurdles.

Finally, advanced AI and foundation models pose both opportunities and challenges. Current learning-based approaches struggle with generalization to unseen scenarios and require vast amounts of domain-specific data. The practical applicability and effectiveness of foundation models (LLMs, VLMs) in global path planning for drones are largely unexplored, and they often lack robust spatial reasoning and real-time perception integration. Furthermore, current AI models are unable to write complex robot operating systems (ROS) from scratch, and the rigorous testing and safety validation of AI-generated code remain open problems. Future work will explore multi-agent SLAM systems for faster mapping and enhanced reconstruction, improve system robustness against false positives, and develop more resilient fault detection and control mechanisms.

Table 1: Summary of key papers in SLAM and drone autonomy, highlighting strengths and weaknesses.

Paper / Source	**Strengths**	**Weaknesses / Gaps**
(Song et al., 2022)	Robust VI-SLAM for dynamic environments; effectively rejects dynamic features; multi-hypothesis feature grouping; outperforms SOTA methods in dynamic scenarios.	Potential for further speed improvements; limited to VI-SLAM (no LiDAR); future extension to LVI SLAM needed.
(Park & Bae, 2022)	Efficient graph optimization reduces map points and computation by 50% while improving pose accuracy; it integrates multiple SLAM map quality properties.	Marginal graph optimization and 3D spatial density considerations remain future work; tested mostly on ORB-SLAM2 pipelines.
(Niculescu et al., 2024)	Enables fully onboard SLAM for nano-UAVs with ultra-low power (87.9mW); accurate indoor mapping (4.5 cm); efficient ICP scan matching in 55ms.	Lacks deep learning integration for object recognition; path planning based on generated maps yet to be integrated.

Continued on next page

Table 1 – continued from previous page

Paper / Source	**Strengths**	**Weaknesses / Gaps**
(Tukan et al., 2023)	Novel outlier removal for sparse ORB-SLAM3 data; converts sparse points into dense maps with clustering; effective real-time indoor navigation.	Focus on toy drones; scalability and adaptation to more resource-limited platforms is future work.
(Chakrabarty & Sarangi, 2024)	Fast and accurate non-ML depth image rectification for edge devices; 31% PSNR improvement; runs at 27 FPS on NVIDIA Jetson Nano.	Focused only on depth image rectification; broader SLAM integration and testing in diverse scenarios needed.
(Davletshin et al., 2024)	Uses DeblurGANv2 to enhance RGB frames, reducing motion blur; improves 3D reconstruction and object segmentation for UAVs in dynamic settings.	Initial position and object correction need optimization; multi-agent SLAM not yet explored.
(Merat et al., 2024)	Aligns VSLAM output with city digital twins to eliminate drift; achieves robust 6-DoF global localization; outperforms VIO-GPS systems.	Susceptible to 3D aliasing and errors from digital twin or local cloud inaccuracies.
(Tyagi & Gaur, 2025)	Integrates ORB-SLAM3, LQR control, fault detection, and onboard vision on Raspberry Pi 4; demonstrated robustness in GPS-denied environments.	Vision module speed limited (2 FPS); hardware acceleration and stereo vision integration planned.

Continued on next page

Table 1 – continued from previous page

Paper / Source	Strengths	Weaknesses / Gaps
(Li et al., 2025)	First online Gaussian Splatting SLAM handling dynamic objects; fast mapping and localization; outperforms existing GS-SLAM in dynamic scenes.	Motion model complexity limited for real-time performance; exploring more complex models is future work.
(Sartori et al., 2025)	AI-based split computing for vision and planning on nano-drones; real-time obstacle avoidance with 8 FPS; good detection accuracy (60.8% mAP).	Limited top speed and obstacle avoidance success rates; multi-point navigation and trajectory optimization planned.
(Xiao et al., 2025)	Integrates LLMs and VLMs for semantic and visual path planning; high success rates (¿90%) even in complex obstacle scenarios.	Average reasoning time (9 sec) may limit high-frequency replanning; real-time efficiency improvements needed.
(Burke, 2025)	AI fully generates drone command/control software (10k lines) in 2.5 weeks, drastically reducing human effort; runs onboard with stable Wi-Fi.	Current AI struggles with larger codebases and complex info flow; safety and reliability of AI-generated code need research.

8 Discussion

The current state of drone technology and its applications, while advanced, faces several persistent challenges and limitations that are key areas for future development. Dynamic and complex environments remain a significant hurdle for SLAM systems, with issues such as managing dynamic Gaussian splats and rejecting persistent false positives from temporarily static objects. Computational and memory constraints severely limit onboard processing for tiny robots, often necessitating offloading computation to external infrastructure, which can introduce latency and security vulnerabilities. Visual-based perception systems, while promising, are still challenged by the large data volumes generated by high-resolution cameras and the limited processing power on nano-UAVs, making depth-sensing alternatives often more practical. Path planning, particularly when relying on sparse point clouds from SLAM systems like ORB-SLAM3, struggles to accurately differentiate true obstacles from noisy data, potentially leading to collisions if robust outlier removal and data transformation are not employed. Learning-based approaches for drone navigation, including deep learning and foundation models, currently lack generalization to unseen scenarios and require extensive domain-specific data, with the practicality and robustness of large language models (LLMs) and vision-language models (VLMs) in global path planning remaining largely unexplored, often lacking robust spatial reasoning and seamless real-time perception integration. Furthermore, despite breakthroughs in AI-generated drone control systems, the rigorous testing and safety validation of AI-authored code are still open problems, especially for high-risk operations, and current AI models are limited by context windows and reasoning depth.

9 Conclusion

In conclusion, from 2020 to 2025, drone autonomy has transitioned from predominantly simulated or controlled environments to deployable, intelligent systems capable of real-time navigation, obstacle avoidance, and specialized tasks in GPS-denied and dynamic real-world scenarios. Concurrently, SLAM has witnessed significant advancements in handling dynamic objects, improving efficiency through map sparsification, enhancing visual data quality through AI-driven deblurring, and successfully adapting for fully onboard execution on tiny, power-constrained platforms. Researchers now possess a richer understanding of the intricate trade-offs between computational resources, sensor modalities, and specific application requirements, which is actively propelling the boundaries of autonomous drone capabilities in increasingly complex and demanding settings. Future research is poised to further refine these capabilities, with a particular emphasis on achieving more robust generalization in AI models, comprehensive 3D dynamic mapping, and optimizing multi-agent systems for broader coverage, increased robustness, and collaborative intelligence.

10 Acknowledgments

I am deeply grateful to the Nepal Research and Collaboration Center (NRCC) for providing me with the opportunity to help me learn to conduct research. This experience has been invaluable, allowing me to significantly expand my knowledge and develop new skills that will be foundational for my future career.

References

Burke, P. (2025). *Robot builds a robot's brain: Ai generated drone command and control station hosted in the sky.* Retrieved from `https://arxiv.org/abs/2508.02962`

Chakrabarty, Y., & Sarangi, S. R. (2024). *Voxdepth: Rectification of depth images on edge devices.* Retrieved from `https://arxiv.org/abs/2407.15067`

Davletshin, D., Zhura, I., Cheremnykh, V., Rybiyanov, M., Fedoseev, A., & Tsetserukou, D. (2024). *Sharpslam: 3d object-oriented visual slam with deblurring for agile drones.* Retrieved from `https://arxiv.org/abs/2410.05405`

Li, R. B., Shaghaghi, M., Suzuki, K., Liu, X., Moparthi, V., Du, B., . . . Nguyen, T. (2025). *Dynagslam: Real-time gaussian-splatting slam for online rendering, tracking, motion predictions of moving objects in dynamic scenes.* Retrieved from https://arxiv.org/abs/2503.11979

Merat, R., Cioffi, G., Bauersfeld, L., & Scaramuzza, D. (2024). *Drift-free visual slam using digital twins.* Retrieved from https://arxiv.org/abs/2412.08496

Niculescu, V., Polonelli, T., Magno, M., & Benini, L. (2024, April). Nanoslam: Enabling fully onboard slam for tiny robots. *IEEE Internet of Things Journal*, *11*(8), 13584–13607. Retrieved from http://dx.doi.org/10.1109/JIOT.2023.3339254 doi: 10.1109/jiot.2023.3339254

Park, Y., & Bae, S. (2022). *Keeping less is more: Point sparsification for visual slam.* Retrieved from https://arxiv.org/abs/2207.00225

Sartori, M., Singhal, C., Roy, N., Brunelli, D., & Gross, J. (2025). *Ai and vision based autonomous navigation of nano-drones in partially-known environments.* Retrieved from https://arxiv.org/abs/2505.04972

Song, S., Lim, H., Lee, A. J., & Myung, H. (2022). *Dynavins: A visual-inertial slam for dynamic environments.* Retrieved from https://arxiv.org/abs/2208.11500

Tukan, M., Fares, F., Grufinkle, Y., Talmor, I., Mualem, L., Braverman, V., & Feldman, D. (2023). *Orbslam3-enhanced autonomous toy drones: Pioneering indoor exploration.* Retrieved from https://arxiv.org/abs/2312.13385

Tyagi, A., & Gaur, C. (2025). *Slam-based navigation and fault resilience in a surveillance quadcopter with embedded vision systems.* Retrieved from https://arxiv.org/abs/2504.15305

Xiao, J., Tsao, C. W., Zhang, Y., & Feroskhan, M. (2025). *Fm-planner: Foundation model guided path planning for autonomous drone navigation.* Retrieved from https://arxiv.org/abs/2505.20783