Proceedings of the 4th International Conference on Mechatronics and Smart Systems
DOI: 10.54254/2755-2721/2026.33602

Pixel-to-Angle Mapping and Angle-Domain Incremental Control for Vision-Guided Gimbal Target Tracking

Shengye Zhou

Glasgow College, University of Electronic Science and Technology of China (UESTC), Chengdu, China
2023190904022@std.uestc.edu.cn

Abstract. This paper presents a vision-guided two-axis gimbal tracking method based on pixel-to-angle mapping and angle-domain incremental control. Instead of directly feeding image-plane pixel errors into a conventional pixel-domain controller, the proposed pipeline first converts the target deviation from the image center into physically meaningful yaw and pitch angle errors using camera calibration parameters. These angle errors are then used to generate angular-velocity commands and motor RPM references through an angle-domain control chain. To improve practical deployability under sensing noise, abrupt target displacement, and actuator limitations, dead-zone logic, command saturation, and low-pass filtering are incorporated into the control loop. The paper further organizes the method into a complete conference-style presentation including system overview, imaging geometry, control architecture, and evaluation protocol. The resulting formulation provides a physically consistent interface between visual measurements and gimbal actuation, and offers a more interpretable basis for future extensions with encoders, IMU feedback, and more advanced front-end detectors and trackers.

Keywords: visual servoing, gimbal target tracking, pixel-to-angle mapping, angle-domain control, active vision

1. Introduction

Vision-guided target tracking is a fundamental capability in surveillance, active perception, autonomous landing, and persistent observation. Its main objective is to keep the target near the image center despite motion, disturbance, and field-of-view limits. Classical visual servoing literature has established the distinction between image-based visual servoing (IBVS) and position-based visual servoing (PBVS), and has shown that the definition of visual error strongly influences controller behavior, stability, and implementation complexity [1-4]. At the same time, camera calibration provides the geometric bridge that relates image coordinates to physically meaningful viewing directions [1].

In many lightweight gimbal systems, however, a simplified engineering pipeline is still commonly adopted: target detection, pixel error computation, and direct pixel-domain control. Although this workflow is easy to implement, pixel error is not a physical quantity with a fixed scale. The same deviation measured in pixels may correspond to different line-of-sight errors under

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.33602

different focal lengths, image resolutions, and fields of view [5-8]. Moreover, the mapping from image displacement to viewing angle becomes increasingly nonlinear near the image boundary. Recent work on inertially stabilized platforms, gimbal-assisted perception, vision-based UAV landing, and airborne pan-tilt-zoom visual search increasingly emphasizes the need to couple imaging geometry [9,10], field-of-view constraints, and controller design within a unified physical framework [11-14].

Meanwhile, the front end of target detection and tracking has progressed rapidly. Modern real-time detectors and trackers, such as ByteTrack, YOLOv7, RT-DETR, and recent track-anything or generalizable tracking models, can provide more stable target localization at practical frame rates [15-18]. This shifts the bottleneck from merely detecting the target to converting visual measurements into a control quantity that remains physically interpretable and transferable across camera configurations [19,20].

To address this issue, this paper proposes a pixel-to-angle mapping and angle-domain incremental control framework for vision-guided gimbal tracking. The main contributions are threefold. First, the target deviation in the image plane is explicitly converted into yaw and pitch angle errors using camera intrinsic parameters. Second, an angle-domain control chain is designed to generate angular-velocity and RPM commands from physically meaningful angular errors. Third, a complete manuscript-level formulation is constructed, including system overview, imaging geometry, controller architecture, and evaluation protocol, so that the method can be cleanly extended to future quantitative comparisons and hardware validation.

2. System overview and hardware platform

Figure 1 summarizes the overall workflow of the proposed vision-guided gimbal tracking system. The closed loop consists of image acquisition, target detection and center extraction, pixel error computation, pixel-to-angle mapping, angle-domain incremental control, RPM conversion, motor command transmission, and gimbal actuation with recentering feedback. Unlike the conventional chain of "pixel error -> controller -> motor command," the proposed method introduces an explicit geometric mapping stage before control. The target deviation is first converted into yaw and pitch angle errors through camera calibration, and only then passed to the controller. This modification turns the control variable from an image-plane quantity into a physically scaled viewing-angle quantity, which improves interpretability and potentially reduces sensitivity to changes in focal length, resolution, and field of view [1-4].

DOI: 10.54254/2755-2721/2026.33602

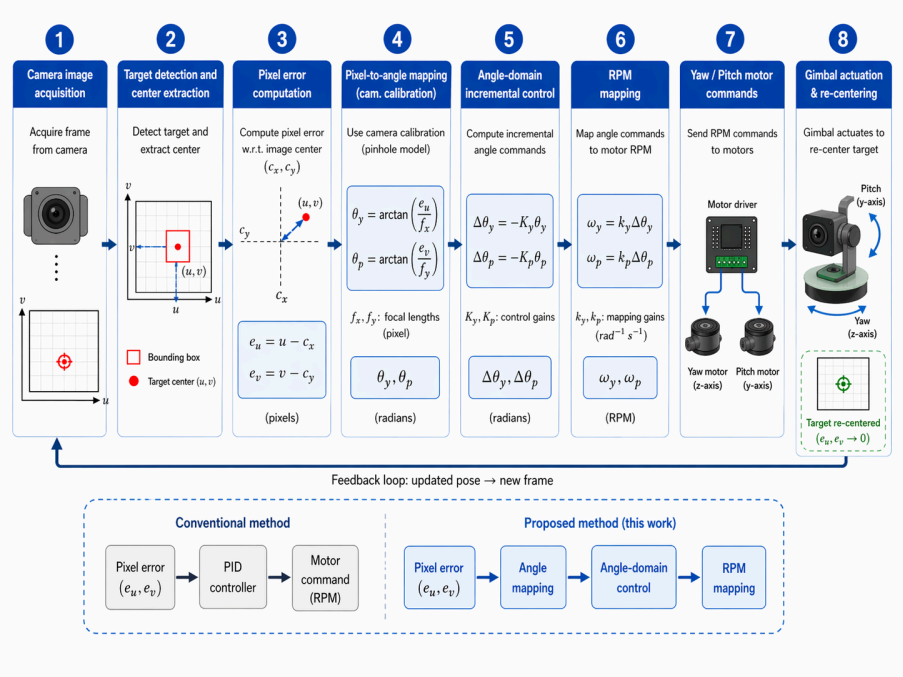

Figure 1. Overall workflow of the proposed vision-guided gimbal tracking system. The image-plane target deviation is converted into angular error before control, and the lower panel compares the proposed angle-domain chain with a conventional pixel-domain pipeline

Figure 2 shows the prototype hardware platform used as the carrier of the proposed method. The system integrates a camera module, a two-axis gimbal actuation unit, an embedded controller, motor-driving electronics, and a mobile base. At the current stage, the main role of this figure is not to claim control performance, but to demonstrate that the proposed method is designed for a real hardware chain rather than being a purely conceptual formulation. This system view is important because recent research on active gimbal mechanisms and gimbal-assisted aerial perception has highlighted the benefits of rapid line-of-sight adjustment, reduced blur, and improved visual persistence when the sensing direction is actively regulated.

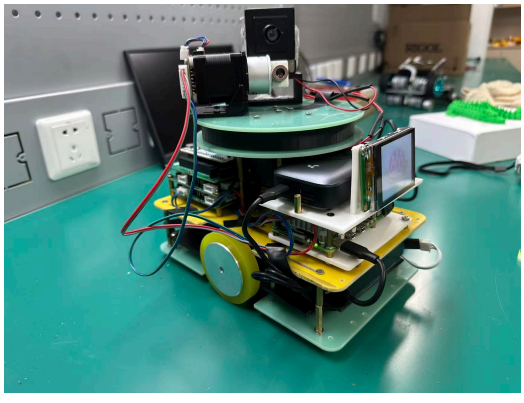

Figure 2. Photograph of the developed vision-guided gimbal tracking platform. The prototype includes the camera, two-axis actuation unit, embedded controller, motor-driving electronics, and mobile carrier

DOI: 10.54254/2755-2721/2026.33602

3. Pixel-to-angle mapping

The proposed method is built on the observation that the controller should regulate the target line of sight rather than the pixel position itself. Figure 3(a) illustrates the imaging geometry. Let the target center in the image plane be denoted by (u, v), and let the principal point be denoted by (c_x, c_y). The horizontal and vertical pixel errors are defined as follows:

$$e_u = u - c_x, \quad e_v = v - c_y. \tag{1}$$

Instead of treating these image-plane quantities as the final control errors, the proposed method interprets them as deviations from the optical axis. Using a pinhole camera model with effective focal lengths f_x and f_y, the yaw and pitch angle errors are computed as

$$\Delta\theta_{yaw} = \arctan(e_u / f_x) \tag{2}$$

$$\Delta\theta_{pitch} = \arctan(e_v / f_y) \tag{3}$$

This formulation allows the controller to operate on a physically interpretable quantity rather than on raw image displacement.

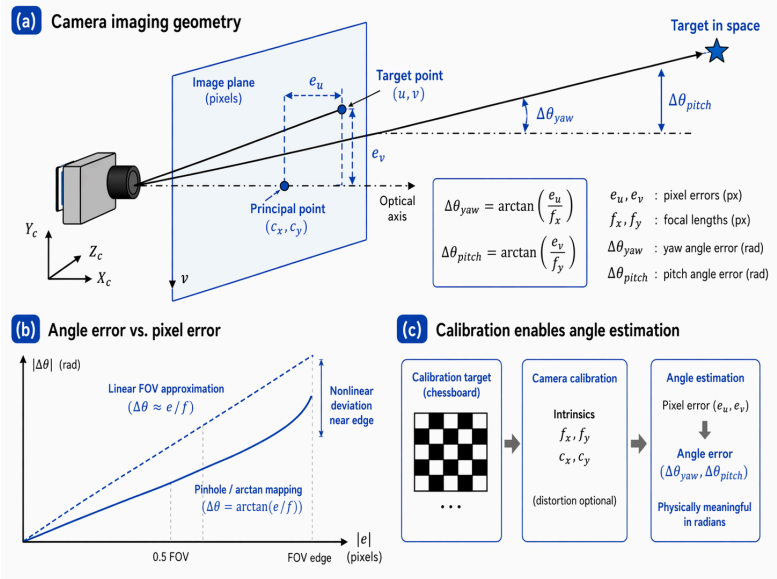

Figure 3. Pixel-to-angle mapping. (a) Imaging geometry and the definition of pixel and angular errors. (b) Comparison between a linear FOV approximation and nonlinear arctangent mapping. (c) Role of camera calibration in obtaining physically meaningful angular error estimates

Figure 3(b) further highlights the difference between a small-angle linear approximation and the nonlinear arctangent relation. When the target is close to the image center, a first-order approximation may be adequate. However, as the target moves toward the image edge, the true mapping becomes increasingly nonlinear. Figure 3(c) places camera calibration within the overall method. Calibration is not merely a preprocessing convenience; it is the prerequisite that turns image-plane deviation into a physically meaningful control input [1].

DOI: 10.54254/2755-2721/2026.33602

4. Angle-domain incremental control and stabilization

Figure 4(a) shows the proposed control chain. Once the angular error Δθ is obtained, the controller does not rely on a full pixel-domain PID structure as the main path. Instead, it uses a proportional incremental law

$$\omega = k \cdot \Delta\theta. \tag{4}$$

Here ω denotes the commanded angular velocity for the corresponding axis, and k is the angle-domain gain. The command is then mapped into motor speed using

$$\mathrm{RPM} = 60\omega / (2\pi). \tag{5}$$

The advantage of this chain is that each variable has an explicit physical meaning: image deviation corresponds to pixel error, line-of-sight deviation corresponds to angular error, and the actuator input is expressed as motor RPM. This makes the controller more interpretable and easier to extend with encoder or IMU feedback in future versions [2]- [7].

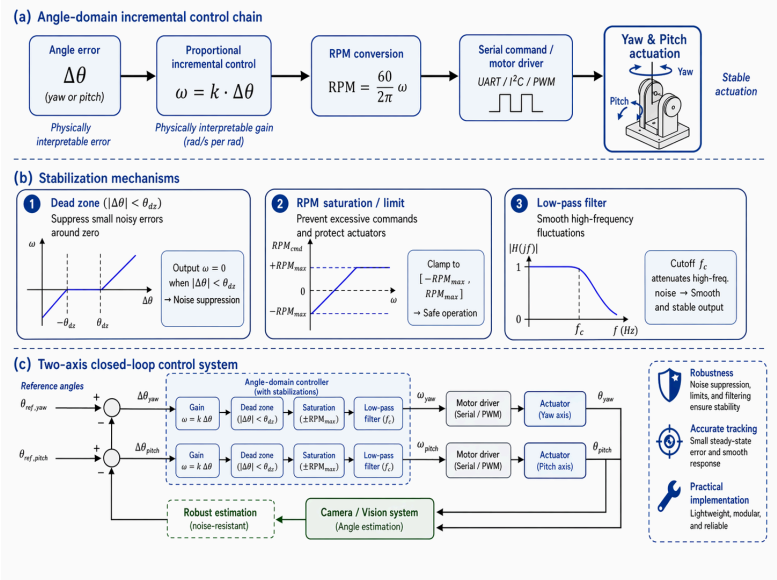

Figure 4. Angle-domain incremental control and stabilization mechanisms. (a) Control chain from angular error to motor command. (b) Dead zone, saturation, and low-pass filtering. (c) Two-axis closed-loop architecture for the proposed visual gimbal control system

Figure 4(b) describes three stabilization mechanisms used to improve deployability. First, a dead zone suppresses small commands caused by pixel jitter, quantization error, or detection noise when the target is already close to the center. Second, RPM saturation prevents excessive commands and protects the actuator from abrupt speed demands. Third, low-pass filtering attenuates high-frequency fluctuations from the visual front end and produces smoother command signals. Similar stabilization considerations are common in recent work on robust visual servoing, gimbal-assisted tracking, and UAV landing with field-of-view constraints.

Figure 4(c) presents the complete two-axis closed-loop structure. The yaw and pitch channels share a common visual front end but maintain independent angle-error computation and command generation paths. The controller acts on the line of sight by adjusting gimbal pose, and the updated

DOI: 10.54254/2755-2721/2026.33602

image is fed back to the camera for the next iteration. This modular structure is particularly suitable for future integration with encoder feedback, IMU measurements, or more sophisticated state estimation blocks.

5. Experimental protocol and evaluation design

At the current stage, the manuscript is positioned as a method-centric conference draft. Accordingly, Figure 5 is defined as an evaluation protocol rather than a result figure. The purpose of this section is to specify the scenarios and metrics that will be used for future quantitative validation.

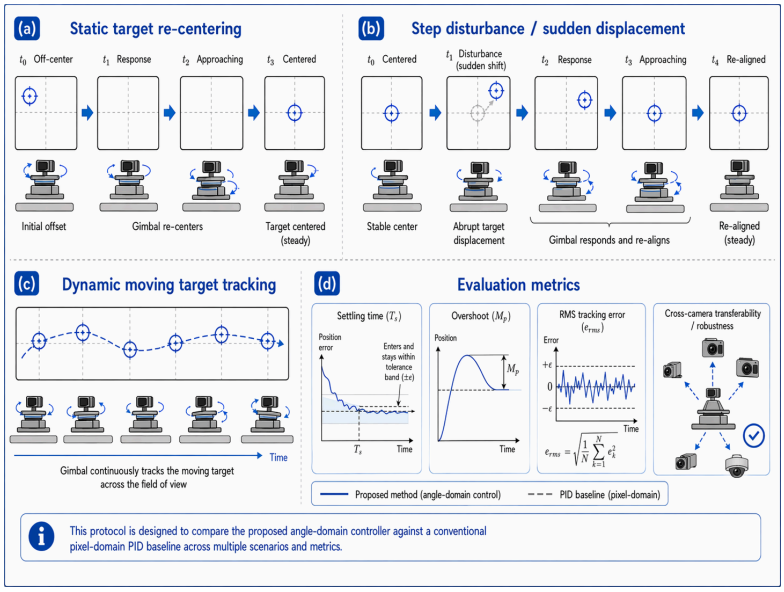

Figure 5. Experimental scenarios and evaluation protocol. (a) Static target recentering. (b) Step disturbance or sudden target displacement. (c) Dynamic moving-target tracking. (d) Evaluation metrics including settling time, overshoot, RMS tracking error, and cross-camera transferability

Figure 5(a) defines a static recentering scenario in which the target starts away from the image center and the gimbal gradually drives it back to the principal point. This test focuses on basic closed-loop convergence and verifies whether the proposed mapping-and-control interface is functioning correctly.

Figure 5(b) defines a step-disturbance scenario in which the target abruptly shifts at a certain instant. This case evaluates transient behavior, including overshoot, recovery speed, and the role of stabilization mechanisms under sudden error changes.

Figure 5(c) defines a dynamic moving-target scenario. In this case, the target follows a time-varying trajectory across the field of view and the gimbal must continuously compensate in yaw and pitch. This setting better reflects realistic operation because it simultaneously tests front-end measurement stability, the real-time behavior of pixel-to-angle conversion, and the response of the controller to continuous visual error inputs.

Figure 5(d) defines the recommended evaluation metrics: settling time T_s, overshoot M_p, RMS tracking error e_rms, and cross-camera transferability or robustness. The last metric is particularly important for this work because the central motivation of the angle-domain formulation is to reduce the dependence of controller tuning on a specific camera configuration. If the control variable is expressed in physically meaningful angular units, similar behavior should be easier to preserve across different focal lengths, resolutions, and fields of view.

DOI: 10.54254/2755-2721/2026.33602

6. Conclusion

This paper presented a vision-guided gimbal target tracking method based on pixel-to-angle mapping and angle-domain incremental control. By explicitly inserting imaging geometry into the control path, the method converts image-plane target deviation into yaw and pitch angular errors before command generation. The resulting chain provides a more physically interpretable interface between visual measurement and gimbal actuation than conventional pixel-domain control. In addition, the use of dead-zone logic, saturation, and low-pass filtering improves the practical structure of the controller for noisy and disturbance-prone visual environments. The current manuscript establishes a coherent conference-style formulation of the method, while future work will focus on real calibration results, motor-speed calibration, and quantitative comparison against a conventional pixel-domain baseline.

References

[1] Z. Zhang, 'A flexible new technique for camera calibration, ' IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 11, pp. 1330-1334, 2000.
[2] S. Hutchinson, G. D. Hager, and P. I. Corke, 'A tutorial on visual servo control, ' IEEE Transactions on Robotics and Automation, vol. 12, no. 5, pp. 651-670, 1996.
[3] F. Chaumette and S. Hutchinson, 'Visual servo control, Part I: Basic approaches, ' IEEE Robotics & Automation Magazine, vol. 13, no. 4, pp. 82-90, 2006.
[4] F. Chaumette and S. Hutchinson, 'Visual servo control, Part II: Advanced approaches, ' IEEE Robotics & Automation Magazine, vol. 14, no. 1, pp. 109-118, 2007.
[5] J. Yang, X. Liu, J. Sun, and S. Li, 'Sampled-data robust visual servoing control for moving target tracking of an inertially stabilized platform with a measurement delay, ' Automatica, vol. 137, art. no. 110105, 2022.
[6] Z. Zhao, P. Han, Y. Xu, W. Xie, W. Zhang, K. Liang, and Q. Zeng, 'Vision-based autonomous landing control of a multi-rotor aerial vehicle on a moving platform with experimental validations, ' IFAC-PapersOnLine, vol. 55, no. 3, pp. 1-6, 2022.
[7] T. Huynh and Y.-B. Kim, 'A study on robust finite-time visual servoing with a gyro-stabilized surveillance system, ' Actuators, vol. 13, no. 3, art. no. 82, 2024.
[8] J. G. Hansen and R. M. H. P. de Figueiredo, 'Active object detection and tracking using gimbal mechanisms for autonomous drone applications, ' Drones, vol. 8, no. 2, art. no. 55, 2024.
[9] S. Yoo, J.-H. Park, and D. E. Chang, 'A two-step controller for vision-based autonomous landing of a multirotor with a gimbal camera, ' Drones, vol. 8, no. 8, art. no. 389, 2024.
[10] J. Kim, Y. Kim, S. Kim, H. Cho, and D. Jung, 'Vision-based geolocation of moving ground targets using Kalman filtering with a gimbal camera on board a UAV, ' Aerospace, vol. 12, no. 12, art. no. 1065, 2025.
[11] W. Mao, Z. Li, and M. Huo, 'Landing control algorithm for gimbal-serviced UAVs based on field-of-view constraints, ' Scientific Reports, vol. 15, art. no. 36013, 2025.
[12] N. R. Chacón Ríos, S. Mondal, and A. Tsourdos, 'Gimbal control of an airborne pan-tilt-zoom camera for visual search, ' in AIAA SciTech Forum, 2025.
[13] P. Łuczak and G. Granosik, 'Autonomous UAV landing and collision avoidance system for unknown terrain utilizing depth camera with actively actuated gimbal, ' Sensors, vol. 25, no. 19, art. no. 6165, 2025.
[14] S. O. Semerikov, P. P. Nechypurenko, T. A. Vakaliuk, I. S. Mintii, et al., 'Vision-Based Autonomous UAV Landing: A Comprehensive Review of Technologies, Techniques, and Applications, ' Journal of Intelligent & Robotic Systems, vol. 111, art. no. 115, 2025.
[15] Y. Zhang, P. Sun, Y. Jiang, D. Yu, F. Weng, Z. Yuan, P. Luo, W. Liu, and X. Wang, 'ByteTrack: Multi-object tracking by associating every detection box, ' in Proc. ECCV, 2022.
[16] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, 'YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, ' in Proc. CVPR, 2023.
[17] H. K. Cheng, S. W. Oh, B. Price, A. Schwing, and J.-Y. Lee, 'Tracking Anything with Decoupled Video Segmentation, ' in Proc. ICCV, 2023.
[18] Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, and J. Chen, 'DETRs Beat YOLOs on Real-time Object Detection, ' in Proc. CVPR, 2024.

DOI: 10.54254/2755-2721/2026.33602

[19] Z. Qin, L. Wang, S. Zhou, P. Fu, G. Hua, and W. Tang, 'Towards Generalizable Multi-Object Tracking, ' in Proc. CVPR, 2024.
[20] Y. Xiong, C. Zhou, X. Xiang, L. Wu, C. Zhu, Z. Liu, S. Suri, B. Varadarajan, R. Akula, F. Iandola, R. Krishnamoorthi, B. Soran, and V. Chandra, 'Efficient Track Anything, ' in Proc. ICCV, 2025.