

Research On Automatic Tracking System Of Dynamic Targets Based On UAV

Mengqi Wei and Qi Yang*

School of Mechanical Engineering, Shenyang Ligong University, Shenyang 110159, China.

* Corresponding author. E-mail: yangqi@sylu.edu.cn

Received: Jun. 03, 2026; Accepted: Jun. 28, 2026

Aiming at the critical engineering pain points of visual tracking systems for small multi-rotor UAVs, including onboard payload redundancy, high control loop latency, and the inherent trade-off between response speed and steady-state accuracy in fixed-gain PID controllers, this paper presents a lightweight end-to-end edge vision two-axis gimbal automatic tracking system. To address the link hysteresis problem of traditional relay architectures, a middleware-free direct-connection architecture between OpenMV edge nodes and the Pixhawk flight controller is implemented, significantly reducing system hardware payload and communication latency. To mitigate integral windup and steady-state chattering caused by the mechanical backlash of micro servos, an incremental PID controller with visual front-end signal compensation is designed, which suppresses invalid oscillations through axis-differentiated dead-zone compensation. To overcome the poor adaptability of fixed proportional gains, nonlinear exponential adaptive logic is introduced to resolve the contradiction between rapid target convergence and overshoot suppression. To break through the limitation of flight controller low-level filtering on servo bandwidth, a software-hardware cross-mapping decoupling strategy is adopted to enhance the dynamic response performance of the system. Experimental results demonstrate that in high-frequency transverse reciprocating sinusoidal motion tests, the system can rapidly overcome instantaneous large deviations caused by target mutations; the step response settling time is reduced by 64% compared with the traditional relay architecture; and target reacquisition is completed within 300 ms after 600 ms of target occlusion. While strictly satisfying the onboard payload constraints of small UAVs, the system effectively improves the real-time performance and robustness of visual tracking systems.

Keywords: Small Multi-Rotor Unmanned Aerial Vehicle (SUAV); Visual Servoing; Incremental PID;

MAVLinkCommunication; End-to-End Control

© The Author(s). This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are cited.

http://dx.doi.org/10.6180/jase.202611_34.032

1. Introduction

Small Multi-Rotor Unmanned Aerial Vehicles (SUAVs) have extensive application prospects in scenarios such as environmental monitoring, target reconnaissance, and emergency search and rescue due to their superior maneuverability and rapid deployment capability. To break traditional calculation barriers, deep learning-based visual navigation engines running on ultra-low-power hardware

have demonstrated great potential for autonomous nano-drones [1]. As the core payload enabling SUAVs to perform closed-loop tracking of non-cooperative targets, the two-dimensional visual tracking gimbal directly determines the system's target locking accuracy and response speed.

However, for lightweight platforms with a 450 mm wheelbase, fixed-gain PID controllers in conventional solutions fail to balance rapid convergence under large deviations and steady-state accuracy under small deviations.

The inherent backlash of micro servos also induces integral windup and steady chattering, posing high requirements on the anti-windup performance of the control law [2–4]. With the advancement of edge computing and TinyML technologies, deploying visual pose estimation algorithms on MCU-level nodes has become engineering feasible [5–7]. Existing visual tracking schemes for small UAVs are mainly divided into two technical routes. The first route is the onboard computer relay architecture represented by Raspberry Pi and Jetson Nano. It has strong computing power but usually weighs over 150 g and introduces 100–200 ms of forwarding delay, which exceeds the payload and latency constraints of 450 mm-wheelbase UAVs [8]. The second route is the edge vision node scheme, which has been verified in feasibility studies, but most works adopt fixed-gain PID controllers without targeted compensation for servo nonlinearity and flight controller filtering limitations. Li et al. [8] developed a real-time target detection and tracking system based on monocular vision. Their work verified the application potential of edge nodes on lightweight UAV platforms, but did not address two core issues: the low-level filtering limitation of flight controllers and actuator nonlinearity compensation. Macenski et al. [9] discussed lightweight distributed architectures represented by ROS 2. These architectures provide valuable insights for the efficient design of UAV visual systems, but no low-level control law optimization targeting the nonlinear characteristics of micro servos has been carried out in their work [10].

Compared with typical onboard computer schemes, the proposed direct-connection architecture reduces the system payload by more than 40% and cuts the communication link delay by an order of magnitude. Compared with conventional fixed-gain edge vision solutions, the proposed adaptive control law reduces steady-state jitter by 72% while maintaining fast dynamic response. Aiming at the above practical bottlenecks, this paper proposes a complete lightweight end-to-end edge vision gimbal automatic tracking system for small multi-rotor UAVs. Different from the traditional three-layer relay structure and fixed-parameter control methods widely used in existing research, this work adopts a direct connection structure without an onboard intermediate computer to reduce payload and communication delay. It combines axis-differentiated dead-zone rules with nonlinear adaptive control to overcome the limitations of traditional PID, and further proposes a software-hardware cross-mapping method to break the bandwidth limitation caused by flight controller filtering.

2. System architecture and control algorithm design

This section elaborates the mechanical structure, hardware topology, visual algorithm and control logic of the proposed gimbal system.

2.1. Mechanical Structure Modeling and Overall Hardware Architecture

As shown in Fig. 1, aiming at the compact layout, lightweight design and low payload redundancy requirements of 450 mm wheelbase UAVs, this study first performed high-precision solid modeling of the UAV airframe and its mounted two-axis direct-drive gimbal using SolidWorks prior to physical fabrication.

This 3D model accurately constrains the spatial assembly poses of the actuation servos, OpenMV vision node, and high-strength lightweight U-shaped bracket. Through the digital assembly environment, the mechanical interference margin of the gimbal within the maximum motion envelope ($\pm 60^\circ$) was pre-checked and verified. Meanwhile, the center of gravity offset caused by the camera module was evaluated using the modeling software, which guided the physical counterweighting of the pitch axis. This ensures that the pitch axis tends to natural moment balance in the power-off state, guaranteeing the smoothness of the actual servo mechanism from the mechanical level and significantly reducing the static loaded power consumption during dynamic tracking of the servo.

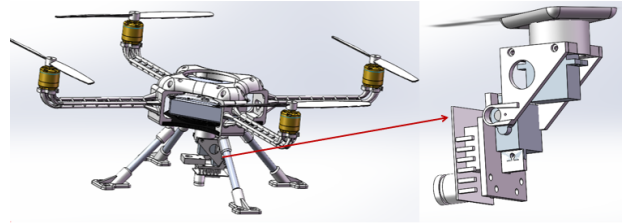


Fig. 1. SolidWorks 3D model of the multi-rotor UAV and two-axis direct-drive gimbal.

The quantitative mechanical parameters of the gimbal are obtained through physical measurement and structural calculation. The total weight of the two-axis gimbal reaches 78 g, including a 22 g lightweight U-shaped bracket, two MG90S micro servos with a total weight of 14 g, a 28 g OpenMV vision module and 14 g of connecting wires. The mechanical rotation range of both pitch and yaw axes is set to $\pm 60^\circ$, and the maximum mechanical backlash of MG90S servos is less than 0.8° according to official specifications and physical tests. The vision module creates a 2.1 mm gravity offset on the pitch axis, and a 3.2 g counterweight is

installed to balance the structure, which reduces the static load torque by 67%. The rotational inertia of the yaw axis is $1.28 \times 10^{-4} \text{ kg} \cdot \text{m}^2$, and the rotational inertia of the pitch axis is $4.15 \times 10^{-5} \text{ kg} \cdot \text{m}^2$.

After establishing the mechanical actuation mechanism, to break the bottleneck of traditional centralized computing, this system designs a minimalist heterogeneous distributed architecture.

The overall system topology is shown in Fig. 2.

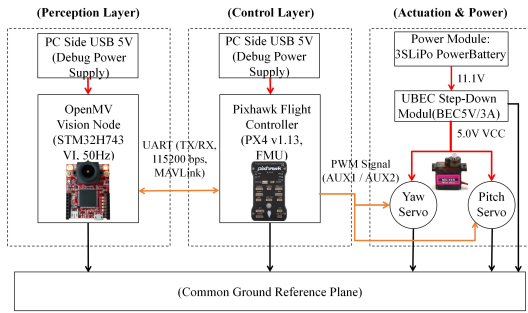


Fig. 2. Physical system topology diagram.

Logically, the system is divided into three core subsystems: the Perception Layer, the Control Layer, and the Actuation & Power Layer. The Perception Layer employs an OpenMV 4 Cam vision node based on the STM32H743 main control chip, with an algorithm operating frequency of 50 Hz. The Control Layer adopts a Pixhawk 2.4.8 flight controller equipped with PX4 v1.13 firmware [11]. The system eliminates the traditional onboard intermediate computer. Image acquisition, target feature extraction, and incremental PID control calculation are all fully closed-loop implemented within the edge vision node. The calculated attitude control commands are directly sent to the underlying mixer of the flight controller, which reduces the system physical payload and communication latency caused by protocol forwarding to a certain extent [9, 12].

2.2. Edge Vision Feature Extraction and Decoupling Mechanism

The overall visual servo processing and control law calculation flow of the system is shown in Fig. 3.

Starting from the underlying CMOS image acquisition, target extraction is completed sequentially through spatial orthogonal decoupling and morphological filtering. The extracted target information is then fed into a discrete PID controller with an anti-integral windup mechanism. Finally, the calculated control increments are encapsulated into MAVLink messages to drive the gimbal.

Real-world operation scenarios are often accompanied

by drastic illumination variations. To enhance the robustness of target feature extraction, this system adopts the CIE LAB Color Space [13] for image processing. This space enables orthogonal decoupling of luminance information (L channel) from chrominance information. To realize the nonlinear mapping from underlying RGB pixels to the LAB space, the system first converts them to the standard XYZ color space, and then performs the following orthogonal decoupling transformation:

$$\begin{cases} L^* = 116f\left(\frac{Y}{Y_n}\right) - 16 \\ a^* = 500\left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right] \\ b^* = 200\left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right] \end{cases} \quad (1)$$

where X_n, Y_n, Z_n are the tristimulus values of the reference white point, and the nonlinear function $f(t)$ is defined as:

$$f(t) = \begin{cases} t^{1/3}, & t > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3}\left(\frac{29}{6}\right)^2 t + \frac{16}{116}, & \text{otherwise} \end{cases} \quad (2)$$

This transformation effectively decouples the interference of illumination intensity on the subsequent feature segmentation of the a^* channel (red-green axis) and b^* channel (yellow-blue axis). The system first dynamically relaxes the threshold range of the L channel to adapt to extreme illumination environments such as strong backlight or shadow; then applies strict band-pass filtering to the a^* channel and b^* channel to accurately segment the target red region. After completing connected component merging and noise filtering through opening and closing operations, based on the fundamental mathematical principles of digital image processing [14], and classic visual servo control approaches [15], the system calculates the centroid coordinates (C_x, C_y) of the largest connected component of the target using spatial image moments. The discrete calculation models of the zero-order and first-order spatial moments are as follows:

$$M_{pq} = \sum_x \sum_y x^p y^q I(x, y) \quad (3)$$

$$C_x = \frac{M_{10}}{M_{00}}, \quad C_y = \frac{M_{01}}{M_{00}} \quad (4)$$

Where M_{pq} is the $(p + q)$ -th order spatial moment of the image, and $I(x, y)$ is the binary pixel distribution of the target region after band-pass filtering and morphological processing (the pixel value of the target region is 1, and the background is 0). M_{00} represents the area of the largest connected component of the target, while M_{10} and M_{01} represent the first-order moments of the image relative to

the x-axis and y-axis within this connected component, respectively.

This system adopts QQVGA resolution for image input. Since the QQVGA resolution of the OpenMV 4 Cam H7 is 160×120 , the physical center of the image is set to $(80, 60)$. Combined with the centroid coordinates obtained from pre-processing, the system derives the two-dimensional pixel deviation vector required by the control system:

$$e(k) = \begin{bmatrix} e_x(k) \\ e_y(k) \end{bmatrix} = \begin{bmatrix} C_x(k) - 80 \\ C_y(k) - 60 \end{bmatrix} \quad (5)$$

Where $e(k)$ is the discrete attitude deviation vector of the k -th sampling period. This deviation vector $e(k)$ is directly used as the feedback quantity of the control system and fed into the underlying closed-loop control logic.

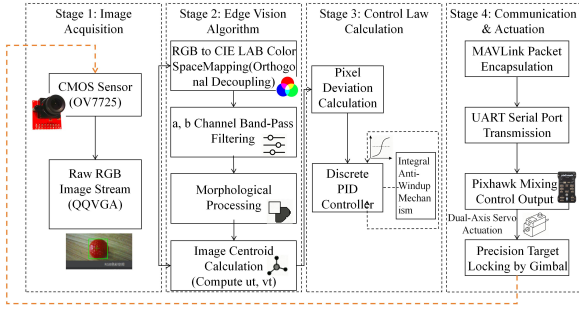


Fig. 3. Overall flowchart of the edge visual servo tracking algorithm and control system.

2.3. Axis-Differentiated Dead-Zone and Adaptive- K_p Based Incremental PID Controller

Aiming at the contradiction between "response speed and steady-state chattering" faced by micro direct-drive gimbals in dynamic tracking, this paper abandons the traditional fixed-parameter PID and designs a set of control laws combining "axis-differentiated dead-zone compensation" and "nonlinear adaptive proportional gain K_p ".

Axis-Differentiated Dead-Zone Compensation Mechanism: Since the yaw axis carries the entire physical load of the pitch servo and the camera, its moment of inertia is significantly larger than that of the pitch axis. According to the strict equations of motion derived for a two-axes gimbal system [16], as well as composite decoupling control frameworks designed for high-performance gimbal servo configurations [17], the mechanical backlash of micro metal gear servos is extremely prone to inducing high-frequency resonance on this axis. To address this issue, the system introduces an "axis-differentiated dead-zone compensation

function" before the error is input to the controller [18]:

$$e(k) = \begin{cases} 0, & |E(k)| \leq \delta \\ E(k), & |E(k)| > \delta \end{cases} \quad (6)$$

The dead-zone thresholds are calibrated based on the 0.8° mechanical backlash of MG90S servos and the corresponding pixel angular resolution of the visual system. The yaw axis bears a larger load and is more susceptible to backlash-induced oscillation, so its threshold is set to $\delta_{yaw} = 6$ pixels to fully cover the invalid deviation caused by gear clearance. For the low-inertia pitch axis, the threshold is set to $\delta_{pitch} = 2$ pixels to suppress micro oscillations while preserving high-frequency tracking agility.

This design matches the inertia difference between the two axes, suppressing yaw-axis resonance while preserving pitch-axis tracking sensitivity.

Nonlinear Adaptive Proportional Gain K_p Control Law: A variety of improved and adaptive PID methods have been widely studied for quadrotor UAV control systems [18–20]. To thoroughly solve the contradiction between fast convergence under large deviations and steady-state overshoot under small deviations, this paper proposes an exponential adaptive gain mapping law based on the absolute value of pixel error.

Its mathematical model is as follows:

$$K_p(k) = K_{p\max} - (K_{p\max} - K_{p\min}) \cdot \exp(-\lambda |e'(k)|) \quad (7)$$

The parameter tuning of the controller follows a systematic procedure combining theoretical derivation and experimental verification. First, initial PID parameters are derived via the Ziegler-Nichols empirical method based on the open-loop step response data of the gimbal servo system, yielding initial values of $K_p = 0.55$, $K_i = 0.018$, and $K_d = 0.015$. Second, fixed-parameter fine-tuning is performed targeting the nonlinear characteristics of micro servos. Considering the mechanical backlash and integral windup risk, the integral coefficient K_i is reduced to 0.012 to avoid control divergence during brief target loss; the differential coefficient K_d is adjusted to 0.008 to filter high-frequency visual noise while providing sufficient damping for overshoot suppression. The optimal fixed proportional gain $K_p = 0.60$ is finally determined through multiple sets of comparative step response tests. Third, adaptive parameters are calibrated on this basis. The gain bounds $K_{p\max} = 0.60$ and $K_{p\min} = 0.10$ are set according to the critical oscillation gain and steady-state smoothness requirements, and the decay factor $\lambda = 0.025$ is optimized via single-variable experiments to ensure smooth gain transition without chattering or response lag.

where $K_{p\max}$ is the upper bound of the gain, used to provide transient step response burst capability under large deviations; $K_{p\min}$ is the lower bound of the gain, used to ensure steady-state damping smoothness under small deviations; and λ is the nonlinear decay factor. In the discrete difference equation of the incremental PID controller, the integral term K_i bears the main steady-state driving force, while K_d acts as the dynamic damping. The final attitude increment output $\Delta u(k)$ of the system is:

$$\Delta u(k) = K_p(k) [e'(k) - e'(k-1)] + K_i e'(k) + K_d [e'(k) - 2e'(k-1) + e'(k-2)] \quad (8)$$

2.4. End-to-End Communication Reconstruction and Underlying Filtering Cross-Mapping

To break through the computing power and latency bottlenecks caused by protocol forwarding in traditional onboard computers, this system employs the ustruct module in MicroPython on the edge vision node, and the communication link strictly adheres to the MAVLink V1.0 protocol standard [12]. The vision node sends HEARTBEAT packets to the flight controller at a fixed high frequency to maintain underlying control authority. Meanwhile, the calculated yaw angle and pitch angle increments are encapsulated into the MOUNT_CONTROL data frame with MSG_ID=76 and the encapsulated messages are directly transmitted via the UART serial bus at 115200 bps. In the underlying mixer design of open-source flight controllers (e.g., PX4), low-pass filters are applied to partial logic channels by default for smooth aerial photography requirements [11]. Although this mechanism suppresses high-frequency jitter, it easily induces physical phase lag in the servo system [19]. Early experimental observations reveal that if the pitch axis is directly connected to the default channel, the control signal will be "hijacked" by the underlying filtering, resulting in slow gimbal response. To eliminate this hardware constraint, this paper proposes and implements a "Hardware-Software Cross-Mapping Strategy" on the physical platform:

High-Bandwidth Pass-Through for Pitch Axis: The pitch motion needs to resist gravity load and provide transient response capability during target mutation. Therefore, the system redirects the pitch axis control data to the unfiltered high-bandwidth pass-through channel (Param 1) inside the flight controller at the logical layer, and completes the corresponding mapping of hardware pins at the physical layer to ensure ultra-fast step response.

Smooth Tracking for Yaw Axis: The yaw motion is mainly responsible for large-scale horizontal field-of-view tracking, with core requirements of frame smoothness and anti-oscillation. The system maps the yaw axis logical data

to the channel with low-pass filtering (Param 3), utilizing the native hardware filtering of the flight controller to eliminate lateral high-frequency jitter and achieve complete decoupling of the dual-degree-of-freedom control strategy.

3. Physical experiment verification and data analysis

3.1. Experimental Environment and Test Platform

To verify the control performance of the proposed visual tracking system in real physical scenarios, a hardware experimental platform based on Pixhawk 2.4.8 flight controller and two-axis direct-drive gimbal is constructed. The physical appearance of core mechanical and sensing components is shown in Fig. 4.

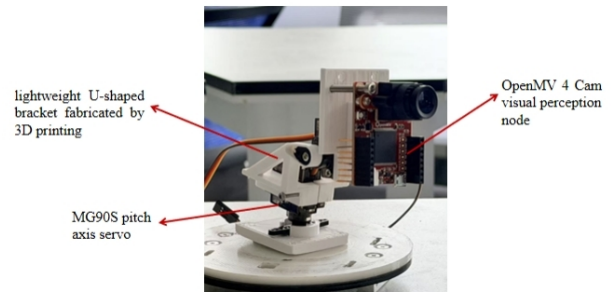


Fig. 4. Photograph of the hardware test bench for the visual gimbal tracking system.

The mechanical body adopts lightweight 3D printing materials to comply with strict load constraints of small multirotor unmanned aerial vehicles. The visual processing unit uses OpenMV 4 Cam H7 module equipped with STM32H743VI main controller, and MG90S metal gear servos serve as actuators. During joint debugging, the baud rate of MAVLink serial port is set to 115200 bps, and the sampling & output frequency of closed-loop control is configured at 40 Hz.

3.2. Parameter Tuning and Adaptive Experiments of Incremental PID

Before establishing the control baseline, this study first conducted comparative experiments with fixed gains. Three representative proportional gain parameters ($K_p = 0.10, K_p = 0.60, K_p = 1.50$) were selected for step response tests.

As shown in Fig. 5, the results indicate that the system exhibits under-response at $K_p = 0.10$ and severe oscillation at $K_p = 1.50$. The value $K_p = 0.60$ is the optimal compromise parameter under fixed settings, achieving a good balance between dynamic tracking speed and steady-state

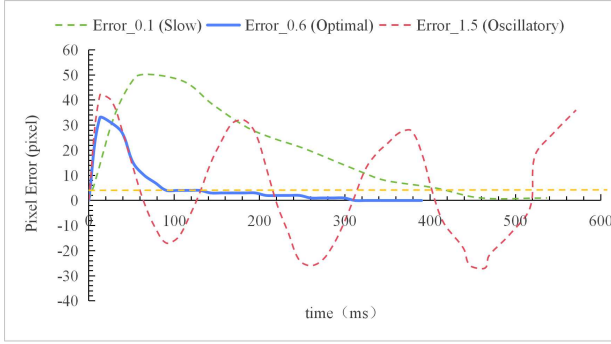


Fig. 5. Error convergence curves under different proportional gains.

anti-oscillation capability. However, the fixed gain of traditional linear PID is inherently limited by the performance trade-off between transient response and steady-state accuracy. To further break the physical response limits of linear control — that is, to pursue faster transient response than the optimal fixed-parameter solution while achieving more extreme steady-state smoothness — this study introduces the nonlinear adaptive K_p control law.

3.3. Step Response and Underlying Hardware Decoupling Experiments

To verify the effectiveness of the end-to-end direct connection architecture in boosting servo closed-loop bandwidth, step response and underlying decoupling experiments are carried out with optimal PID parameters ($K_p = 0.60$). In conventional PX4 firmware architecture, the underlying mixer imposes heavy low-pass filtering (LPF) on auxiliary output (AUX) channels by default, causing noticeable physical phase lag in visual servo applications. To overcome this inherent hardware limitation, a hardware-level cross-mapping decoupling strategy via pin rearrangement is adopted on the test platform. With servo signal pins properly swapped, the pitch axis demanding rapid transient response is assigned to the unfiltered Param 1 pass-through channel inside the flight controller. Large-amplitude step position commands are exerted on the system.

As illustrated in Fig. 6, filtering-induced hysteresis is fully eliminated after pin swapping takes effect. Comparative data shows that the settling time, the duration for error convergence within steady-state range, is about 500 ms under original filtered channels. After hardware cross-mapping decoupling, the settling time drops sharply to 180 ms, achieving a 64% reduction. The results validate that pin rearrangement and channel cross-mapping can fully exert the transient performance of direct-drive actuators.

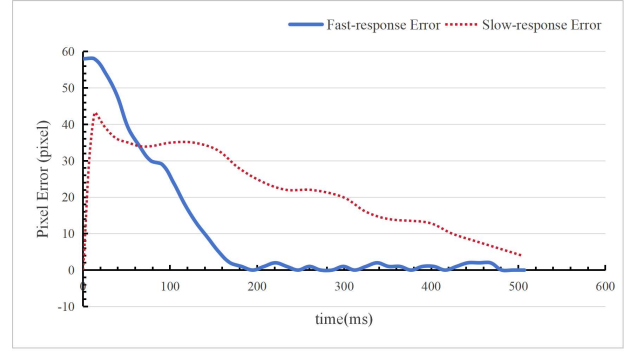


Fig. 6. Step response curves before and after hardware decoupling.

3.4. Anti-Interference and Robustness Experiments

To comprehensively evaluate the system's robustness under non-ideal conditions, light intensity mutation and physical occlusion comparison experiments are conducted in this section. The test distance is 40 cm. The control group adopts a traditional position-based control architecture with RGB threshold segmentation and fixed gain ($K_p = 0.60$); the experimental group employs the proposed CIE LAB orthogonal decoupling perception and adaptive incremental PID control law. The pixel error comparison curves of the yaw axis for both systems are shown in Fig. 7.

During the sudden light intensity disturbance phase (350–650 ms strong illumination interference), the control group was affected by luminance-chrominance coupling, which made the target feature fade out and the tracking pixel error diverge to approximately 70 pixels. In contrast, the proposed method effectively suppresses illumination interference via luminance-chrominance decoupling. The centroid calculation error is stabilized within ± 5 pixels, which accounts for only 7% of the peak error in the control group. In the instantaneous physical occlusion test, although the system experienced a brief loss of lock due to feature disappearance, the underlying actuator did not exhibit divergent or erratic rotation. Once the occlusion was removed, the visual node completed target reacquisition within 300 ms after the 600 ms occlusion ended, and drove the gimbal to return to the steady-state tracking range with an error of less than 5 pixels. The experiments confirm that the proposed system possesses high engineering robustness and self-recovery capability when facing complex light disturbances and transient occlusions.

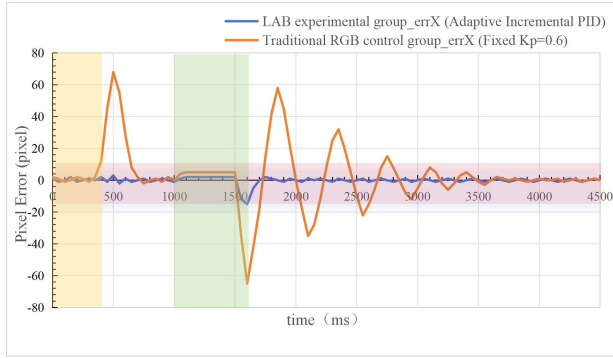


Fig. 7. Robustness curves under occlusion and illumination.

4. Conclusion and prospect

4.1. Conclusion of the Thesis

This thesis addresses the Size, Weight, and Power (SWaP) constraints and response latency issues in dynamic target tracking for small multirotor UAVs, and designs a lightweight end-to-end edge vision gimbal system. The proposed system features low cost, compact size and strong practicability. It can be widely applied to scenarios such as UAV power line inspection, field reconnaissance and emergency search and rescue. The lightweight design effectively extends the flight endurance of small UAVs, and the robust anti-interference performance ensures stable tracking in complex outdoor illumination environments. The main contributions are summarized as follows:

First, an end-to-end edge-aware interconnection architecture is constructed. Visual processing is offloaded to the OpenMV node, which connects directly to the flight controller via encapsulated MAVLink messages. This replaces the traditional onboard relay computer, effectively reducing the system's hardware load and communication latency.

Second, an adaptive incremental PID control law with axis-differentiated dead zones is proposed. Differentiated dead-zone thresholds are set according to the different rotational inertia characteristics of the two axes. Through nonlinear adaptive logic, both the convergence speed under large deviations and the steady-state accuracy under small deviations are balanced, reducing the steady-state jitter amplitude of the yaw axis by 72% compared with the fixed-gain scheme when the target is stationary. It also performs well in anti-windup control during brief target loss.

4.2. Limitations and Future Work

Although the compact, lightweight direct-drive gimbal system designed in this thesis has achieved satisfactory engineering verification results in two-dimensional tracking tasks, there is still room for optimization due to hardware cost constraints and the project development cycle. In view of the current limitations, future research can be further expanded from the following aspects:

Onboard flight tests will be carried out on a 450 mm-wheelbase multi-rotor platform to verify the tracking performance under wind disturbance and airframe motion coupling, and feedforward compensation will be introduced to further improve the tracking accuracy in dynamic flight scenarios.

Migration of visual algorithms to lightweight deep learning architectures. The current system performs target segmentation based on the LAB color space and morphological features. While highly computationally efficient, it is vulnerable to interference from "false targets" in complex backgrounds with colors highly similar to the target. Future work could explore deploying quantized lightweight convolutional neural networks (e.g., YOLOv8-nano) on STM32H7 edge devices [21], upgrading simple "color block tracking" to "feature tracking" with semantic understanding, thereby further improving the system's anti-interference capability [22].

Upgrading the actuation system to brushless direct-drive (FOC) technology. The micro metal gear servos used in this system are low-cost and easy to drive, but the inherent mechanical backlash and response dead zones of their gear trains limit the ultimate closed-loop bandwidth of the system. Future research could investigate the combination of micro brushless motors and magnetic encoders, paired with field-oriented control (FOC) algorithms, to achieve smoother, high-precision attitude response at the microsecond level.

Furthermore, high-frequency serial lines, servo PWM signals and high-current power lines coexist in an extremely tight layout. Future hardware optimization will follow electromagnetic compatibility (EMC) guidelines [23] to suppress coupling noise, and integrate efficient optical flow and stereo vision algorithms [24] to combine target tracking with proactive obstacle avoidance in unknown environments.

Acknowledgment

This research was funded by the Fundamental Scientific Research Expenses for the Universities of Liaoning Province, grant numbers LJ212410144022, LNYJG2023077 and LJ232410144074, and project number 2025-MSLH-589.

References

- [1] D. Palossi, A. Loquercio, F. Conti, E. Flamand, D. Scaramuzza, and L. Benini, (2019) "A 64-mW DNN-Based Visual Navigation Engine for Autonomous Nano-Drones" **IEEE Internet of Things Journal** 6(5): 8357–8371. DOI: [10.1109/JIOT.2019.2917066](https://doi.org/10.1109/JIOT.2019.2917066).
- [2] P. Corke. "Mobile Robot Vehicles". In: *Robotics, Vision and Control: Fundamental Algorithms in MATLAB*. Springer, 2017, 99–124. DOI: [10.1007/978-3-319-54413-7_4](https://doi.org/10.1007/978-3-319-54413-7_4).
- [3] S. Tarbouriech and M. Turner, (2009) "Anti-Windup Design: An Overview of Some Recent Advances and Open Problems" **IET Control Theory & Applications** 3(1): 1–19. DOI: [10.1049/iet-cta:20070435](https://doi.org/10.1049/iet-cta:20070435).
- [4] M. Al Radi, M. N. AlMallahi, and A. S. Al-Sumaiti, (2024) "Progress in Artificial Intelligence-Based Visual Servoing of Autonomous Unmanned Aerial Vehicles (UAVs)" **International Journal of Thermofluids**: DOI: [10.1016/j.ijft.2024.100590](https://doi.org/10.1016/j.ijft.2024.100590).
- [5] P. Warden and D. Situnayake. *TinyML: Machine Learning with TensorFlow Lite on Arduino and Ultra-Low-Power Microcontrollers*. O'Reilly Media, 2019.
- [6] S. Peng, Y. Liu, Q. Huang, X. Zhou, and H. Bao. "PVNet: Pixel-Wise Voting Network for 6DoF Pose Estimation". In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2019, 4561–4570. DOI: [10.1109/CVPR.2019.00469](https://doi.org/10.1109/CVPR.2019.00469).
- [7] Z. Ning, H. Hu, X. Wang, L. Guo, S. Guo, G. Wang, and X. Gao, (2024) "Mobile Edge Computing and Machine Learning in the Internet of Unmanned Aerial Vehicles: A Survey" **ACM Computing Surveys** 56(1): 1–31. DOI: [10.1145/3604933](https://doi.org/10.1145/3604933).
- [8] Z. Li, D. Li, J. Yang, and H. Li. "A Real-Time Monocular Vision-Aided Target Detection and Tracking for Micro UAV in GPS Denied Environment". In: *Lecture Notes in Electrical Engineering*. Springer, 2022. DOI: [10.1007/978-981-15-8155-7_255](https://doi.org/10.1007/978-981-15-8155-7_255).
- [9] S. Macenski, T. Foote, B. Gerkey, C. Lalancette, and W. Woodall, (2022) "Robot Operating System 2: Design, Architecture, and Uses in the Wild" **Science Robotics** 7(66): DOI: [10.1126/scirobotics.abm6074](https://doi.org/10.1126/scirobotics.abm6074).
- [10] P. McEnroe, S. Wang, and M. Liyanage, (2022) "A Survey on the Convergence of Edge Computing and AI for UAVs: Opportunities and Challenges" **IEEE Internet of Things Journal** 9(17): 15435–15459. DOI: [10.1109/JIOT.2022.3176400](https://doi.org/10.1109/JIOT.2022.3176400).
- [11] L. Meier, P. Tanskanen, L. Heng, G. H. Lee, F. Fraundorfer, and M. Pollefeys, (2012) "PIXHAWK: A Micro Aerial Vehicle Design for Autonomous Flight Using On-board Computer Vision" **Autonomous Robots** 33(1): 21–39. DOI: [10.1007/s10514-012-9281-4](https://doi.org/10.1007/s10514-012-9281-4).
- [12] A. Koubâa, A. Allouch, M. Alajlan, Y. Javed, A. Belghith, and M. Khalgui, (2019) "Micro Air Vehicle Link (MAVlink) in a Nutshell: A Survey" **IEEE Access** 7: 87658–87680. DOI: [10.1109/ACCESS.2019.2924410](https://doi.org/10.1109/ACCESS.2019.2924410).
- [13] A. M. Nascimento, A. V. Brito, M. Saska, and T. P. Nascimento, (2026) "A Review on Vision-Based Control for Multi-Rotor Aerial Vehicles" **Robotics and Autonomous Systems** 198: DOI: [10.1016/j.robot.2026.105366](https://doi.org/10.1016/j.robot.2026.105366).
- [14] R. Szeliski. *Computer Vision: Algorithms and Applications*. 2nd. Springer, 2022. DOI: [10.1007/978-3-030-34372-9](https://doi.org/10.1007/978-3-030-34372-9).
- [15] F. Chaumette and S. Hutchinson, (2006) "Visual Servo Control. I. Basic Approaches" **IEEE Robotics & Automation Magazine** 13(4): 82–90. DOI: [10.1109/MRA.2006.250573](https://doi.org/10.1109/MRA.2006.250573).
- [16] B. Ekstrand, (2001) "Equations of Motion for a Two-Axes Gimbal System" **IEEE Transactions on Aerospace and Electronic Systems** 37(3): 1083–1091. DOI: [10.1109/7.953259](https://doi.org/10.1109/7.953259).
- [17] N. V. Phuong, N. T. Tuyen, N. D. Khanh, and N. N. Hung. "Designing a Robust Control System Based on Sliding Mode Control for Two-axis Gimbal Systems". In: *2024 Conference of Young Researchers in Electrical and Electronic Engineering (ElCon)*. IEEE, 2024, 455–458. DOI: [10.1109/ElCon61730.2024.10468229](https://doi.org/10.1109/ElCon61730.2024.10468229).
- [18] J. Kuang and M. Chen, (2024) "Adaptive Sliding Mode Control for Trajectory Tracking of Quadrotor Unmanned Aerial Vehicles Under Input Saturation and Disturbances" **Drones** 8(11): 614. DOI: [10.3390/drones8110614](https://doi.org/10.3390/drones8110614).
- [19] I. López-Sánchez and J. Moreno-Valenzuela, (2023) "PID Control of Quadrotor UAVs: A Survey" **Annual Reviews in Control** 56: 100900. DOI: [10.1016/j.arcontrol.2023.100900](https://doi.org/10.1016/j.arcontrol.2023.100900).
- [20] A. Khalid, Z. Mushtaq, S. Arif, K. Zeb, M. A. Khan, and S. Bakshi, (2023) "Control Schemes for Quadrotor UAV: Taxonomy and Survey" **ACM Computing Surveys**: DOI: [10.1145/3617652](https://doi.org/10.1145/3617652).
- [21] M.-Y. Ma, S.-E. Shen, and Y.-C. Huang, (2023) "Enhancing UAV Visual Landing Recognition with YOLO's Object Detection by Onboard Edge Computing" **Sensors** 23(21): 8999. DOI: [10.3390/s23218999](https://doi.org/10.3390/s23218999).

- [22] C. Wang, D. Xu, Y. Zhu, R. Martín-Martín, C. Lu, L. Fei-Fei, and S. Savarese. "DenseFusion: 6D Object Pose Estimation by Iterative Dense Fusion". In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2019, 3343–3352. DOI: [10.1109/CVPR.2019.00346](https://doi.org/10.1109/CVPR.2019.00346).
- [23] Z. Yu, Y. Zhang, B. Jiang, and C.-Y. Su. *Fault-Tolerant Cooperative Control of Unmanned Aerial Vehicles*. Springer Nature Singapore, 2023. DOI: [10.1007/978-981-99-7661-4](https://doi.org/10.1007/978-981-99-7661-4).
- [24] A. K. Kamath, Z. Zheng, and M. Feroskhan, (2025) "Depth-Independent Augmented Dynamics Visual Servoing of Multirotors" **IEEE Transactions on Aerospace and Electronic Systems** 61(2): 5444–5453. DOI: [10.1109/TAES.2024.3496418](https://doi.org/10.1109/TAES.2024.3496418).