Exploration on the Application of Intelligent Perception Technology in UAV Autonomous Flight

Yu Fengyou

Chongqing Bashu Secondary School, Chongqing, 400010, China

Abstract: As the core support for UAV autonomous flight, intelligent perception technology is driving the evolution of aircraft from remote control operation to fully autonomous decision-making. Through multi-source sensor fusion and deep learning algorithms, this technical system endows aircraft with environmental cognition and situation understanding capabilities, enabling them to independently complete navigation, obstacle avoidance and task execution in complex scenarios. Current research focuses on the collaborative optimization of key modalities such as visual perception, LiDAR, millimeter-wave radar and infrared detection, so as to improve the robustness of the system in low-visibility and high-dynamic environments. Technological breakthroughs have significantly expanded the application boundaries of UAVs in fields such as logistics distribution, agricultural plant protection, emergency rescue and infrastructure inspection, giving birth to new industrial forms and business models. However, the real-time bottleneck of perception systems, communication delay in multi-UAV coordination and reliability issues under complex meteorological conditions are still the main obstacles restricting large-scale deployment. The future development trend will point to the lightweight design of edge intelligent computing architecture, semantic-level fusion of multi-modal perception information, and the construction of distributed cooperative perception mechanisms for swarm intelligence, so as to achieve truly fully autonomous flight capabilities.

Keywords: Intelligent Perception; UAV Autonomous Flight; Multi-source Sensor Fusion; Environmental Cognition; Edge Computing; Swarm Intelligence

0. Introduction

The rapid development of UAV technology is reshaping the pattern of the modern aviation industry, and its core driving force has shifted from traditional platform manufacturing to the deep integration of intelligent systems. In this transformation process, intelligent perception technology, as a key bridge connecting the physical world and digital decision-making, undertakes multiple missions of environmental information acquisition, real-time situation analysis and flight safety control. Different from the early operation mode relying on manual remote control or preset waypoints, contemporary UAV systems urgently need to possess human-like environmental understanding capabilities to cope with the challenges of complex scenarios such as urban canyons, dense forest shelter and electromagnetic interference. The introduction of intelligent perception technology enables aircraft to independently complete obstacle identification, dynamic path planning and abnormal state disposal through on-board computing units, marking a qualitative leap of UAV technology from automation to autonomy. At present, the research hotspots in this field focus on the diversified expansion of perception modalities, real-time optimization of fusion algorithms and the design of coordination mechanisms for heterogeneous systems. These explorations not only provide a technical path for the intelligent upgrading of a single aircraft, but also lay a theoretical foundation for the construction of cluster operations and air-ground integrated networks. An in-depth analysis of the internal mechanism and application paradigm of intelligent perception technology is of important theoretical value and practical significance for promoting the high-quality development of the UAV industry.

1. Technical Architecture and Optimization Path of Visual Perception System

1.1 Algorithm Evolution of Monocular Visual Depth Estimation

Monocular visual depth estimation technology infers the three–dimensional structure of a scene from a single two–dimensional image, providing a lightweight spatial perception solution for UAVs. Early methods relied on geometric constraints and manual feature extraction, which were prone to estimation failure in texture–deficient areas. In recent years, end–to–end deep learning models based on encoder–decoder architecture have significantly improved prediction accuracy. By introducing attention mechanism and multi–scale feature fusion strategy, the network can adaptively focus on key areas and integrate context information. The rise of self–supervised learning framework further reduces the dependence on labeled data. The loss function is constructed by using the temporal consistency of image sequences, enabling the model to maintain stable depth inference ability in dynamic scenes. Aiming at the embedded deployment requirements of UAV platforms, knowledge distillation and network pruning techniques are widely used in model compression, controlling inference delay to millisecond level while maintaining accuracy.

1.2 Real–time Matching Strategy of Binocular Stereo Vision

Binocular stereo vision simulates the principle of human binocular parallax and obtains dense depth maps through triangulation, whose accuracy is closely related to baseline length and camera calibration quality. Traditional local matching algorithms have matching ambiguity in weak texture and repetitive pattern areas, and semi–global matching algorithm alleviates this problem to a certain extent through energy function optimization. Stereo matching networks in the deep learning era adopt 3D convolutional neural networks to construct cost volumes, and realize sub–pixel level disparity estimation using softmax regression. To meet the real–time requirements of UAVs during high–speed flight, researchers have proposed sparse cost volume construction and hierarchical refinement strategies, reducing the computational complexity from cubic to linear level. In terms of hardware acceleration, the collaborative design of field programmable gate arrays and application–specific integrated circuits enables binocular systems to output high–frame–rate depth streams under low–power conditions, meeting the timeliness requirements of obstacle avoidance decision–making.

2. Environmental Modeling Mechanism of LiDAR Perception

2.1 Point Cloud Processing Flow of Mechanical Rotating Radar

Mechanical rotating LiDAR obtains 3D point cloud data of the surrounding environment through 360° rotating scanning, with ranging accuracy up to centimeter level, making it the preferred sensor for high–precision environmental perception of UAVs. Raw point cloud data contains a large number of ground points and noise interference. Ground segmentation algorithms extract passable areas based on point cloud height distribution features or plane fitting methods. Cluster analysis uses Euclidean distance or density peak algorithm to group point clouds into independent targets, providing a basis for subsequent classification and tracking. The introduction of deep learning methods has realized end–to–end semantic analysis of point clouds. PointNet series networks directly process unordered point sets through permutation–invariant operations, while voxelization or projection–based representation methods improve computational efficiency by leveraging mature 2D convolution architectures. Aiming at the computing power constraints of airborne platforms, the balance between lightweight network design and point cloud downsampling strategy becomes a key consideration in engineering practice.

2.2 Integrated Design Trend of Solid–state LiDAR

Solid–state LiDAR abandons the mechanical rotating structure and adopts optical phased array or micro–electro–mechanical system to realize beam scanning, with the advantages of small size, high reliability and low cost, which fits the integration needs of consumer–grade UAVs. Optical phased array technology realizes rapid deflection of beam pointing by regulating optical wave phase, and its inertialess scanning characteristic makes it suitable for high–frequency scene perception. The micro–electro–mechanical system scheme uses micro–mirror arrays to reflect laser beams, realizing chip–level packaging while maintaining high angular resolution. The field of view of solid–state radar is relatively limited, and multi–radar array layout has become a solution to expand the perception range, realizing spatial alignment of multi–source data through external parameter calibration and point cloud registration algorithms. With the maturity of silicon photonics technology, solid–state LiDAR is evolving towards higher line count, longer detection distance and lower power consumption, providing a hardware foundation for all–weather autonomous flight of UAVs.

3. Construction of All–weather Perception Capability of Millimeter–wave Radar

3.1 Signal Processing Principle of Frequency Modulated Continuous Wave Radar

Frequency modulated continuous wave millimeter–wave radar realizes target ranging and velocity measurement by transmitting continuous waves with linear frequency modulation and analyzing echo difference frequency signals. Its working band is usually 24 GHz or 77 GHz, with physical characteristics of penetrating rain, fog, smoke and dust. The signal processing link includes mixing demodulation, fast Fourier transform, constant false alarm rate detection and other links, converting raw intermediate frequency signals into range–Doppler spectrograms. The application of multiple–input multiple–output technology significantly improves angular resolution through virtual array expansion, enabling the radar to distinguish multiple targets in the same range unit. Micro–Doppler feature analysis uses spectral modulation caused by vibration or rotation of target components to realize fine classification of targets such as UAVs, birds and ground pedestrians, providing a unique information dimension for situation understanding in complex airspace.

3.2 High–resolution Reconstruction Technology of 4D Imaging Radar

4D imaging radar adds the capability of pitch angle measurement on the basis of traditional three–dimensional perception, outputting point cloud data including range, azimuth, height and velocity information, with angular resolution up to degree level or even sub–degree level. High–resolution imaging relies on large–scale antenna arrays and advanced signal processing algorithms. Compressed sensing and super–resolution reconstruction technologies are used to break through physical aperture limitations. The application of deep learning in radar signal processing is increasingly in–depth. Networks based on convolutional neural networks or Transformer architecture can directly estimate target parameters from raw echoes, avoiding information loss and error accumulation in the traditional processing chain. For UAV applications, 4D imaging radar plays an irreplaceable role in visual failure scenarios such as night flight, smoke environment and strong light interference, and its fusion with optical sensors constructs an all–weather perception system.

4. Thermal Imaging Application of Infrared and Multispectral Perception

4.1 Performance Improvement of Uncooled Infrared Focal Plane Detectors

Uncooled infrared focal plane detectors are based on the microbolometer principle and can work in the 8–14 μm long–wave infrared band without low–temperature cooling devices, greatly reducing the volume and power consumption of thermal imaging systems, which is suitable for UAV platform integration. Vanadium oxide and amorphous silicon are mainstream sensitive materials, and their temperature coefficient of resistance determines the sensitivity index of the detector. In recent years, the pixel size has been reduced from 17 μm to 12 μm or even smaller, realizing chip area reduction and cost reduction while maintaining resolution. The optimization of noise suppression technology and digital integration algorithm of readout integrated circuits further improves thermal sensitivity, enabling the system to distinguish smaller temperature differences. The advancement of packaging technology has promoted the integrated integration of wafer–level optics and detectors, providing technical support for the lightweight design of UAV payloads.

4.2 Cross–spectrum Fusion Method of Thermal Imaging and Visible Light

The fusion of thermal imaging and visible light images aims to integrate the thermal radiation information of the former and the texture details of the latter to generate synthetic images with more interpretability for human eyes or machines. Pixel–level fusion methods include weighted average, Laplacian pyramid and multi–scale transformation, but are prone to problems such as reduced contrast or artifacts. Feature–level fusion uses deep learning networks to extract hierarchical features of the two modalities respectively, and realizes information integration by designing fusion rules or training fusion networks. In UAV applications, cross–spectrum fusion significantly improves the efficiency of tasks such as night search and rescue, fire monitoring and concealed target detection. Registration algorithm is the key to fusion quality. Registration methods based on feature points or mutual information need to cope with the challenges brought by nonlinear gray–scale differences and perspective changes between modalities. The development of unsupervised and self–supervised learning frameworks reduces the dependence on paired training data and enhances the practicability and deployment flexibility of the system.

4.3 Practice of Multispectral Imaging in Agriculture and Environmental Monitoring

Multispectral imaging obtains the reflection information of targets in specific bands through narrow–band filters or spectroscopic elements, and realizes crop growth assessment, pest identification and water stress monitoring combined with vegetation index calculation. Multispectral cameras carried by UAVs usually cover blue, green, red, red–edge and near–infrared bands, with spatial resolution up to centimeter level, meeting the refined needs of precision agriculture. In the field of environmental monitoring, multispectral data is used for water eutrophication assessment, forest cover change detection, geological mineral identification and other applications. Radiometric calibration and geometric correction preprocessing ensure the quantitative reliability of data, and machine learning algorithms mine patterns related to target attributes from massive spectral–spatial features.

5. Multi–source Information Fusion and Intelligent Decision System

5.1 Sensor Spatio–temporal Synchronization and Calibration Technology

The premise of multi–source sensor fusion is precise time synchronization and spatial calibration. Any deviation will lead to data association errors and reduced fusion performance. Time synchronization is usually realized by hardware trigger or network time protocol. Global navigation satellite system timing and pulse per second signals provide a unified time reference for each sensor. Spatial calibration needs to estimate the rotation and translation transformation relationship between sensors. Methods based on calibration boards or natural scenes are suitable for laboratory and field environments respectively. Online calibration algorithms can estimate and compensate parameter drift during system operation, which is particularly important for long–term operating UAV systems.

5.2 End–to–end Perception Network Based on Deep Learning

End–to–end deep learning networks directly map raw sensor data to high–level semantic output, simplifying the cumbersome manual design process in traditional perception systems. Multi–modal fusion network architectures include early fusion, middle fusion and late fusion paradigms, among which the Transformer architecture based on attention mechanism shows powerful cross–modal information integration capabilities. The bird's–eye view representation method converts multi–view images or point clouds into a unified top–down grid map, facilitating downstream prediction heads for target detection and trajectory prediction. Occupancy grid mapping models the occupied state of three-dimensional space in a voxel manner, avoiding the limitations of explicit target detection in occluded or special-shaped object scenarios. Self-supervised and semi-supervised learning technologies use large-scale unlabeled data to improve feature representation ability and reduce the dependence on expensive labeling. Network architecture search and quantization technologies are committed to seeking the optimal balance between accuracy and efficiency to meet the real-time constraints of airborne computing.

5.3 Real-time Optimization Strategy of Edge Computing Architecture

The on-board computing resources of UAVs are strictly limited, and the design of edge intelligent architecture requires a fine trade-off between computing power, power consumption and delay. Heterogeneous computing platforms integrate computing units such as central processing units, graphics processing units and neural network accelerators, and realize load balancing through task division and scheduling. Model optimization technologies include network pruning to remove redundant connections, knowledge distillation to transfer large model knowledge to lightweight networks, and quantitative inference to reduce numerical accuracy requirements. Neural architecture search automatically explores the optimal network topology adapted to hardware constraints. Software-level optimization involves operator fusion, memory layout optimization and compiler automatic tuning. Part of the computing tasks are offloaded to ground stations or the cloud for execution, and the impact of communication bandwidth fluctuations is alleviated through adaptive code rate control and predictive transmission strategies. The federated learning framework supports multi-machine collaborative model training, improving the generalization performance of perception models while protecting data privacy.

6. Conclusion

As the nerve center of UAV autonomous flight, the development context of intelligent perception technology clearly outlines the evolution trajectory from single modality to multi-source fusion, from rule-driven to data-driven, and from single-machine intelligence to swarm collaboration. The deep integration of multi-modal sensors such as vision, LiDAR, millimeter-wave radar and infrared has endowed UAVs with the ability of accurate cognition and autonomous decision-making in complex and dynamic environments, significantly expanding their application boundaries and task efficiency. The current technical system still faces practical challenges in extreme weather adaptability, long-endurance energy management and high-density airspace coordination. The breakthrough of these bottlenecks requires interdisciplinary collaborative innovation of materials science, microelectronics and artificial intelligence algorithms. Looking forward to the future, the maturity of neuromorphic computing chips will promote the energy efficiency ratio of perception systems to achieve an order of magnitude improvement, the virtual test environment constructed by digital twin technology will accelerate algorithm iteration and system verification, and the development of swarm intelligence theory will give birth to a decentralized collaborative perception network. With the continuous expansion of technical boundaries and the increasingly perfect industrial ecology, intelligent UAVs with fully autonomous flight capabilities will surely play an irreplaceable strategic value in the construction of smart cities, the implementation of precision agriculture and the construction of emergency response systems, becoming an important enabling technology to promote the digital transformation of the economy and society.

References:

[1] Zhang Aopeng. Research on Autonomous Flight Control Method of Coaxial Rotor UAV[D]. Shenyang Ligong University, 2025.

[2] Pan Minglong, Liu Bo. Research on Autonomous Flight Control of Aerial Detection UAV[C]// Chongqing Big Data and Artificial Intelligence Industry Association. Proceedings of the Academic

Symposium on Artificial Intelligence and Economic Engineering Development (III). Nanjing University of Aeronautics and Astronautics, 2025: 680−682.

[3] Pan Zhengxiao. Research on UAV Autonomous Flight Control Based on Multi−dimensional Data Fusion[D]. Civil Aviation Flight University of China, 2022.

[4] Song Yuxuan. Design and Implementation of Quadrotor UAV Autonomous Flight System in Unknown Environment[D]. Harbin Institute of Technology, 2022.

[5] Wang Congbao, Zhang Ansi, Yang Lei, et al. A Review of Deep Vision−based Perception and Obstacle Avoidance for Quadrotor UAV Autonomous Flight[J]. Radio Engineering, 2023, 53(10): 2233−2243.