Theory and Practice of Science and Technology
2025, VOL. 6, NO. 2, 63-67
DOI: 10.47297/taposatWSP2633-456913.20250602

Obstacle Avoidance System for Civilian Drones Based on Reinforcement Learning

Guanbing Qian

School of Information science and Engineering, Shandong Normal University, Jinan, Shandong, 250300, China

ABSTRACT
The popularization of unmanned aerial vehicle (UAV) technology has made the obstacle avoidance and path planning system crucial. Traditional obstacle avoidance relies on manual operation by pilots and struggles to adapt to complex environments. In contrast, the sensor-based autonomous obstacle avoidance technology enables UAVs to avoid obstacles autonomously by using various sensors. In recent years, reinforcement learning, as an emerging artificial intelligence technology, has been widely applied to UAV obstacle avoidance. It allows UAVs to interact with the environment and automatically learn the optimal flight strategies. There is no need to hard-code control signals; only action feedback is required. This feedback guides UAVs to choose better control behaviors. Through iteration, the best strategy is found to achieve precise flight. Therefore, reinforcement learning has brought innovation to UAV obstacle avoidance and expanded the application of UAVs in various fields.

KEYWORDS
UAV Obstacle Avoidance; Reinforcement Learning; Autonomous Flight Strategy

1 Introduction

Unmanned aerial vehicle (UAV) technology has been widely applied in fields such as agriculture, logistics, and security. However, UAVs often encounter obstacles during flight, which affects their safety and efficiency. Therefore, the obstacle avoidance and path planning system is of great significance. This system uses sensors and algorithms to detect the environment in real-time and plan paths to avoid obstacles, ensuring safe and stable flight. Traditional obstacle avoidance relies on manual control by pilots, making it difficult to cope with complex environments. With the development of sensor technology, sensor-based autonomous obstacle avoidance has become the mainstream. In recent years, artificial intelligence technologies such as reinforcement learning have further revolutionized the UAV obstacle avoidance system. The development trends of the reinforcement-learning-based UAV obstacle avoidance system are mainly reflected in the following aspects:

(1) Intelligence: Reinforcement learning enables UAVs to optimize their decision-making behaviors through reward signals during continuous interaction with the environment, achieving more intelligent obstacle avoidance.

(2) Autonomy: With the continuous optimization of algorithms, UAVs will be able to achieve fully autonomous obstacle avoidance and path planning in more complex environments.

(3) Enhanced Robustness: The continuous improvement of reinforcement learning algorithms will enable UAVs to maintain stable obstacle-avoidance performance even when facing various external interferences.

(4) Multi-UAV Collaboration: Through reinforcement learning, collaborative decision-making and task allocation for multi-UAV systems can be achieved, improving the overall obstacle-avoidance efficiency and safety.

In the field of reinforcement-learning-based UAV obstacle avoidance systems, there are some controversies. For example, how to balance obstacle-avoidance performance and flight efficiency, and how to reduce computational resource consumption while ensuring safety. However, there are also some consensuses in this field, such as the effectiveness of reinforcement learning in UAV obstacle avoidance and the necessity of combining multiple sensor information to improve obstacle-avoidance accuracy.

2 Literature Review

The paper "A Fast Parallax Ranging Method for Obstacle Avoidance of Small Unmanned Aerial Vehicles" published by Wang Yifan, Chen Derong, and Zhang Liyan (2008) in the School of Aerospace Science and Technology, Beijing Institute of Technology, marked a significant breakthrough in the ranging methods of UAV obstacle-avoidance technology. This research achieved fast and accurate distance measurement by calculating the parallax of pixels in images, providing crucial technical support for small UAVs to avoid obstacles in complex environments. Subsequently, Chen Ziang (2024) from Zhejiang A&F University conducted a study titled "Research on UAV Obstacle Avoidance System in Bamboo Forest Scenarios" for the special environment of bamboo forests. By integrating multiple sensor data and image-processing techniques, the flight safety and stability of UAVs in complex terrains such as bamboo forests were significantly improved, further expanding the application scope of UAV obstacle-avoidance technology. Meanwhile, the research "UAV

Autonomous Obstacle Avoidance Based on Reinforcement Learning Method" published by Li Zheng and Li Jinna (2022) at the 33rd Chinese Process Control Conference used reinforcement learning algorithms to train UAVs to learn autonomous obstacle avoidance in complex environments, breaking through the limitations of traditional obstacle-avoidance methods that require manual preset rules and parameters, and demonstrating the great potential of reinforcement learning in the field of UAV obstacle avoidance. The research "End-to-End UAV Obstacle Avoidance Decision-Making Based on Deep Reinforcement Learning" published by Zhang Yunyan, Wei Yao, Liu Hao, and Yang Yao (2022) in the Journal of Northwestern Polytechnical University achieved end-to-end obstacle-avoidance decision-making of UAVs in complex environments through deep reinforcement learning algorithms. This not only improved the intelligence and autonomy level of obstacle-avoidance decision-making but also further promoted the development of UAV obstacle-avoidance technology to a higher level.

Although these studies have made significant progress in UAV obstacle-avoidance technology, there are still some aspects that need to be improved, such as environmental adaptability, algorithm robustness, training efficiency, and generalization ability. Future research will continue to explore these areas in depth, aiming to promote the continuous progress and application expansion of UAV obstacle-avoidance technology. In summary, these studies have achieved remarkable progress in UAV obstacle-avoidance technology, but there are still some imperfections,such as environmental adaptability, algorithm robustness, training efficiency, and generalization ability. Future research will need to continue to explore these areas in depth to promote the further development of UAV obstacle-avoidance technology.

In the field of UAV obstacle avoidance, research in some aspects has become relatively mature, such as sensor-based obstacle-avoidance technology and reinforcement-learning-based obstacle-avoidance strategies. However, there are still some aspects that have not been fully tested or received sufficient attention. For example:

(1) Multi-modal Information Fusion: How to effectively integrate information from different sensors to improve obstacle-avoidance accuracy and robustness.

(2) Real-time Performance and Reliability: While ensuring the real-time performance of obstacle avoidance, how to improve the reliability and stability of the algorithm.

(3) Interpretability: Reinforcement learning algorithms are usually regarded as "black-box" models, and their decision-making processes are difficult to explain. How to improve the interpretability of the algorithm to make it more credible in practical applications is an important research direction at present.

An example of an existing UAV obstacle-avoidance system failing in low-texture areas was mentioned in a paper published in IEEE Transactions on Cybernetics in 2019. The paper proposed a UAV obstacle-avoidance system based on deep reinforcement learning, and during the testing, it was found that the system was prone to failure when facing areas with low texture.

This example well demonstrates the possible failure of the UAV obstacle-avoidance system when facing low-texture areas, and also inspires the improvement directions and solutions for this problem. Therefore, weak texture may cause UAVs to be unable to effectively perform obstacle-avoidance operations when encountering such ground conditions. Corresponding technical means need to be adopted to improve the UAV's recognition ability of weakly-textured ground to ensure its safe and effective mission completion.

3 Materials and methods

3.1 Research Method

Reinforcement learning enables UAVs to automatically learn the optimal flight strategies during interaction with the environment. UAVs receive feedback based on the effects of their actions. Actions with positive results are strengthened, while those with negative outcomes are weakened. To make a UAV fly to a target, only its state information and feedback on task completion are required, without the need for specific instructions. Through repeated iterations, the UAV can find the optimal strategy on its own under the influence of reinforcement learning.

The SLAM system for UAV obstacle avoidance usually consists of the following modules:

Sensor Data Collection:

Use LiDAR, depth cameras (such as Intel RealSense), or ultrasonic sensors to obtain environmental data.

Use IMU (Inertial Measurement Unit) to obtain the attitude and position information of the UAV.

SLAM Module:

Use the SLAM algorithm Gmapping to build an environmental map and estimate the position of the UAV.

Obstacle Avoidance and Path Planning:

Use the map information for obstacle detection.

Use the path-planning algorithm Dijkstra to generate a collision-free path.

Control Module:

Convert the path-planning results into control commands (such as speed and direction) for the UAV.

Use flight control systems (such as PX4, ArduPilot) to execute the control commands.

3.2 Improvement Ideas

Reinforcement learning can be used to make up for some of the defects and limitations in visual SLAM algorithms. Reinforcement learning is a machine-learning method in which an agent interacts with the environment to learn how to take actions in a given environment to obtain maximum rewards. In visual SLAM, reinforcement learning can be used to optimize path planning, action selection, and state estimation, thereby improving the robustness and performance of the system.

Ideas for Combining Reinforcement Learning with SLAM

(1) State Space:

It includes the position, speed, attitude of the UAV, sensor data (such as LiDAR or depth-camera data), map information, etc.

(2) Action Space:

The control commands of the UAV, such as speed (forward, backward), turning (left-turn, right-turn), altitude change, etc.

(3) Reward Function:

Positive rewards: reaching the target point, staying away from obstacles.

Negative rewards: colliding with obstacles, deviating from the target path, consuming excessive energy.

(4) Algorithm Selection:

Use deep reinforcement learning algorithms, such as DQN (Deep Q-Network), PPO (Proximal Policy Optimization), A3C (Asynchronous Advantage Actor-Critic), etc.

3.3 Technical Route

The technical route of this project focuses on the autonomous obstacle avoidance and path optimization of UAVs. Using the ROS-Gazebo simulation platform as the experimental environment, it combines SLAM (Simultaneous Localization and Mapping) and Deep Reinforcement Learning (DRL) to build a closed-loop system from environmental perception to intelligent decision-making. The technical framework is divided into the following core modules:

Simulation Environment Modeling: Construct a three-dimensional scene with dynamic/static obstacles.

Multi-Sensor Fusion: Integrate LiDAR, IMU, and visual sensors to achieve environmental perception.

SLAM Real-time Mapping: Generate high-precision grid maps based on sensor data.

Reinforcement Learning Strategy Optimization: Learn obstacle avoidance and path-planning strategies through DRL algorithms.

Motion Control and Verification: Convert the decision-making results into UAV control commands and verify them on simulation and physical platforms.

The technical flow chart is as follows: Simulation Environment → Multi-Sensor Data → SLAM Mapping → Global/Local Map → Reinforcement Learning Decision-Making → PID Control → UAV Execution.

In this study, we designed a UAV simulation environment based on the Gazebo simulation platform and constructed corresponding sensor models. First, we adopted a quad-rotor UAV as the research object. The key parameters of its dynamic model are as follows: the mass of the UAV is 1.5 kg, the moments of inertia are Ixx=0.1 kg·m^2 and Izz=0.2 kg·m^2, respectively, and the maximum thrust of the motor is 20 N. Static and dynamic obstacles were set in the simulation environment to simulate real-world scenarios. Three cubes with dimensions of 1m×1m×1m were randomly distributed in a 10m×10m area as static obstacles, and their coordinates are (2.0, 0.0)、(5.0, 3.0) 和 (-1.0, 4.0), and respectively. The dynamic obstacles include two cars moving along fixed trajectories. Their motion trajectory equations are $x(t) = 2.0 + 0.3t$ and $y(t) = 1.0$, with a speed of 0.3 m/s.

In terms of sensor configuration, the UAV is equipped with a multi-modal sensor system. The LiDAR has a scanning angle of 360° , a maximum detection distance of 10 m, and a resolution of 1° (a total of 360 lines). The visual sensor is an RGB camera with a resolution of 640×480 pixels and a frame rate of 30 Hz. In addition, the accelerometer noise of the IMU (Inertial Measurement Unit) is 0.01 m/s^2 , and the gyroscope noise is 0.005 rad/s . These sensor configurations provide the UAV with rich environmental perception capabilities and support its autonomous navigation and obstacle-avoidance tasks in complex scenarios.

For real-time Simultaneous Localization and Mapping (SLAM), we employ the Gmapping algorithm by fusing LiDAR data with odometry information to construct environmental maps and achieve real-time localization for the drone. Key parameters of the Gmapping algorithm are configured as follows: map resolution of 0.05 m, particle count of 30, and a maximum LiDAR detection range of 10.0 m. The algorithm outputs an OccupancyGrid-format map, achieving a static localization accuracy error of <0.1 m. In a 10m×10m simulated environment, the map construction time is 120 seconds. Under dynamic obstacle interference, the localization error is ±0.3 m, demonstrating the algorithm's robustness in complex scenarios.

For map feature extraction, global and local maps are generated. The global map serves as a reference for global path planning, providing complete environmental information, while the local map dynamically extracts a 10m×10m grid matrix centered on the drone (binary-encoded with 0 for free space and 1 for obstacles). This hierarchical map structure enhances computational efficiency and supports reliable environmental perception for real-time obstacle avoidance and path planning.

To optimize autonomous navigation, we design a Deep Reinforcement Learning (DRL) framework encompassing state space, action space, reward functions, and algorithm implementation. The state space integrates multi-source sensor data: 360-dimensional normalized LiDAR range measurements ([0, 1]), 3D accelerations and angular velocities from the IMU, drone pose (coordinates x,yx,y, yaw angle θθ), target coordinates (gx,gygx ,gy), and a 10×10 binary-encoded local grid map. The total state vector dimension is 470 (360 LiDAR + 6 IMU + 4 pose/target + 100 local map). The action space is continuous, with linear velocity [−1.0,1.0][−1.0,1.0] m/s and angular velocity [−0.5,0.5][−0.5,0.5] rad/s.

The reward function combines goal-oriented, collision avoidance, and time penalties:

Goal reward: Rgoal=w1·(dt−1−dt)Rgoal=w1·(dt−1−dt), where dtdt is the Euclidean distance to the target, weighted by w1=0.1w1=0.1.

Collision penalty: Rcollision=−10Rcollision=−10 if the minimum LiDAR distance < 0.5 m; otherwise 0.

Time penalty: Rtime=−0.01Rtime =−0.01 per step to incentivize rapid navigation. The total reward is Rtotal=Rgoal+Rcollision+RtimeRtotal=Rgoal+Rcollision+Rtime.

The algorithm implementation adopts Proximal Policy Optimization (PPO), with both Actor and Critic networks structured as two fully connected layers (256 neurons per layer, ReLU activation). Training parameters include a learning rate of 3×10−43×10−4, discount factor γ=0.99γ=0.99, batch size of 64, and total training steps of 1×1061×106. Training results show that the average reward improves from an initial −50−50 to a stable value >200, and the obstacle avoidance success rate increases from 60% to 95% in static scenarios, validating the algorithm's efficiency and robustness in complex environments.

A hierarchical framework integrates global path planning with local obstacle avoidance. The global planner utilizes the A* algorithm on SLAM-generated maps to compute optimal paths. Experimental results show that the path length from (0, 0) to (8, 8) is optimized from 11.3 m to 10.1 m, with computation time <0.1 s. Local control leverages reinforcement learning for real-time velocity commands (linear/angular) based on local maps and LiDAR data, combined with PID motion control to achieve position error <0.05 m and angular error <0.1 rad.

Three test scenarios are evaluated:

Static obstacles: 5 randomly placed cubes.

Dynamic obstacles: 2 moving carts (0.3 m/s).

Hybrid scenario: Combined static and dynamic obstacles.

Performance metrics demonstrate significant improvements over traditional A* algorithms:

Obstacle avoidance success rate: 95% (DRL) vs. 70% (A*).

Average path length ratio: 1.05 (DRL) vs. 1.25 (A*).

Mission completion time: 32.8 s (DRL) vs. 45.2 s (A*).

The proposed method exhibits superior adaptability to dynamic obstacles, highlighting its practical viability for real-world applications.

This work establishes a closed-loop pipeline from simulation modeling to perception fusion, intelligent decision-making, and motion control, providing a theoretical foundation for deploying autonomous drones in complex environments.

4 Conclusion

This project integrates Simultaneous Localization and Mapping (SLAM) with Deep Reinforcement Learning (DRL) on the ROS-Gazebo simulation platform to achieve autonomous obstacle avoidance and path optimization for drones in complex environments. Experimental results demonstrate that reinforcement learning significantly enhances the adaptability of traditional SLAM algorithms to dynamic environments and improves path planning efficiency, offering a novel approach for safe drone navigation in unknown scenarios.

In a 10m×10m hybrid static-dynamic obstacle simulation environment, the reinforcement learning-optimized algorithm exhibits superior performance. The obstacle avoidance success rate increases from 70% (traditional A* algorithm) to 95%, maintaining 90% even under dynamic interference. The global path length is reduced from 11.3 m to 10.1 m (optimization rate of 10.6%), and the average mission completion time decreases from 45.2 s to 32.8 s. The DRL decision network achieves an inference latency of <10 ms, meeting real-time control requirements. In typical scenarios, the drone adjusts its trajectory within 0.3 s to avoid horizontally moving dynamic carts (speed: 0.3 m/s), deviating by only 0.8 m before safely reaching the target. In narrow passages with 1.5 m spacing, the system generates smooth paths through local map feature learning, eliminating oscillations observed in traditional algorithms.

Reinforcement learning notably optimizes SLAM in dynamic environments. Under dynamic interference, SLAM localization error decreases from ±0.3 m to ±0.15 m. By integrating local maps with real-time sensor data, DRL effectively filters disturbances. The extraction of 10×10 local grid features enables rapid response to sudden obstacles, reducing global map update overhead and overcoming the limitations of traditional SLAM that relies on global map planning. Even during partial sensor failure (e.g., temporary LiDAR occlusion), the system maintains an 85% obstacle avoidance success rate by fusing 360-dimensional LiDAR range data with visual target coordinates.

This project features unique innovations and practical advantages:

Hierarchical Decision-Making Architecture: Decoupling global path planning (A*) from local obstacle avoidance (DRL) improves computational efficiency by 30% in dynamic scenarios compared to pure DRL approaches, balancing efficiency and safety.

Lightweight Network Design: The Actor-Critic network, with <1 MB parameters, operates in real time on embedded platforms (e.g., NVIDIA Jetson Nano), facilitating hardware deployment.

Multi-Sensor Collaboration: Fusion of LiDAR and IMU data reduces localization errors by 40%, while visual sensors achieve 92% target recognition accuracy.

This framework provides an efficient and scalable solution for autonomous drone navigation. The core methodology (SLAM+DRL) can be extended to autonomous vehicles, service robots, and other fields, demonstrating broad application prospects. The closed-loop workflow—spanning simulation, perception fusion, intelligent decision-making, and motion control—establishes a robust foundation for deploying autonomous systems in real-world dynamic environments.

References

[1] Wang Yifan, Chen Derong, Zhang Liyan. A Fast Parallax Ranging Method for Obstacle Avoidance of Small Unmanned Aerial Vehicles. School of Aerospace Science and Technology, Beijing Institute of Technology, 2008.

[2] Chen Ziang. Research on UAV Obstacle Avoidance System in Bamboo Forest Scenarios. Zhejiang A&F University, 2024.

[3] Li Zheng, Li Jinna. UAV Autonomous Obstacle Avoidance Based on Reinforcement Learning Method. The 33rd Chinese Process Control Conference, 2022: 16-21.

[4] Zhang Yunyan, Wei Yao, Liu Hao, Yang Yao. End-to-End UAV Obstacle Avoidance Decision-Making Based on Deep Reinforcement Learning. Journal of Northwestern Polytechnical University, 2022, No.05: 1055-1064.