

Fusion Algorithms and Anti-Interference Stability Optimization for SLAM in Inspection UAVs under Complex Occlusion Conditions

Chengzhan Li^{1,*}, Rui Peng²,

{lotush200625@gmail.com^{1,*}}

Mechanical Design and Automation, Shanghai Normal University, Shanghai, 201418, China¹,
Artificial Intelligence, Macau University of Science and Technology, Macau, 999078, China²

Abstract. To solve the positioning drift and control response inaccuracy problems that are prevalent in Simultaneous Localization and Mapping (SLAM) of inspection UAVs in complex occlusion and multi-interference coupled situations, this paper systematically reviews multi-sensor fusion SLAM algorithms and their anti-interference stability optimization. It first examines the technical aspects and shortcomings of visual SLAM and LiDAR SLAM and covers multi-sensor fusion schemes like vision-LiDAR and LiDAR-IMU. Secondly, this paper summarizes a collaborative framework on the basis of filtering-based state sharing, model predictive control (MPC) integration, and event-triggered mechanisms to resolve the problem of inadequate coordination between SLAM and the control system. Thirdly, it concludes an anti-interference plan which incorporates both conventional robust control and learning-reinforced control. The results of this review offer technical assistance to enhance the working reliability of inspection drones in complex environments.

Keywords: Drones; Multi-sensor fusion; Anti-interference; SLAM technology

1 Introduction

As the ever-growing demand of global infrastructure operation and maintenance grows, unmanned aerial vehicle (UAV) autonomous inspection technology has emerged as a central support technology in areas like power, transportation, and forestry. Its operational reliability is directly reliant on two core modules: The first one is Simultaneous Localization and Mapping (SLAM) technology which breaks through the limitations of traditional navigation systems, allowing robots to do real-time self-localization in unknown environments and simultaneously create environmental maps, significantly enhancing the autonomy and adaptability of the robot. The effective implementation of SLAM technology does not only rely on efficient algorithms, but also on the collaborative efforts of various sensors [1]; The second is a strong (anti-interference) control system is another fundamental aspect to guarantee continuous and reliable functioning. According to the variations in the features of sensing devices, the SLAM

technology system may be classified into two significant branches lidar SLAM and visual SLAM [2].

In real inspection scenarios, complex obstructions such as those from power transmission lines and trees can lead to unreliable environmental perception. Uncertainties caused by sudden strong interference such as electromagnetic radiation from substations and airflow disturbances often cause traditional SLAM systems to experience positioning drift and inaccurate control system response, which seriously affects the autonomous operation effect of UAVs [3]. Therefore, studying multi-sensor fusion SLAM positioning technology and achieving its synergistic optimization with the anti-interference stability of the control system is of great theoretical and practical value for improving the adaptability and operational reliability of inspection UAVs in complex environments.

Based on the existing literature, visual SLAM (such as ORB-SLAM3) relies on texture features to achieve localization, while laser SLAM (such as Cartographer) relies on geometric information to ensure the robustness of mapping. However, single-sensor solutions all have inherent defects of poor environmental adaptability. Multi-sensor fusion SLAM improves robustness by integrating complementary information such as vision, LiDAR, and IMU [4], and has become a current research hotspot. Internationally, the focus is on the optimization of fusion architecture and the lightweighting of algorithms, while domestically, the focus is more on the customized adaptation of inspection scenarios, such as feature-level fusion schemes for power transmission lines. However, existing research still has obvious shortcomings: on the one hand, the weight allocation of SLAM fusion is mostly a fixed mechanism, which cannot dynamically adapt to the changes in the confidence of sensor data under occlusion scenarios; on the other hand, SLAM positioning and control systems are often designed independently, and insufficient coordination leads to asynchronous state estimation delay and control response. Moreover, anti-interference technology is mostly aimed at a single interference source, which is difficult to cope with multi-interference coupled scenarios; at the same time, cross-modal feature association is easily affected by occlusion interference, and the feature extraction efficiency for key targets such as power transmission lines is insufficient. All of these problems restrict the improvement of the overall system performance.

The core research content in this field mainly focuses on three directions: First, the optimization of SLAM fusion technology, designing a dynamic weight allocation mechanism adapted to complex occlusion scenarios to improve positioning accuracy; second, the UAV SLAM and control coordination strategy, constructing a positioning-control integrated framework to solve the problem of asynchronous response between the two; and third, anti-interference stability optimization methods, establishing a multi-interference collaborative suppression strategy to enhance system robustness.

The significance of this research is reflected in two aspects: at the theoretical level, it provides targeted solutions for improving the coordination and anti-interference stability of UAV SLAM and control in complex scenarios, and enriches the theoretical system of UAV autonomous navigation; at the practical application level, it can effectively improve inspection efficiency, reduce the risk of high-risk environment operations, provide technical support for the engineering implementation and performance iteration of inspection UAVs, and help upgrade the intelligent operation and maintenance of infrastructure.

2 SLAM technology based on the fusion of multiple systems

2.1 SLAM methods for vision and lidar

Visual SLAM. Visual SLAM has become the mainstream SLAM method due to its advantages of simple structure, low cost and extraction of semantic information. UAVs using vision as the main sensor need to complete map fusion based on the common field of view. In recent years, related research has focused on improving the efficiency, accuracy, robustness and compatibility of the algorithm. Deep learning methods have also been gradually integrated into swarm SLAM systems, mainly in implicit map representation and semantic information integration [5]. Visual SLAM is divided into two categories according to dimension: 2D and 3D. 2D-SLAM commonly uses monocular or binocular cameras, while 3D-SLAM relies on stereo cameras and RGB-D cameras to obtain depth information. Its core algorithms can be divided into three categories: feature point-based, direct method-based and deep learning-based visual SLAM [6].

Laser SLAM. Compared with visual information, LiDAR can directly acquire distance and orientation information of the environment, which has significant advantages in complex scenes and is more suitable for the point cloud map construction requirements of active swarm SLAM. At present, deep neural networks (DNN) based on point cloud information are developing rapidly and are gradually being applied to swarm SLAM systems [5]. LiDAR SLAM is categorized into 2D and 3D. The distance information of high precision offered by LiDAR allows it to be accurate and stable in localization and mapping in complex environments [6].

2.2 Fusion SLAM technology from different sensors

SLAM technology is based on the input of information provided by several sensors to perceive the environment and position accurately. Nevertheless, current single-sensor SLAM systems tend to be unreliable in terms of system robustness, low environmental perception flexibility, and lack of data analysis functionality, which makes it challenging to sustain the long-term performance needed to conduct UAV inspections. Thus, sensor fusion technology is important in SLAM systems. The disadvantages of various sensors can be offset by integrating the benefits of various sensors. Complementary or redundant information may enhance the strength and precision of the system [1]. Over the past few years, there has been a lot of interest in fusion SLAM that is based on lidar and vision modules as opposed to coupling IMU modules to enhance their respective robustness and accuracy.

Visual sensor and LiDAR fusion SLAM. Visual SLAM and laser SLAM have achieved remarkable results in their respective application fields, but single-sensor solutions have inherent limitations. Visual SLAM obtains accurate correspondences by matching image feature points, but RGB-D cameras are sensitive to changes in lighting in outdoor environments and have difficulty effectively extracting scene depth information. Laser radar SLAM is based on scanning matching methods and has high precision and strong depth information acquisition capabilities, but it is prone to generating "ghost points" in dynamic scenes, and has problems such as insufficient stability and unclear environmental features. The close coupling of the two can achieve complementary advantages and make up for their respective defects, which has become a key focus of researchers [1].

With the development of SLAM technology, related methods for the fusion of vision and lidar have emerged. Some studies have proposed a scheme to integrate lidar data and visual information [7]. First, lidar point cloud data is processed by Rao-Blackwellized particle filtering technology to estimate pose, then the camera is used to capture real-time three-dimensional point cloud information of the environment, and finally the Bayesian method is used to perform map fusion and update on the similar features extracted from the two types of point clouds to achieve high-precision environmental map construction [2].

The core premise of the fusion of lidar and visual sensors is the measurement of information of the same object at the same time. Therefore, the calibration and fusion of data from different sensors is the key issue of this technology. Li et al. proposed a tightly coupled integration method of PPP, inertial navigation, vision and lidar, which directly fuses sparse visual landmarks and lidar edge features through a centralized extended Kalman filter, and uses two-point RANSAC and spatiotemporal loop matching mechanism to remove outliers [8].

LiDAR and IMU fusion SLAM. As a precision sensing device, LiDAR is based on the transmission and reception of laser pulses. As a core component of motion sensing system, IMU is specifically designed for the perception and measurement of acceleration, attitude orientation and rotation parameters in three-dimensional space [9]. LiDAR uses IMU fusion for assisted positioning, thereby solving the problems of distortion and low update rate caused by motion in LiDAR SLAM. In recent years, a number of high-precision and real-time LiDAR fusion inertial measurement unit SLAM studies have emerged [1]. Hassan et al. proposed a method for large-scale three-dimensional mapping of construction cranes by attaching 2D LiDAR and IMU to the crane arm, and introduced a complementary filter with moving average filtering function for LiDAR attitude estimation, which has strong robustness to severe vibration [10]. The scanning and local map are matched by building a local map and pre-integrating the inertial measurement unit (IMU), and the IMU measurement values are obtained. The joint cost function is minimized to get odometry information, and key frames pose is optimized with radar-inertial odometry .

Other fusion schemes and adaptive fusion. In addition to vision-LiDAR and LiDAR-IMU fusion, there is also a popular and effective solution of vision-IMU fusion SLAM. This technique successfully addresses the issue of tracking loss in pure vision systems under fast motion or texture loss through the close integration of high-frequency motion information provided by the IMU with geometric measurements provided by the vision system. It further compensates the integration drift of the IMU using pre-integration methods and nonlinear optimization which has greatly enhanced robustness and accuracy of the system in dynamic conditions.

Moreover, to handle very complicated situations like substations and forests, researchers are investigating heterogeneous architectures that integrate more sensors (GNSS and barometers). The conventional multi-sensor fusion tends to use fixed weights or strategies which find it hard to respond to real-time confidence variations in the sensor data in dynamic situations where there are multiple coupled factors like occlusion and interference. Dynamic adaptive fusion mechanisms have emerged as a significant research line. These techniques seek to modify the fusion weights of various sensor information online depending on environmental features (e.g., texture richness and geometric structure), motion state, or estimation uncertainty. As an illustration, when vision is blocked, the weight of vision can be lowered and the contribution of

IMU or vision can be increased when laser point cloud degrades, thus making intelligent adaptation to complex environments and enhancing the robustness of the system as a whole.

2.3 SLAM optimization examples for complex environments

The above multi-sensor fusion architecture and methods give a theoretical basis and technical strategy to enhance the SLAM performance of UAVs in complex environments. In real engineering applications, researchers have suggested more specific optimization schemes to address the challenges of particular complex situations (such as wild or forest environments with undulating terrain, dense vegetation, and strong motion interference), validating the effectiveness of the adaptive, tightly coupled fusion framework.

To overcome the problem of point cloud deformation and positioning drift due to extreme conditions like complex terrain and violent motion, researchers have suggested several tightly coupled and adaptive optimization methods. The mainstream method usually uses inertial navigation system (INS)-assisted motion compensation techniques (like adaptive frame segmentation with IMU pre-integration) at the front end to correct point cloud distortion; and uses tightly coupled filtering (like iterative error state Kalman filtering, IESKF) to integrate laser and IMU data to get highly robust initial estimation. At the back end, a factor graph optimization framework is used to combine multi-source observations such as laser inertial odometry, closed-loop detection, and GNSS for global consistency. For example, Lu Chunxiao et al. used the above techniques in combination and verified their effectiveness in scenarios such as wild forests [11]. In general, this kind of technical route of "front-end motion compensation and tight coupling + back-end multi-source factor graph optimization" has become an effective paradigm for improving the robustness of SLAM in complex environments.

3 Positioning control coordination and robust optimization

3.1 Co-design of SLAM and control system

In the conventional UAV autonomous inspection architecture, the SLAM module and flight controller usually take a perception-decision-execution serial pipeline design, which runs separately at both software and hardware levels. Although this decoupled architecture is helpful in developing modules and debugging them, it will inevitably result in delays of state estimation and control command lags. As an example, where visual features are lost because of intense reflections or total occlusion in regions with high power towers or within tunnels, the SLAM system can lose pose tracking temporarily. The controller will cause trajectory deviations and even collision accidents when it continues to make feedback adjustments on outdated information. To overcome this, researchers have suggested different collaborative frameworks:

Filter-based state sharing mechanism. Filter methods are able to estimate the state and manage uncertainty by updating and estimating system states recursively, extracting true states out of noisy observation data [1]. Multi-source sensor information (IMU, vision, and lidar) is combined using Extended Kalman Filter (EKF) or Unscented Kalman Filter (UKF) to provide a uniform estimate of the position, velocity, attitude, and covariance of the UAV. The entire state vector is fed into the underlying PID or LQR controller in real time enhancing the consistency of the state and speed of response of the closed-loop system [12]. This approach

has been implemented on engineering level on PX4 flight control platform and ROS system. In nonlinear filtering, UKF is able to deal with more cases that have high-order statistical properties. It does not minimize the number of computational steps but it is simpler to implement and its technical development has become a research hotspot in the SLAM field. Zhang et al. suggested a better featureless Kalman filter algorithm using Radial Basis Function-Neural Network (RBF-NN), which can correct deviations of UKF filtering with a constant coefficient model, successfully addressing the issue of weak signal extraction under complex noise and considerably enhancing the accuracy of the system.

Model predictive control (MPC) integration. MPC is an online optimization of control inputs over a future time period, which makes it inherently appropriate to constrained dynamic systems. Recent work uses the pose sequence and environmental map produced by SLAM as state observations and obstacle avoidance constraints for MPC, achieving perception-based replanning of trajectories. For example, in forest inspection applications, MPC can adaptively update paths using real-time point cloud maps to avoid sudden tree branch obstacles. However, this approach requires high computing resources and must balance algorithm efficiency and control performance.

Event-triggered mechanism. Unlike the conventional fixed-frequency control, event-triggered mechanism updates the control commands only when system state error is above a certain level (e.g. position deviation $> 0.3\text{m}$ or abrupt change in attitude angle change rate). This architecture significantly decreases the communication burden between SLAM and the controller, eliminates unnecessary movements of actuators, and increases battery life which is especially applicable to bandwidth-limited or low-power edge computing situations.

3.2 Strategies for enhancing anti-interference stability

With UAV missions expanding into complex environments, including urban canyons, mountain valleys, and high-voltage electric fields, the intensity and uncertainty of external disturbances (e.g., gusts, electromagnetic interference, and airflow vortices) have grown significantly. Anti-interference capability has therefore become a core measure for evaluating the robustness and safety of autonomous inspection systems. At present, research mainly follows two technical directions:

Traditional robust control methods. The traditional robust control techniques such as H-infinity control, sliding mode control and adaptive control are the classic early solutions in the area of UAV flight control. The H₂-infinity control is a frequency domain analysis and linear matrix inequality tool which reduces the norm between disturbance input and performance output using an augmented model to achieve closed-loop stability even under worst-case disturbances and is applicable at low-to-medium speed cruise. Sliding mode control depends on discontinuous switching control laws that enable quick convergence of system state towards the sliding surface, being invariant to matched disturbances and parameter variations, and is appropriate to high-stability demands like hovering and maneuvering flight. Adaptive control is dynamic in its nature since it modifies the control law through online estimation of unknown parameters or disturbance limits, which has proven to be very effective when dealing with time-varying parameter problems due to load variation and fuel usage.

These techniques have a full stability proof mechanism and mature engineering implementation procedures, and have been used in attitude control and trajectory tracking of fixed-wing and

multi-rotor UAVs. But their drawbacks are evident: they require correct dynamic models of UAVs, the simplification of models or mismatch of parameters will directly influence the effects of control; sliding mode control is susceptible to chattering which further aggravates actuator wear and excites unmodeled dynamics; the convergence speed of adaptive control parameter and accuracy of control cannot be balanced, and does not satisfy the real-time demands of fast maneuvering situations.

Learning-enhanced control. In order to address the model dependence and lack of flexibility to unknown disturbances of traditional robust control, researchers have strongly integrated data-driven approaches like reinforcement learning, imitation learning, and transfer learning with control policies to create learning-enhanced control schemes, which has become a significant research direction in UAV anti-interference control in complex environments. This is not an approach that needs pre-establishment of a correct dynamic model. The underlying concept is to develop a multi-condition disturbance database within a high-fidelity simulation environment with different wind speed perturbations, electromagnetic interference, and terrain occlusion, and train intelligent agents including Deep Q-Network (DQN) and Proximal Policy Optimization (PPO). These agents learn attitude adjustment, trajectory correction, and dynamic compensation strategies on their own by constantly interacting with the environment to real-time counteract the effects of complex disturbances. In the meantime, domain adaptation, meta-learning, and other technologies are employed to minimize the distribution gap between simulation and real world so as to effectively transfer simulated models to real platforms [13], and obtain more effective anti-interference results in unknown situations like urban canyon turbulence and high-voltage electric field electromagnetic interference.

In comparison to traditional approaches, learning-enhanced control has robust online adaptive potential. It is able to optimize the strategies of control by interacting and incrementally learning continuously, dynamically adjusting to alterations in disturbance properties; a few studies have proposed a hierarchical reinforcement learning structure that separates high-level task planning and low-level control implementation, which enhances the speed and accuracy of responses in cases with multiple coupled disturbances. Also, certain scholars have developed a hybrid framework that is learning-robust, where the learning module is used to identify nonlinear features of unknown disturbances and the robust module is used to maintain basic system stability, balancing adaptability and safety redundancy.

Nevertheless, this technical path is still confronted with numerous issues: first of all, low sample efficiency. To train high-performance intelligent agents, it takes a lot of simulated or actual samples of flight which leads to long research and development cycles and high costs. In addition, the gathering of real-scenario samples is restricted by safety in flying and environmental factors; second, lack of safety and interpretability. The intelligent models decision-making process is a black box process, and it is hard to trace and interfere with abnormal decisions, which presents risks of application in tasks of high safety levels; third, poor generalization capacity in real environments. Simulation and real situations have inherent differences in disturbance distribution and physical properties, which may result in performance degradation following transfer of the model. Hardware-in-the-loop testing, incremental learning, and other techniques are required to enhance generalization performance; fourth, high computational resource consumption. Deep reinforcement learning models inference needs high computing power support and cannot be easily implemented directly on small UAV embedded

platforms that do not have much computing power. Light-weight design and acceleration technologies of inference have become critical bottlenecks in engineering implementation.

3.3 Application examples of collaborative anti-interference in typical scenarios

The above collaborative frameworks and anti-interference strategies should be tested on particular complex engineering situations. The autonomous inspection of a substation is an example of a high-risk and strong-interference application scenario of UAVs. It has heavy internal machinery and complicated electromagnetic conditions, which present extreme difficulties to the online perception, real-time planning, and anti-interference control of UAVs, making it an ideal test platform to confirm the engineering worthiness of the integrated collaborative strategy of perception-planning-control.

In the substation case, path planning must avoid static obstacles, deal with the effect of electromagnetic interference on sensor data, and meet strict flight safety requirements. Traditional global map - based path planning algorithms are not up to dynamic updates and real - time obstacle avoidance. Relying solely on reactive collision avoidance is hard to ensure global optimality and task efficiency. Under the collaborative framework, researchers have closely integrated high - precision environmental maps built in real - time or preloaded by the SLAM system with advanced path - planning algorithms to improve planning efficiency and flight safety. Wu Tian et al. put forward the IS-RRT intelligent optimized Rapidly Exploring Random Tree algorithm for complex substation environments [14]. The effectiveness of this algorithm is founded on collaboration with the perception system.

The experimental findings indicate that the collaborative scheme of perception and planning is much better than the conventional single planning algorithms in the main indicators like path length, speed of planning, and the number of nodes of a path which is an effective solution to the problem of path planning in complex confined spaces. This team-based mode creates a full circle loop between SLAM environmental perception and planning algorithm decision optimization and trajectory tracking by underlying anti-interference controller. It not only enhances performance of each module through collaboration adaptation among modules but also increases the overall operational reliability, safety and efficiency of UAVs in extreme situations such as substations, which gives a successful example of how the collaborative and anti-interference theories can be applied in practice in typical industrial situations through collaborative adaptation among modules and system-level overall optimization.

4 Current challenges and future development trends

4.1 Main existing challenges

Dynamic occlusion leads to nonlinear changes in the confidence of sensor data at the perception level. Most existing fusion algorithms employ fixed weights, and do not provide real-time and accurate assessment of sensor reliability and dynamic fusion mechanisms resulting in many failures in cross-modal feature association. Data-driven learning-enhanced control is problematic at the decision-making and control level because it has a black box issue, lack of interpretability and safety, and this makes it hard to trust completely on safety-critical inspection activities. At the system and verification level, simulation environments and real situations are

very different. The absence of standard test datasets and benchmarks with typical complex disturbances lead to a gap between the performance of an algorithm in simulation and its performance in practice that does not support fair evaluation and iteration of technologies.

4.2 Future development trends

Integrated intelligent architectures. The future will be characterized by a closed system of perception-decision-control, which is end-to-end and integrated with the previous modular approach that was fragmented in its own right. In an attempt to use large models (Transformers) as the end-to-end multimodal unified modeling, different input types (vision, LiDAR, IMU) can be transformed into the same semantic representation space to combine them and associate them. At the same time, state estimation of SLAM, control commands, and anti-interference compensation measures can be integrated into a single model. High real-time performance and system stability can be guaranteed via Time-Sensitive Networking (TSN), which enables millisecond-level time synchronization.

Environment-adaptive fusion mechanism. The future systems will become more and more dependent on environment-adaptive fusion designs which are able to dynamically change the weights of sensors in response to the quality of data of the sensor (e.g., visual feature matching rates, effective LiDAR point cloud content, and IMU noise) that is being monitored in real time. The establishment of dynamic correlation models among occlusion rate, interference intensity and sensor reliability is essential for adaptive multi-sensor weight optimization. As an example, these types of mechanisms would also be used to raise the weight of the LiDARs when a wire was occluded and compensate the filter compensation with the IMU in cases of electromagnetic disturbance, so that the system could be much better adapted to itself in the case of dynamic and complex environments.

Safe and reliable learning-based control system. One of the most important future developments will be a hierarchical, safe, and robust control scheme based on a deep learning approach with classical control theory where the bottom layer of the control is based on a combination of deep learning and the upper one is based on learning compensation. The bottom level of this architecture has to use the algorithm of the robust control (integral sliding mode control) with rigorous stability proofs in order to provide the basic safety. Simultaneously, the upper layer would be trained using the deep reinforcement learning technique to counteract unstructured and uncertain perturbations. The integration of interpretable algorithms (e.g., attention mechanisms and causal reasoning) can significantly improve the interpretability of model learning processes. A full-scenario validation framework combining digital twin technology and hardware-in-the-loop tests is required to validate the practical effectiveness, safety and robustness of learning-based control methods in actual inspection scenarios.

Standardized testing and open-source ecosystem. Specialized datasets and hardware-in-the-loop testing systems for typical inspection scenarios (e.g., substations, forests, and mountainous areas) need to be constructed to support algorithm iteration and optimization. Industrial open-source frameworks are urgently required to promote the industrialization of inspection technologies. The continuous iteration of SLAM, control algorithms and anti-interference technologies can be effectively accelerated via standardized resource libraries. The construction of a comprehensive testing ecosystem can provide strong support for the large-scale deployment of UAV inspection systems.

5 Conclusion

The perception accuracy and control stability of inspection drones are challenged by complex environments with complicated obstructions and multiple interferences. Conventional single-sensor SLAM systems tend to be prone to positioning drift, and the separate design of the perception system and the control system can result in incorrect responses, which is a serious limitation on the reliability of autonomous drone operations.

To overcome these issues, multi-sensor fusion SLAM technology has emerged as a fundamental solution. In this paper, how a multi-modal fusion architecture that combines vision-LiDAR and LiDAR-IMU with filtering and factor graph optimization techniques is reviewed, can effectively compensate the environmental adaptability limitations of single sensors and significantly improve perception robustness in complex situations. Moreover, a collaborative framework based on filtering-based state sharing and model predictive control integration allows deep integration between SLAM and the control system. Together with an anti-interference strategy that integrates traditional robust control with learning-reinforced control, the overall system resilience is further enhanced. Its practical value has been verified in common inspection scenarios such as substations and rough terrain.

The traditional robust control has well-developed stability proofs and good engineering application base, whereas the emerging learning-reinforced control has high adaptability to unknown disturbances. Both of them have their benefits and major complementarity, which is why integrated innovation is one of the directions in the technological development. In the future, there are four fundamental areas that need constant breakthroughs, including creating an integrated intelligent architecture of perception-decision-control, enhancing environmental adaptive fusion mechanisms, creating safe and reliable learning-based control systems, and creating standardized test datasets and open-source ecosystem. This will make inspection drones more adaptable in very complex situations and offer a strong technical backing to the smart modernization of infrastructure operation and maintenance.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

- [1] Qiao, R.: A review of SLAM technology development from a sensor perspective. *Softw. Guide*. **24**, 67 (2025)
- [2] Yang, H., Liang, Y., Yu, H., Ji, X., Zeng, X.: Robot SLAM technology based on lidar and vision fusion. *Modern. Inf. Technol.* **9**, 194 (2025)
- [3] Campos, C., Elvira, R., Rodríguez, J., Montiel, J., Tardós, J.: Orb-slam3: An accurate open-source library for visual, visual-inertial, and multi-map SLAM. *IEEE Trans. Robot.* **37**, 1874 (2021)

- [4] Qin, T., Li, P., Shen, S.: Vins-mono: A robust and universal monocular visual inertial state estimator. *IEEE Trans. Robot.* **34**, 1004 (2018)
- [5] Yu, Z., Wu, X., Liu, H., Zhu, K., Liu, Y., Zhang, C., Tu, R.: Review of advances in UAV vision/LiDAR cluster SLAM technology. *Navig. Position. Timing.* **12**, 15 (2025)
- [6] Cheng, L., Luo, Z., He, Y., Chen, Y.: Review and outlook on visual and laser SLAM development. *China Testing* **51**, 1 (2025)
- [7] Ma, Q., Wang, J., Zhang, J., Zou, Z.: Three-dimensional mapping of industrial parks through synchronous fusion of lidar and inertial measurement units. *Opt. Precis. Eng.* **32**, 422 (2024)
- [8] Li, S., Li, X., Wang, H., et al.: Multi-GNSS PPP/INS/vision/LiDAR tightly integrated system for precise navigation in urban environments. *Inf. Fusion* **90**, 218 (2023)
- [9] Xia, L., Zhang, B., Wang, Y.: Robot positioning based on inertial sensors and visual odometry. *Chin. J. Instrum. Meas.* **34**, 166 (2013)
- [10] Hassan, M., Das, D., Miura, J.: 3D mapping for a large crane using rotating 2D-LiDAR and IMU attached to the crane boom. *IEEE Access.* **11**, 21104 (2023)
- [11] Lu, C., Zhong, H., Liu, W., Zhou, Y., Cui, Z., Li, W.: Multi-sensor fusion SLAM technology in complex terrain environments. *Robot* **46**, 425 (2024)
- [12] Chen, Y., Huang, S., Fankhauser, P.: Autonomous navigation in complex environments based on deep vision inertial fusion and model predictive control. *IEEE Robot. Autom. Lett.* **6**, 2159 (2021)
- [13] Koch, W., Özgüner, Ü.: Reinforcement learning for UAV control: A survey. *Annu. Rev. Control.* **52**, 197 (2021)
- [14] Wu, T., Wang, Y., Xu, G., Qiu, Z., Pu, Z., Li, P.: Path planning for inspection UAVs in complex substation environments based on IS-RRT*. *J. Mech. Des.* **1** (2025)