Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 16, No. 2, 2026

Design and Implementation of a Search-and-Rescue UAV System Based on Active Disturbance Rejection Control and Vision Technology

Yuan Jiang [1, #], Weicai Wang [2, #, *], Yichen Hua [1]

[1] Army Engineering University, Shijiazhuang, Hebei, China
[2] Army Engineering University, Nanjing, Jiangsu, China

*** Corresponding author:** Weicai Wang

[#]These authors also contributed equally to this work

Abstract: This paper presents a search-and-rescue unmanned aerial vehicle (UAV) system that integrates active disturbance rejection control (ADRC), multimodal visual perception, multi-source sensor fusion, and precision payload delivery. On the basis of the supplied outline and the uploaded simulation material, a complete conference-style manuscript is constructed to describe the system architecture, mission workflow, key algorithms, and simulation-oriented evaluation. The proposed platform uses a ZD550 quadrotor, an AIxBoard onboard computer, Ubuntu, and ROS2 to coordinate sensing, localization, recognition, control, and mission execution. The perception module combines visible-light and infrared sensing with lightweight detection, re-identification, and tracking. The localization module fuses GPS/BeiDou, inertial sensing, visual SLAM, and laser cues to improve robustness in open and weak-GNSS environments. The control module applies ADRC to suppress disturbances caused by wind, maneuvering, and model uncertainty. Ten simulation scenario groups are incorporated into the paper, including lighting variation, complex backgrounds, observation distance, frame rate, overlap tracking, posture variation, carried-item interference, multimodal fusion, UAV trajectory evolution, and positioning error comparison. The results show that multimodal perception and sensor fusion improve recognition and localization consistency under environmental uncertainty, while ADRC enhances platform stability and therefore supports safer delivery execution.

Keywords: Search-and-rescue UAV; Active Disturbance Rejection Control; Multimodal Vision; Sensor Fusion; Visual Navigation; Precision Delivery.

1. Introduction

Search-and-rescue UAVs have become an important component of modern emergency response because they can quickly access hazardous areas, expand situational awareness, and support time-critical decision-making [1]. Their value is particularly significant during the golden rescue period after earthquakes, landslides, fires, and urban accidents, where fast aerial observation can reduce search time and improve the allocation of rescue resources. However, realistic rescue missions are rarely executed under ideal conditions. Instead, they are characterized by GNSS denial or multipath effects, strong wind disturbance, abrupt illumination changes, dynamic obstacles, dense backgrounds, and uncertain victim postures. These factors jointly challenge the stability, perception reliability, and operational accuracy of UAV systems.

Recent studies have addressed these challenges from several perspectives. In the domain of multimodal perception, Lee et al. showed that visible-infrared sensor fusion improves human detection in visually degraded disaster environments [3]. For multi-target perception, Wang et al. combined YOLO-based detection with DeepSORT-style tracking to improve robustness in complex backgrounds [4]. For autonomous localization in GNSS-denied scenarios, visual SLAM has been demonstrated as an effective support for camera-based UAV navigation [2], [5]. On the control side, ADRC provides a practical way to reject unmodeled disturbances in uncertain flight systems [9], [10], while altitude and attitude optimization methods continue to improve rotorcraft stability [6], [8].

To address the remaining gaps, this paper proposes a system-level search-and-rescue UAV architecture that tightly integrates ADRC, multimodal visual recognition, multi-source sensor fusion, and precision delivery. Compared with loosely connected pipeline designs, the proposed framework emphasizes cross-stage coordination among perception, navigation, and execution. The main contributions are fourfold: a practical airborne architecture built around the ZD550 quadrotor, AIxBoard, Ubuntu, and ROS2; a multimodal person-recognition scheme using visible-light and infrared sensing; a localization strategy that combines satellite navigation, inertial sensing, visual SLAM, and range sensing; and a precision delivery strategy with ballistic prediction, visual correction, and safety-trigger mechanisms.

2. Overall System Design

2.1. System Principle and Architecture

The proposed platform is based on a ZD550 quadrotor retrofitted with a dedicated flight-control unit, an AIxBoard onboard computer, and a multimodal sensing suite. The operating system is Ubuntu, and ROS2 organizes sensing, perception, state estimation, task planning, and control as modular nodes. The flight-control loop employs ADRC to estimate the lumped disturbance caused by model uncertainty, payload variation, wind, and transient maneuvering. The estimated disturbance is compensated online so that altitude and attitude stability can be maintained under uncertain rescue conditions [9], [10]. The main system modules and their functions are summarized in Table 1.

For a quadrotor, the translational and rotational dynamics can be written in compact form as $m * x^{\cdot\cdot} = Rf_T + mg + d_tandJ * \Omega^{\cdot} = \tau - \Omega \times J\Omega + d_r$, where m denotes mass, R is the body-to-world rotation matrix, f_T is the total thrust vector, J is the inertia matrix, Ω is the body angular-rate vector, τ is the control torque, and d_t and d_r represent translational and rotational disturbances, respectively. In the proposed design, ADRC treats the unknown terms as generalized disturbances and estimates them with an extended state observer, allowing the controller to remain effective even when the exact plant model is uncertain.

Table 1. Main modules of the proposed search-and-rescue UAV system.

Module	Core components	Function in the mission loop
Flight platform	ZD550 quadrotor, propulsion system, dedicated flight controller	Provides lift, maneuverability, and execution of ADRC-based motion control
Onboard computing	Intel AIxBoard, Ubuntu, ROS2	Runs mission management, perception nodes, state estimation, and communication
Perception subsystem	Visible-light camera, infrared sensor, YOLOv8n/MobileNet-style detector, ReID, SORT/DeepSORT	Detects, verifies, and tracks persons under complex illumination and occlusion
Localization subsystem	GPS/BeiDou, IMU, visual SLAM, laser/LiDAR SLAM	Estimates UAV pose and target-relative position under open and GNSS-limited environments
Delivery subsystem	Release actuator, ballistic model, visual correction, safety trigger	Computes release timing and improves the safety and precision of material delivery

2.2. Operational Workflow

The mission workflow consists of five stages: takeoff and route initialization, autonomous search, visual target confirmation, precise approach and delivery, and return or relay guidance. During the initial stage, the UAV loads a preset search region and establishes its baseline localization state. During the search stage, the platform combines planned waypoints with onboard visual perception to detect possible victims or demand points. Once a candidate is observed, the multimodal recognition module verifies the target and updates its position. The flight controller then performs approach and hover adjustment, after which the precision delivery module computes the release timing and executes a safeguarded drop. Finally, the UAV either returns to base or continues cooperative search according to mission commands.

3. Key Technologies and Implementation

3.1. Multimodal Visual Recognition

Traditional UAV vision systems are highly sensitive to lighting changes, motion blur, target occlusion, and hardware constraints. Backlighting can darken the target region, nighttime scenes can suppress visible textures, and fast-moving targets can cause severe blur. In dense rescue scenes, cluttered backgrounds, overlapping people, and irregular victim poses further increase the difficulty of robust recognition. To cope with these challenges, the proposed framework combines visible-light and infrared sensing so that thermal cues can compensate for the weaknesses of RGB images under low illumination, smoke, fog, and partial shadow [3].

On the algorithmic side, lightweight convolutional detectors such as MobileNet-based backbones and YOLOv8n are suitable for real-time deployment on airborne platforms. Person re-identification (ReID) features are extracted to distinguish individuals across frames, and SORT/DeepSORT-style tracking is used to maintain identity consistency in dynamic scenes [4]. The system also follows an edge-assisted workflow: the airborne side performs fast detection and primary association, while computationally heavier refinement can be executed on a ground station when communication permits.

3.2. Multi-source Sensor-fusion Localization

Localization accuracy is a decisive factor for rescue navigation and precise delivery. In open outdoor areas, GPS provides global coordinates, but it is susceptible to drift, signal blockage, and multipath interference. The proposed method therefore fuses GPS and BeiDou with IMU measurements to stabilize state estimation. Visual SLAM further refines motion estimation by using ground features and map consistency, thereby compensating for long-term drift in satellite-only localization [2], [5]. In indoor or weak-GNSS conditions, the system transitions to a perception-dominant mode in which visual SLAM and laser SLAM jointly provide map-based localization and target-relative positioning.

State fusion and prediction can be implemented with Kalman filtering for approximately linear state updates and particle filtering for strongly nonlinear or non-Gaussian motion scenarios. This design supports both self-localization of the UAV and relative localization of the target, which is crucial for payload release near victims or supply-demand points. RTK can be introduced as a higher-precision reference when correction services are available [7].

3.3. Precision Delivery Strategy

Payload release in rescue missions is difficult because the final landing point depends on aircraft height, velocity, attitude, wind, and the physical properties of the dropped object. The proposed strategy constructs a projectile model using real-time flight height, speed, attitude, and environmental wind estimates to predict the optimal release point. The strategy distinguishes between lightweight and heavyweight payloads and applies differentiated release parameters to reduce trajectory bias. In addition, the system implements a multi-layer safety mechanism, including space-clearance inspection, dual-condition triggering, and protected payload packaging.

A closed-loop correction mechanism is then formed by visual tracking and flight-control micro-adjustment. During the final approach, the onboard vision module updates the relative target position, and the controller compensates for the residual release error caused by crosswind and short-term attitude transients. Because ADRC suppresses disturbance-driven state fluctuation, the release state is more repeatable, which is beneficial for precision delivery [9], [10].

4. Simulation Experiments and Result Analysis

4.1. Experimental Setup

Table 2. Simulation scenario groups incorporated into the manuscript.

Scenario group	Main variable	Observed engineering implication
Lighting-dependent detection	Sunny, rainy, and night conditions	Infrared cues are important when visible-light contrast is weak
Complex-background recognition	Open area, urban street, forest, ruins	Background clutter and occlusion motivate stronger feature robustness
Distance-dependent recognition	Near-to-far target range	Layered search and close-range confirmation are necessary
Inference frame rate	Lighting-dependent computational load	Model lightweighting is required for stable real-time operation
Multi-person overlap tracking	From no overlap to dense crowd	Occlusion handling and ReID remain critical in rescue crowds
Pose-dependent recognition	Standing, walking, lying, back view	Victim-oriented datasets should include noncanonical postures
Recognition with carried items	Backpack and large obstructions	Partial-body occlusion can change body contour and reduce confidence
Multimodal-fusion comparison	Visible only, infrared only, fusion	Complementary sensing outperforms any single modality
UAV flight-state simulation	3D trajectory, altitude, speed	Dynamic maneuvers must be considered when validating perception
Outdoor positioning error comparison	GPS, BeiDou, fused GNSS, RTK reference	Multi-GNSS fusion improves accuracy; RTK remains the highest-precision baseline

The experimental evaluation was designed as a simulation-oriented campaign following the mission logic of the proposed rescue system. The scene library covered outdoor plazas, parks, and residential-community spaces, together with illumination and background variations relevant to rescue operations. The platform configuration was based on the ZD550 quadrotor described in Section 2, equipped with a visible-light camera, an infrared sensing channel, GNSS modules, IMU, and onboard computing resources for ROS2-based perception and control. Because the uploaded simulation notes primarily reported scenario-wise trends rather than raw point-by-point datasets, the analysis in this section emphasizes comparative behavior, consistency across scenarios, and engineering implications rather than large-sample statistical significance testing. The ten simulation scenario groups incorporated into this manuscript are summarized in Table 2.

4.2. Simulation Results

To make the simulation evidence explicit, each scenario group is discussed with its corresponding figure. The text therefore refers directly to the figure number in the main body and explains the engineering meaning of every plot rather than placing the figures in a cluster.

4.2.1. Recognition Robustness under Environmental Variation

As shown in Fig. 1, person detection remains strongest in favorable illumination such as sunny daytime, where sufficient light and clear texture improve the separability of human features. Detection performance degrades under rainy and nighttime conditions, which confirms the limitations of visible-light sensing in low-contrast scenes. This trend is consistent with the motivation for visible-infrared fusion reported in the literature [3].

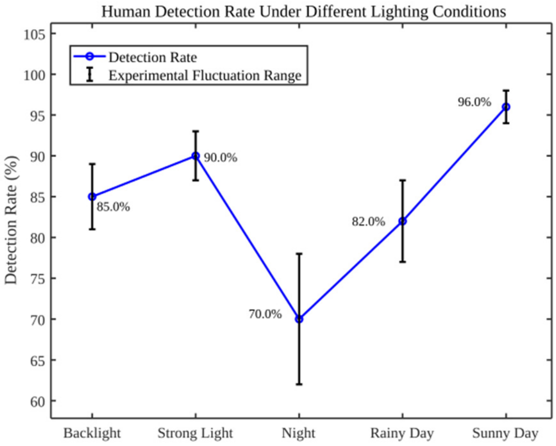

Fig 1. Human detection rate under different lighting conditions.

The practical implication of Fig. 1 is that rescue missions should not rely on a fixed visual threshold across all weather and time conditions. Instead, the sensing pipeline should adapt its confidence strategy according to environmental visibility, while infrared cues help preserve target saliency when visible textures weaken.

Figure 2 presents the recognition results under different background complexities. Open areas provide the highest recognition accuracy because the target contour is clear and background interference is limited, whereas forests, ruins, and dense urban scenes introduce clutter, occlusion, and target-background similarity. The result supports the need for robust feature extraction and tracking in rescue scenes with structural debris and vegetation.

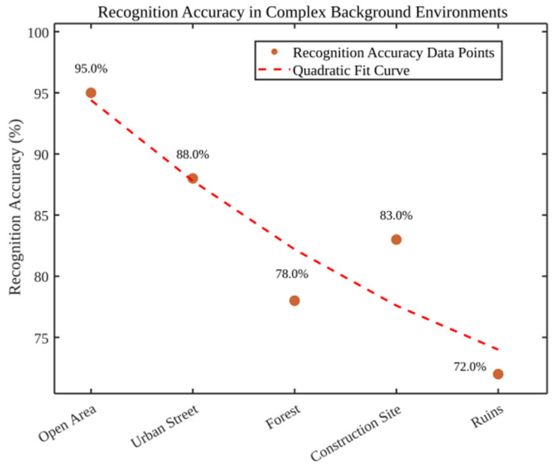

Fig 2. Recognition accuracy in complex background environments.

The engineering interpretation of Fig. 2 is that dataset design and model tuning must include clutter-heavy scenarios rather than only clean pedestrian scenes. Otherwise, the recognition model may overfit to simple backgrounds and underperform in real rescue deployment.

As shown in Fig. 3, recognition reliability gradually decreases as the UAV-to-target distance increases. Once the target occupies fewer pixels, human shape cues and thermal boundaries become less informative. This observation supports a layered search strategy in which long-range scanning is followed by closer confirmation before decision-making or payload release.

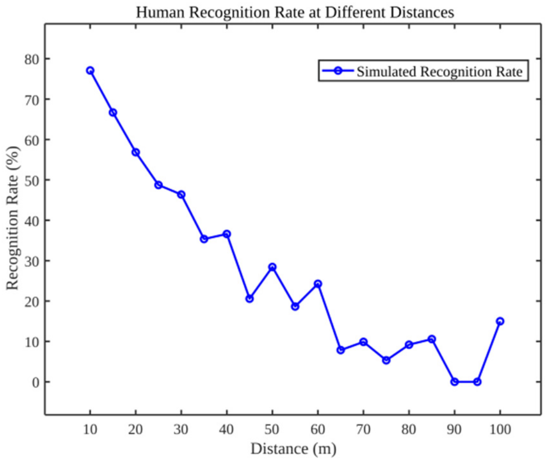

Fig 3. Human recognition rate at different distances

Figure 3 also indicates that perception quality is closely linked to flight planning. Search altitude, camera field of view, and waypoint density should be co-designed so that the target enters a distance band that still supports reliable recognition before the mission transitions to the approach stage.

The computational side is summarized in Fig. 4. The real-time inference frame rate varies with lighting conditions because darker or visually degraded scenes often require stronger enhancement or denoising before detection. Even when recognition remains acceptable, a frame-rate drop can weaken temporal continuity for moving targets.

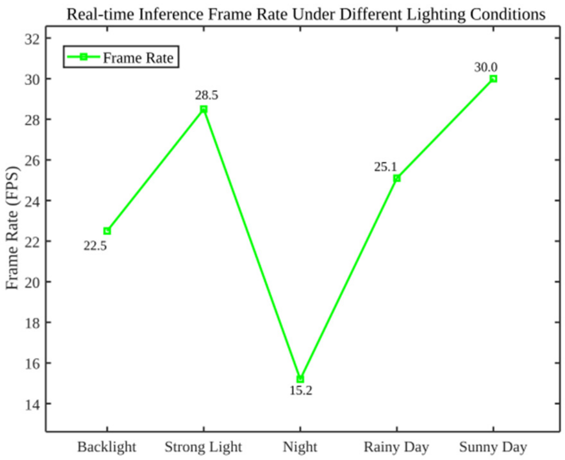

Fig 4. Real-time inference frame rate under different lighting conditions

The implication of Fig. 4 is that lightweight models and efficient preprocessing are indispensable for embedded airborne deployment. This conclusion is aligned with prior work on real-time UAV perception, where system latency must remain bounded to maintain stable tracking [4].

Figure 5 focuses on tracking success when multiple people overlap. The success rate decreases from the non-overlap case to dense-crowd conditions because identity switching and short-term occlusion become more frequent. This is a typical challenge for multi-object tracking and further justifies the use of ReID-supported association [4].

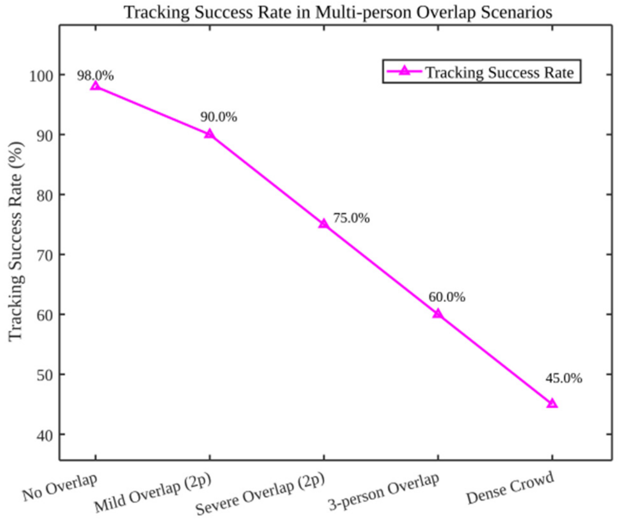

Fig 5. Tracking success rate in multi-person overlap scenarios.

From the perspective of search-and-rescue deployment, Fig. 5 suggests that the system should maintain multi-frame confirmation logic rather than trusting a single-frame identity assignment in crowd-like conditions. Such temporal smoothing can reduce false alarms when victims and rescuers gather in the same area.

Figure 6 compares recognition performance across different human postures. Standing and walking postures remain easier to classify, whereas lying-down and back-view states reduce recognition confidence. This observation is especially important for rescue missions because injured victims are likely to appear in noncanonical postures instead of standard upright poses.

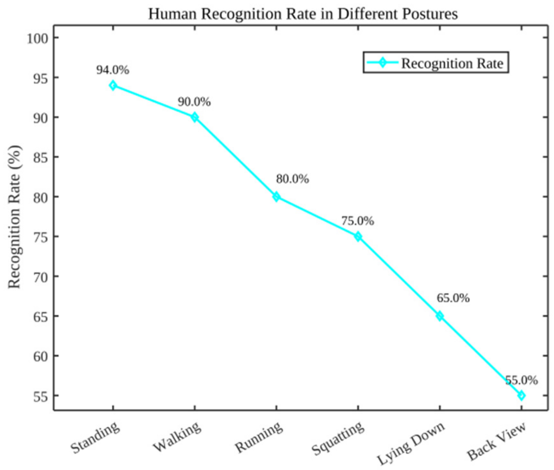

Fig 6. Human recognition rate in different postures.

The main lesson from Fig. 6 is that model training and validation should include victim-oriented pose distributions. Benchmarks dominated by normal pedestrian images may produce an overly optimistic estimate of field performance in disaster scenes.

As illustrated in Fig. 7, carried items and local obstructions also reduce recognition accuracy because they alter body contour and create partial occlusion. Backpacks cause a moderate drop, while larger obstructions have a much stronger influence on the final confidence score.

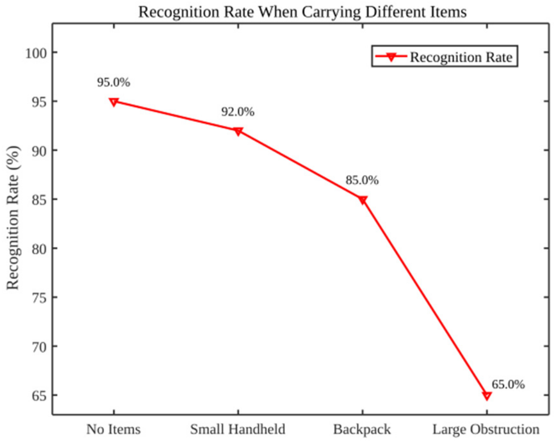

Fig 7. Recognition rate when carrying different items.

Figure 7 reinforces the idea that a rescue-oriented perception system should emphasize part-based or multimodal cues rather than depending solely on full-body appearance. Infrared responses and temporal motion continuity can partially compensate for visual shape disruption [3].

4.2.2. Multimodal Fusion, Flight-state Simulation, and Localization

Figure 8 provides a direct comparison among visible-light-only, infrared-only, and visible-plus-infrared fusion. The fusion scheme delivers the most stable overall performance because the two modalities are complementary: visible-light images preserve fine structure in normal illumination, whereas infrared images maintain target saliency under darkness or partial visual obstruction [3].

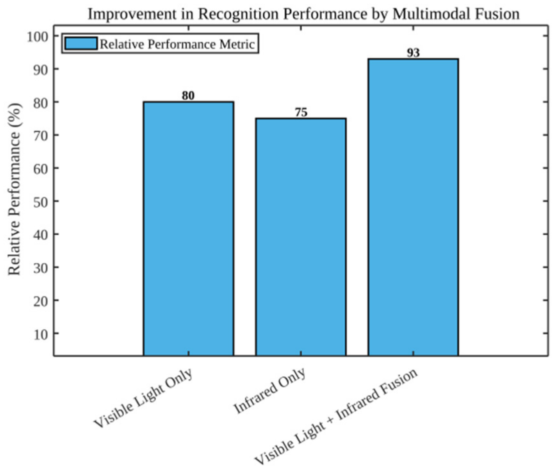

Fig 8. Improvement in recognition performance by multimodal fusion.

The analysis of Fig. 8 confirms the central sensing choice of this paper. Instead of forcing a single modality to handle all weather and all illumination conditions, the system improves robustness by allowing one modality to compensate for the weakness of the other.

The dynamic flight-state simulation is shown in Fig. 9. The trajectory includes climb, descent, turning, acceleration, and deceleration phases, thereby reproducing realistic motion perturbations that can affect onboard imaging and state estimation. Such dynamic validation is important because perception algorithms tested only under static or smooth-motion conditions often overestimate practical field performance.

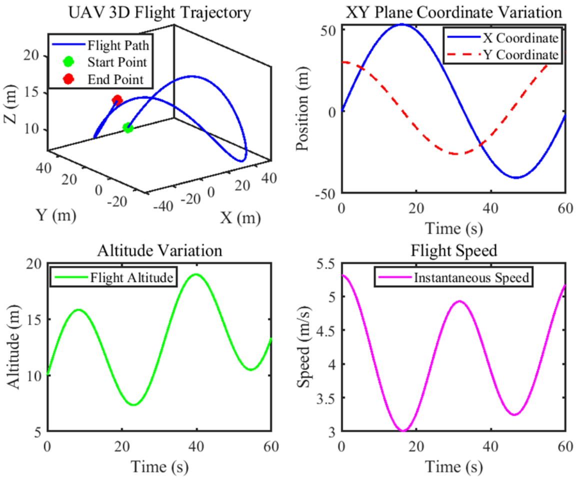

Fig 9. UAV 3D flight trajectory and altitude/speed variation.

Figure 9 also clarifies why control quality matters for perception quality. When the platform experiences aggressive state changes, image blur, viewpoint drift, and localization transients become more likely. ADRC-based disturbance suppression therefore supports not only flight stability but also cleaner visual input for downstream modules [9], [10].

Finally, Fig. 10 compares positioning error in outdoor open-scene conditions. The RTK reference delivers the highest precision, while GPS+BeiDou fusion substantially outperforms either single-satellite solution. This pattern is consistent with the established value of multi-source navigation and high-precision correction in aerial positioning [7].

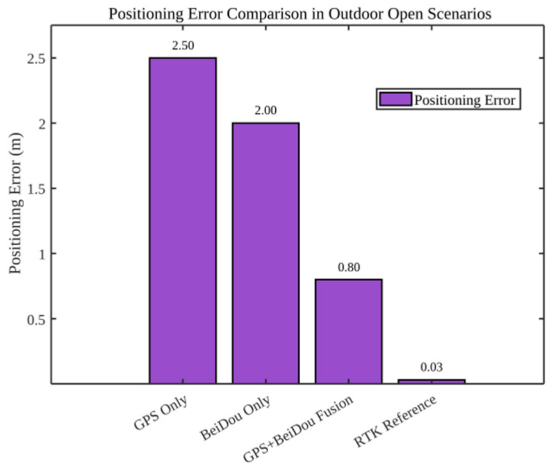

Fig 10. Positioning error comparison in outdoor open scenarios.

The engineering conclusion drawn from Fig. 10 is that the localization stack should be hierarchical. RTK can be used when correction service and infrastructure are available, whereas GPS/BeiDou fusion with visual-inertial refinement provides a practical baseline for broader deployment, especially in semi-structured rescue areas.

4.2.3. Implications for Precision Delivery

Although the uploaded simulation notes focus more heavily on recognition and localization than on payload-drop statistics, the integrated mission analysis still supports several conclusions about delivery performance. First, accurate target confirmation and stable target-relative localization are prerequisites for reducing drop-point bias. Second, ADRC-based stabilization improves the quality of the release state by suppressing attitude fluctuation and short-period disturbance. Third, the ballistic prediction and visual correction loop is especially important when the target is moving or when wind changes during the final approach. In engineering terms, the delivery subsystem benefits directly from the perception and control improvements documented in Figs. 1–10.

4.3. Discussion

From a systems perspective, the most important result is not a single best-case metric, but the consistency of performance gains across heterogeneous scenarios. Multimodal perception improved robustness under illumination changes and occlusion; sensor fusion improved localization reliability; and ADRC strengthened platform stability under disturbance. These gains are mutually reinforcing. Better flight stability yields cleaner image sequences; better perception yields better relative target localization; and better localization enables safer and more accurate delivery. This cross-stage coupling explains why a closed-loop system architecture is preferable to a loosely connected module stack for search-and-rescue UAVs.

Compared with conventional RGB-only recognition and single-source localization pipelines, the proposed system is more suitable for rescue scenes characterized by uncertainty and environmental variation. Nevertheless, several limitations remain. First, the current results are largely simulation-derived and scenario-driven; future work should report standardized quantitative benchmarks and field trials. Second, communication quality may constrain edge-ground collaboration in remote areas. Third, extreme conditions such as heavy smoke, severe wind gusts, and large-scale crowd dynamics still require further study.

5. Conclusion and Future Work

5.1. Conclusion

This paper developed a search-and-rescue UAV system that integrates ADRC with multimodal vision, sensor-fusion localization, and precision delivery. The design addresses three core rescue requirements: robust target recognition, accurate self- and target-localization, and safe payload release. The simulation-oriented evaluation shows that the system performs favorably in difficult conditions including illumination variation, complex backgrounds, target overlap, pose diversity, and long observation distance. The work demonstrates that lightweight multimodal vision can enhance recognition robustness and real-time capability, that multi-source sensor fusion can improve three-dimensional positioning accuracy, and that ballistic prediction combined with closed-loop visual correction can strengthen delivery precision. More broadly, the paper shows that system-level coordination among perception, localization, and control is essential for reliable rescue autonomy.

5.2. Future Work

Future research will focus on three directions. The first is robustness under more extreme environments, including severe weather, smoke, and strong electromagnetic interference. The second is long-range accurate localization and delivery through tighter integration of vision, communication, and map priors. The third is multi-UAV cooperation for simultaneous multi-target detection, localization, and supply delivery. In addition, lightweight large-model techniques may be introduced for higher-level target-intent understanding, while satellite communication and 5G links may support low-latency beyond-visual-line-of-sight coordination.

References

[1] D. Floreano and R. J. Wood, "Science, technology and the future of small autonomous drones," Nature, vol. 521, no. 7553, pp. 460–466, 2015.

[2] D. Scaramuzza, M. C. Achtelik, L. Doitsidis, et al., "Vision-controlled micro flying robots: From system design to autonomous navigation and mapping in GPS-denied environments," IEEE Robotics & Automation Magazine, vol. 21, no. 3, pp. 26–40, 2014.

[3] S. Lee, J. Park, and H. Kim, "Multimodal sensor fusion for human detection in disaster environments using UAVs," IEEE Sensors Journal, vol. 22, no. 15, pp. 15234–15243, 2022.

[4] M. Wang, Q. Zhang, and N. Li, "Research on a multi-target tracking algorithm based on improved YOLOv5 and DeepSORT," Acta Electronica Sinica, vol. 50, no. 8, pp. 1896–1904, 2022.

[5] A. Smith, B. Johnson, and C. Williams, "Visual SLAM in low-light conditions for UAV navigation," in Proc. IEEE International Conference on Robotics and Automation (ICRA), 2021, pp. 4521–4527.

[6] R. Patel, A. Kumar, and S. Gupta, "Deep reinforcement learning for wind disturbance rejection in UAVs," IEEE Transactions on Aerospace and Electronic Systems, vol. 61, no. 1, pp. 234–245, 2025.

[7] W. Zhang, Y. Liu, and L. Wang, "Design of a high-precision agricultural UAV navigation system based on RTK-GPS," Transactions of the Chinese Society of Agricultural Engineering, vol. 37, no. 14, pp. 156–163, 2021.

[8] X.-M. Nguyen and S.-K. Hong, “Improved altitude control algorithm for quadcopter unmanned aerial vehicles,” Applied Sciences, vol. 9, no. 10, Art. no. 2122, 2019.

[9] Z. Gao, “Active disturbance rejection control: A paradigm shift in feedback control system design,” in Proc. American Control Conference, 2006, pp. 2399–2405.

[10] J. Han, “From PID to active disturbance rejection control,” IEEE Transactions on Industrial Electronics, vol. 56, no. 3, pp. 900–906, 2009.