Investigation into Vision-Based Navigation Systems for Unmanned Aerial Vehicles

Tianhao Huang

North Alabama International College of Engineering and Technology, Guizhou University, Guiyang, Guizhou, 550000, China

cme.thhuang22@gzu.edu.cn

Abstract. With the rapid advancement of unmanned aerial vehicle (UAV) technology, the limitations of traditional Global Positioning System (GPS) in complex low-altitude environments have become increasingly apparent, driving research demands for vision-based navigation technologies. This paper systematically investigates the fundamental principles, sensor types, and mainstream algorithms of UAV visual navigation systems. Through analysis of practical cases, solutions are proposed to enhance UAVs' autonomous navigation and intelligent decision-making capabilities in complex environments. The study concludes that vision-based navigation systems enable efficient data acquisition and processing in challenging low-altitude scenarios, effectively compensating for GPS deficiencies. To address these issues, the paper proposes solutions such as multimodal sensor fusion, algorithm structure optimization, and parallel computing resource utilization. Future research on vision-based navigation should focus on algorithm optimization and innovation to reduce computational complexity and improve system adaptability. This work aims to address critical challenges in UAV visual navigation systems and provides perspectives on improving real-time performance and positioning accuracy for future developments.

Keywords: Vision-Based Navigation, Unmanned Aerial Vehicles, Autonomous Navigation, Deep Learning

1 Introduction

The rapid development of unmanned aerial vehicle (UAV) technology has enabled its widespread application across diverse domains. However, in low-altitude environments, Global Positioning System (GPS) signals are highly susceptible to degradation or distortion due to building obstructions, vegetation coverage, or multipath effects, significantly compromising UAVs' positioning accuracy and operational stability. In urban areas with dense high-rise building clusters, GPS signals face severe penetration challenges, rendering them unreliable for precise UAV navigation during mission execution. These limitations in complex low-altitude scenarios have driven urgent demands for vision-based navigation methods. As a

J. Iqbal (ed.), *Proceedings of the 2025 2nd International Conference on Mechanics, Electronics Engineering and Automation (ICMEEA 2025)*, Advances in Engineering Research 272,
https://doi.org/10.2991/978-94-6463-821-9_21

critical technology enabling autonomous navigation and intelligent decision-making, vision-based navigation systems mimic biological visual perception mechanisms to efficiently acquire and process environmental data, thereby achieving precise localization and path planning even under GPS-denied conditions.

Global research in UAV vision-based navigation has achieved substantial progress. For instance, Pintér et al. [1] investigated visual odometry (VO) for UAV position estimation; Wang Lianyou et al. [2] analyzed improved Visual-Inertial Navigation Systems (VINS) for indoor localization and autonomous navigation; Zhao Chunhui et al. [3] comprehensively reviewed scene-matching visual navigation techniques for UAVs. Nevertheless, persistent challenges remain in ensuring navigation stability in complex environments, optimizing computational efficiency for real-time performance, and developing robust multi-sensor data fusion strategies, all of which require further systematic investigation.

This paper aims to analyze the principles of UAV vision-based navigation systems and propose solutions to existing challenges. First, this paper systematically elaborates on the fundamental principles of vision-based navigation, including visual perception mechanisms, localization frameworks, and navigation paradigms, while critically evaluating the characteristics and application scenarios of diverse visual sensors. Subsequently, this paper conducts an in-depth analysis of mainstream algorithms such as VO, Simultaneous Localization and Mapping (SLAM), VINS, and deep learning-based approaches, discussing their advantages, limitations, and implementation considerations. Building upon this foundation, this paper proposes targeted solutions to address current technical bottlenecks. Finally, through practical case studies, this paper demonstrates the efficacy of vision-based navigation in environmental perception, obstacle detection, and target tracking/recognition, while outlining future research directions to enhance real-time responsiveness and operational precision.

2 Overview of Vision-Based Navigation Technology

With the continuous advancement of UAV technology, vision-based navigation systems, as its core component, have gradually become a research hotspot. This section aims to systematically elaborate on the basic principles of vision-based navigation technology, sensor types, and their characteristics.

2.1 Fundamental Principles of Vision-Based Navigation

Vision-based navigation technology relies on computer vision and image processing theories to achieve UAV autonomous positioning, navigation, and obstacle avoidance by analyzing image information captured by visual sensors. Its basic principles can be subdivided into two levels: visual perception & processing, and positioning & navigation mechanisms.

In the visual perception and processing stage, visual sensors equipped on the UAV sample the surrounding environment and capture image data rich in scene details. Building on this, a series of image preprocessing techniques—such as filtering, image enhancement, and feature extraction—are applied to denoise raw images, enhance contrast, and extract key features, ensuring subsequent analyses are based on highly reliable datasets.

After establishing environmental perception and processing capabilities, the UAV utilizes algorithms such as VO and SLAM, combined with its motion model, to estimate its position and attitude. Through path planning algorithms, the UAV generates optimal flight paths to achieve autonomous navigation. Additionally, as a critical component of the vision-based navigation system, the obstacle avoidance mechanism ensures safe flight in complex environments by detecting and responding to obstacles in real-time.

2.2 Types of Visual Sensors and Their Characteristics

The selection of visual sensors critically determines the performance of vision-based navigation systems. Currently, common visual sensors include monocular cameras, stereo cameras, RGB-D cameras, and fisheye cameras. In practical applications, sensor types must be selected based on specific requirements and environmental characteristics (Table 1).

Table 1. summarizes the different sensor types used in UAV vision-based navigation systems.

Sensor Type	Advantages	Disadvantages
Monocular Camera [4, 5]	Simple structure; low cost; easy installation and calibration	Lacks depth information; requires motion for depth estimation; limited positioning accuracy
Stereo Camera [6]	Provides 3D information; high positioning accuracy	Complex system; computationally intensive; high real-time demands; calibration challenges
RGB-D Camera [7]	Directly acquires color and depth data; high precision; real-time performance.	Limited effective range; sensitive to ambient lighting; higher cost

3 Vision-Based Navigation Algorithms

In the current field of UAV technology, research on vision-based navigation systems is critical, with algorithmic design and implementation forming the core of such systems. This survey aims to comprehensively explore several key vision-based navigation algorithms, detailing their working principles, unique characteristics, technical implementation specifics, applicable scenarios, and recent research advances. This review focuses on the following traditional vision-based navigation methods: VO, SLAM, VINS, and scene-matching navigation technology. Additionally, due to their significant advantages in navigation accuracy and innovation potential, this paper also discusses multi-sensor data fusion and deep learning-based methods, which are emerging as research hotspots in vision-based navigation.

3.1 VO

VO estimates camera motion by analyzing consecutive image frames to enable UAV autonomous navigation [1]. The feature-based method is suitable for texture-rich environments, offering high accuracy but requiring moderate computational resources with limited real-time performance, primarily applied in complex UAV navigation scenarios. The direct method is better suited for texture-deficient environments, providing good real-time performance when computational resources are sufficient, though with lower accuracy requirements, making it applicable to simpler UAV navigation scenarios.

In feature-based methods, Scale-Invariant Feature Transform (SIFT), Speeded Robust Features(SURF), and Oriented FAST and Rotated BRIEF(ORB)are three representative algorithms. SIFT exhibits scale invariance and adapts well to complex environments but suffers from high computational complexity [8]. SURF achieves higher computational efficiency while maintaining scale invariance, making it suitable for real-time applications [9]. ORB offers real-time performance, rotation invariance, and low computational complexity, ideal for resource-constrained UAV platforms [10].

Direct Sparse Odometry (DSO), a typical implementation of the direct method, optimizes camera motion through sparse direct techniques, demonstrating robust real-time performance and accuracy. However, its performance may degrade under dramatic illumination changes or in texture-rich environments, making it suitable for UAV navigation scenarios requiring real-time performance and moderate accuracy.

3.2 SLAM

SLAM algorithms address UAV autonomous navigation in unknown environments by simultaneously constructing maps and estimating localization in real time.

Filter-based SLAM employs filters to estimate camera poses and map point locations, effectively handling noise and uncertainties in dynamic environments. For instance, Extended Kalman Filter SLAM (EKF-SLAM) uses an extended Kalman filter for estimation, offering high computational efficiency and real-time performance [11]. However, it struggles in large-scale environments and exhibits reduced performance in highly nonlinear scenarios.

Optimization-based SLAM leverages nonlinear optimization techniques to enhance pose and map accuracy. This method fully utilizes all observational data, significantly improving navigation precision [12].

3.3 VINS

VINS integrates visual data with Inertial Measurement Unit (IMU) measurements to achieve more precise and stable navigation. Its operational frameworks are categorized into tightly-coupled and loosely-coupled methods.

Tightly coupled VINS enhances navigation accuracy and stability by jointly optimizing visual and inertial data, effectively mitigating sensor noise and drift. While computationally intensive, it is ideal for applications demanding high-precision navigation, such as UAV flights and autonomous vehicles [2, 13].

Loosely-coupled VINS processes visual and inertial data independently, reducing computational complexity for real-time applications at the cost of lower accuracy. It is primarily used in scenarios requiring rapid response, such as UAV path planning with relaxed precision requirements.

3.4 Deep Learning-Based Methods

Deep learning has been widely applied in vision-based navigation, primarily through Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Generative Adversarial Networks (GANs). These architectures enhance navigation accuracy, robustness, and environmental perception by addressing tasks such as image feature extraction, object recognition, trajectory prediction, and environmental model generation.

CNNs are predominantly used for image feature extraction and object recognition in vision-based navigation [14]. By training deep CNNs, hierarchical image features with stronger representational power and robustness can be learned. CNNs apply to tasks like visual odometry, SLAM, and target tracking.

RNNs excel at processing sequential data, such as UAV trajectory prediction [15]. Training RNN models enables the learning of temporal patterns from historical data, improving trajectory forecasting and navigation control. RNNs are suitable for trajectory smoothing, prediction, and motion planning.

GANs, composed of a generator and discriminator, generate realistic environmental models through adversarial training [16]. In vision-based navigation, GANs synthesize photorealistic environmental images to assist in environmental perception and localization. They are particularly useful for simulation training, loop closure detection, and dynamic map updating.

3.5 Scene-Matching Navigation

Scene-matching navigation [3] achieves localization through image region matching, categorized into region-based matching and deep learning-based matching.

Region-based matching methods employ techniques like template matching for localization. While template matching algorithms are computationally simple and real-time capable, they are sensitive to illumination changes, rotation, and scale variations. To enhance robustness, improved algorithms such as Normalized Cross-Correlation (NCC) or SURF can be adopted.

Deep learning-based methods have gained prominence in scene-matching navigation. Training deep CNNs enables the extraction of more discriminative and robust feature representations for image matching, significantly improving accuracy. These methods also handle complex scenarios, including dynamic environments, varying illumination, and partial occlusions.

3.6 Multi-Sensor Data Fusion

Multi-sensor data fusion integrates information from heterogeneous sensors (e.g., visual cameras, Inertial Measurement Units (IMUs)) to achieve more accurate and reliable navigation outcomes. This fusion enhances both system precision and adaptability to complex environmental changes.

The Kalman filter is an efficient data fusion algorithm for linear systems [17]. In vision-based navigation, fuses visual and inertial sensor data to estimate high-precision poses and velocities. By recursively updating state estimates and covariance matrices, it effectively mitigates system and measurement noise.

Particle filters are suitable for nonlinear, non-Gaussian systems [18]. In vision-based navigation, they handle complex nonlinear relationships and uncertainties. Through sequential sampling and weight updating, particle filters approximate posterior probability distributions for precise state estimation.

Graph optimization leverages factor graphs to model sensor constraints [19]. In vision-based navigation, it constructs factor graphs incorporating visual and inertial constraints, optimizing nodes and edges to estimate poses and maps with high accuracy. This method offers flexibility and scalability for large-scale, complex environments.

4 Applications of UAV Vision-Based Navigation Technology

This section discusses applications of UAV vision-based navigation systems in areas such as drone delivery, security surveillance, indoor/outdoor navigation, and search-and-rescue operations, emphasizing how they enhance navigation precision and environmental adaptability through stereo vision, deep learning, multi-sensor data fusion, and advanced filtering techniques.

Vision-based navigation systems in delivery drones achieve high-precision environmental perception and obstacle detection by integrating stereo vision with deep learning. For example, deep learning-enabled visual processing algorithms allow drones to detect and locate river-crossing pipelines without site constraints [20]. To address dynamic obstacle detection challenges, enhanced versions of Kalman filters and particle filters are applied for multi-sensor data fusion and real-time processing, significantly improving detection accuracy and responsiveness.

UAV vision-based navigation systems combined with computer vision and machine learning algorithms enable continuous tracking and precise identification of suspicious targets. A security-focused UAV system developed by a surveillance company utilizes improved Kalman filtering and Support Vector Machine (SVM) classifiers to handle target occlusion and deformation, providing real-time, reliable intelligence to security personnel. The integration of reinforcement learning and adaptive filtering further enhances tracking stability and recognition accuracy.

Diverse environmental demands for UAV navigation drive versatile applications of vision-based technologies. In indoor warehouse settings, drones achieve autonomous navigation and inventory management via visual SLAM. For outdoor agricultural applications, UAVs leverage the Robot Operating System (ROS) for visual data processing to execute precision agricultural spraying [21].

UAV vision-based navigation systems demonstrate exceptional environmental adaptability and robustness in dynamic rescue scenarios. For instance, deep learning models and reinforcement learning-based controls extract positional data of unmanned vessels and floating objects from aerial imagery to facilitate maritime search-and-rescue missions [22]. Advanced techniques like deep reinforcement learning-based path planning algorithms effectively improve system reliability and performance in rapidly changing environments.

5 Challenges and Solutions

Vision-based navigation faces numerous challenges in complex environments, including visual feature changes caused by dynamic objects, image errors induced by illumination variations, and occlusion issues, all of which may degrade navigation accuracy and stability.

In complex environments, vision-based navigation encounters challenges such as visual feature variations caused by moving objects (e.g., pedestrians, vehicles) in dynamic settings, image characteristic changes due to illumination fluctuations, and critical information loss from occlusions, collectively affecting navigation precision, feature extraction/matching accuracy, and system stability. To address these challenges, multiple solutions can be implemented: integrating multimodal sensors (e.g., LiDAR, infrared sensors) to obtain comprehensive environmental information and enhance robustness; developing adaptive feature extraction algorithms such as deep learning-based methods to adapt to varying illumination and dynamic conditions; and employing robust matching algorithms like RANSAC (Random Sample Consensus) and graph optimization-based methods to handle mismatches and occlusions. These approaches collectively improve the performance and reliability of vision-based navigation in complex environments.

Vision-based navigation also faces challenges in computational efficiency and real-time performance. Since algorithms typically involve extensive image processing and optimization computations, they exhibit high computational complexity. Simultaneously, UAV navigation imposes strict real-time requirements, necessitating rapid processing of images and data. Solutions include optimizing algorithm structures to reduce computational complexity (e.g., adopting fast feature extraction and matching algorithms), leveraging parallel computing resources such as Graphics Processing Units (GPUs) to accelerate image processing and optimization, and offloading partial computational tasks to edge devices to minimize data transmission latency. These methods synergistically enhance computational efficiency and real-time performance.

In multi-sensor data fusion, vision-based navigation systems confront challenges of data inconsistency and fusion algorithm complexity. Data from different sensors may suffer from timestamp desynchronization or scale mismatches, while fusion algorithms themselves are computationally intensive, struggling to meet real-time demands. Solutions involve adopting synchronization and calibration techniques to ensure data consistency, researching efficient fusion algorithms such as improved Kalman filters, particle filters, or graph optimization algorithms to reduce computational complexity, and utilizing distributed computing architectures for decentralized fusion to improve computational efficiency.

6 Conclusions

This paper systematically investigates vision-based UAV navigation technologies, analyzing the fundamental principles of vision-based navigation, sensor types, and mainstream algorithms, while exploring their application effectiveness in complex environments. The study concludes that vision-based navigation significantly enhances UAV autonomous navigation capabilities in complex environments through multimodal sensor fusion and adaptive feature extraction. However, challenges persist, including dynamic environmental interference, illumination variations, occlusions, and computational inefficiencies. To address these issues, the paper proposes solutions such as multimodal sensor fusion, algorithm structure optimization, and parallel computing resource utilization. Future research on vision-based navigation should focus on algorithm optimization and innovation to reduce computational complexity and improve system adaptability. Concurrently, integrating artificial intelligence and machine learning technologies will drive the development of more intelligent navigation systems capable of addressing complex and dynamic environmental challenges.

References

1. Pintér, M., Janoušek, J., Klouda, J., et al.: 'Using Visual Odometry to Determine the Position of a UAV', IFAC-PapersOnLine, 2024, 58, (9), pp 281-286
2. Wang, L., Yang, G., Li, B., et al.: 'Indoor positioning and autonomous navigation method of UAV based on improved VINS', Control Engineering, 2024, 1, (9), pp 1-9. DOI:10.14107/j.cnki.kzgc.20231064
3. Zhao, C., Zhou, H., Lin, Z., et al.: 'A review of UAV scene matching visual navigation technology', Chinese Science: Information Science, 2019, 49, (5), pp 507-519
4. Nabavi-Chashmi, S. Y., Asadi, D., Ahmadi, K.: 'Image-based UAV position and velocity estimation using a monocular camera', Control Engineering Practice, 2023, 134, pp 105460
5. Lee, B., Sangkyung, S.: 'A study on dual quaternion-based cooperative relative navigation of multiple UAVs with monocular vision-inertial integration', Chinese Journal of Aeronautics, 2024, 37, (11), pp 335-354
6. Zhang, R., Lian, S., Li, L., et al.: 'Design and experiment of a binocular vision-based canopy volume extraction system for precision pesticide application by UAVs', Computers and Electronics in Agriculture, 2023, 213, pp 108197
7. Cai, W., Du, S., Yang, W.: 'UAV image stitching by estimating orthograph with RGB cameras', Journal of Visual Communication and Image Representation, 2023, 94, pp 103835
8. Wei, X., Gan, S., Yuan, X., et al.: 'Research on edge response optimization based SIFT algorithm in UAV image matching', Surveying and Mapping Engineering, 2024, 33, (6), pp 1-10. DOI:10.19349/j.cnki.issn1006-7949.2024.06.001
9. Zhang, S., Nie, W., Rao, W., et al.: 'UAV image fast matching algorithm based on improved SURF based on overlapping regions', Radio Engineering, 2024, 54, (8), pp 1978-1985
10. Zhu, F., Li, H., Li, J., et al.: 'Unmanned aerial vehicle remote sensing image registration based on an improved oriented FAST and rotated BRIEF-random sample consensus algorithm', Engineering Applications of Artificial Intelligence, 2023, 126, pp 106944

11. Hammia, S., Hatim, A., Haijoub, A.: 'Enhancing Real-time Simultaneous Localization and Mapping with FPGA-based EKF-SLAM's Hardware Architecture', Procedia Computer Science, 2024, 236, pp 517-526
12. Wang, L.: 'Indoor mapping and positioning of quadrotor drone based on multi-information fusion under ORB-SLAM2', North China Electric Power University (Beijing), 2024
13. Wang, F., Zou, Y., Zhang, C., et al.: 'UAV navigation in large-scale GPS-denied bridge environments using fiducial marker-corrected stereo visual-inertial localisation', Automation in Construction, 2023, 156, pp 105139
14. Cui, J., Zhang, X., Zhang, J., et al.: 'Weed identification in soybean seedling stage based on UAV images and Faster R-CNN', Computers and Electronics in Agriculture, 2024, 227, pp 109533
15. Wu, J., Luo, Z., Li, W., et al.: 'Research on UAV audio signal recognition algorithm based on recurrent neural network model', Electroacoustic Technology, 2024, 48, (5), pp 58-63. DOI:10.16311/j.audioe.2024.05.017
16. Li, Y., Feng, L., Yang, Y., et al.: 'GAN-powered heterogeneous multi-agent reinforcement learning for UAV-assisted task offloading', Ad Hoc Networks, 2024, 153, pp 103341
17. Tang, X., Yang, L., Wang, D., et al.: 'A collaborative navigation algorithm for UAV Ad Hoc network based on improved sequence quadratic programming and unscented Kalman filtering in GNSS denied area', Measurement, 2025, 242, pp 115977
18. Guo, X., Zeng, D., Li, W., et al.: 'Marginalized particle filtering for online parameter estimation of PEMFC applied to hydrogen UAVs', Sustainable Energy Technologies and Assessments, 2023, 57, pp 103265
19. Song, C., Lu, W., Song, H., et al.: 'Onboard cooperative relative positioning system for Micro-UAV swarm based on UWB/Vision/INS fusion through distributed graph optimization', Measurement, 2024, 234, pp 114897
20. Wang, Q., Wang, Z., Gao, Y.: 'Application of river pipeline inspection technology based on UAV vision', Water Supply Technology, 2024, 18, (4), pp 43-46
21. Tang, F.: 'Research on visual navigation technology of agricultural UAV based on ROS', Qilu University of Technology, 2023. DOI:10.27278/d.cnki.gsdqc.2023.000676
22. Wang, Y., Liu, W., Liu, J., et al.: 'Cooperative USV–UAV marine search and rescue with visual navigation and reinforcement learning-based control', ISA Transactions, 2023, 137, pp 222-235