Journal of Beijing Institute of Technology, 2026, Vol. 35, No. 2

Vision-Based Tracking Control for a Gimbal in UAV Strike Scenario

Lihaoqi Zheng, Yichen Tao, Chen Wei✉

(Science and Technology on Aircraft Control Laboratory, School of Automation Science and Electrical Engineering, Beihang University, Beijing 100083, China)

Abstract: In an unmanned aerial vehicle (UAV) strike scenario, vision-based tracking control for a gimbal with a monocular camera is studied in this paper. First, a vision-based localization method is proposed. Image coordinates are converted into geodetic coordinates, and a multi-frame transformation model is established to account for UAV attitude, gimbal angles, and camera parameters. Then, to address the inaccuracy of single-frame observations, a data-driven adaptive covariance Kalman filter (DAC-KF) is introduced to achieve accurate and robust estimation of target positions. After that, a feedforward-proportional velocity gimbal controller based on damped Jacobian inversion is proposed to ensure the camera remains locked on the moving target. Furthermore, a physics-based trajectory model is established to guide the release of unguided projectiles, facilitating accurate strikes on moving targets. Finally, field experiments show that the proposed method effectively locks and strikes moving targets, with a final deviation within one meter.

Keywords: unmanned aerial vehicle (UAV); Kalman filter; gimbal tracking; visual localization; target fusion

1 Introduction

With the rapid advancement of unmanned aerial vehicles (UAVs), their applications in military reconnaissance, disaster monitoring, and logistics transportation have become increasingly widespread [1–3]. While enabling UAVs to perform complex tasks in unstructured environments (e.g., precision guidance, autonomous landing [4,5], and target strike missions [6]), vision-based target localization demonstrates significant potential in real-world scenarios.

Two mainstream strategies have been proposed for UAV vision-based localization: multi-platform cooperative localization [7,8] and single-platform autonomous localization. The multi-platform method involves multiple UAVs observing a target from different perspectives to improve robustness and localization accuracy. However, it relies heavily on high-bandwidth, low-latency communication systems.

In contrast, the single-platform approach features a simpler structure, more flexible deployment, and lower communication requirements, and often utilizes a fixed camera. For example, in [9], the camera is mounted vertically beneath the UAV, and the relative position of the target is estimated by observing the displacement in the image, combined with known flight pose and altitude information. While this method is straightforward, the field of view is limited. In tasks involving dynamic targets or wide-area search, the system is prone to losing the target or interrupting tracking.

To enhance visual perception, researchers have increasingly integrated gimbal systems into

Manuscript received Mar. 28, 2025; revised Jun. 16, 2025; accepted Jul. 21, 2025. The associate editor coordinating the review of this manuscript was Dr. Haibin Duan. This work was supported by the National Natural Science Foundation of China (Nos. U24B20156, T2121003).

✉ Corresponding author. Email: weichen@buaa.edu.cn

DOI: 10.15918/j.jbit1004-0579.2025.046

UAV platforms [10]. Gimbals offer active control of the camera's orientation, allowing for continuous tracking and stabilized imaging. This significantly extends the camera's coverage range and mitigates image shake or target loss caused by UAV maneuvers, making it particularly effective for complex tracking tasks. With precise gimbal control, a single UAV can track wide-angle targets without changing its own heading, thereby enhancing reconnaissance and localization capabilities.

However, the inclusion of a gimbal introduces new challenges. Rapid motion can lead to image distortion, target blurring and system errors, making it difficult for traditional image-based tracking methods to maintain localization accuracy. In [11], a robust control system based on sliding mode control was designed for two-axis gimbals, while [12] proposed a proportional–derivative (PD)-based angular tracking method. However, in our scenario, directly applying angle tracking can easily lead to unstable and shaky imagery.

Furthermore, due to the increased observation chain caused by gimbal motion, filtering techniques are widely used to enhance the stability and robustness of vision-based localization systems. Among these, the Kalman filter (KF) stands out for its strong dynamic modeling and recursive structure, and has been extensively applied in target tracking and state estimation [13–16].

Some approaches apply filtering directly on the image to track the object in pixel coordinates, while others operate in the physical world to track the actual position. Traditional Kalman filters typically assume fixed observation noise covariance, making them ineffective in scenarios where noise distribution changes with motion states. In [17], an adaptive covariance adjustment method based on observation distance and angular information was proposed for mobile platforms, enhancing localization in complex environments. Unfortunately, such methods cannot be directly used for UAV strike missions.

To address these challenges, this paper proposes a novel approach tailored to gimbal-equipped UAV systems: the data-driven adaptive covariance Kalman filter (DAC-KF). By collecting a large amount of flight and image data, we model the relationship between system states and localization errors, enabling dynamic adjustment of noise parameters in the filter to better match real-world observations. This improves both accuracy and robustness, with strong generalization performance. In addition, to mitigate oscillations caused by gimbal motion near kinematic singularities, we design a velocity-to-angular velocity conversion model based on a damped Jacobian matrix, enhancing control-level stability during tracking. The main contributions of this paper are summarized as follows:

1) A DAC-KF is proposed for target fusion. An error model is first trained on the collected dataset to estimate visual positioning errors. The model then estimates the errors of subsequent data and adjusts the covariance matrix of the filter dynamically, which leads to improved visual fusion accuracy.

2) A velocity gimbal controller is developed based on damped Jacobian. By incorporating damping into the Jacobian, it suppresses undesired oscillatory motions during gimbal maneuvers.

3) A comprehensive solution is designed to address the challenges of visual positioning in UAV strike scenarios, and its effectiveness is validated through flight experiments.

2 Problem Formation

In UAV bombing missions, achieving precise strikes on moving target poses a significant challenge. In such scenario, the target's position and velocity are initially unknown to the UAV. To estimate the target state, a gimbal-mounted camera is employed to extend the visual perception

range. Simultaneously, a sandbag payload is suspended beneath the UAV to simulate the bomb. Once the target state is accurately obtained, the UAV releases the payload accordingly, with the objective of minimizing the landing error and ensuring that the payload lands as close as possible to the target.

The study focuses on two key aspects: modeling the UAV bomb trajectory, and solving the visual coordinate transformation and target localization. The primary innovation is proposed in the target localization component.

2.1 Multi-Coordinate System Definition

In this task, the target's position is identified and tracked using a vision sensor mounted on the UAV. Since the camera is installed on a two-axis gimbal, the image coordinates captured by the sensor must be transformed through a series of coordinate frames to obtain the target's position in the global reference frame. This section establishes the complete relationship among these multiple coordinate systems.

Fig. 1 illustrates the coordinate frames used in the flight verification platform: the geodetic coordinate system Σ, the body coordinate system Σ_b and the camera coordinate system Σ_c. The pixel position of the detected target's center is defined as (ρ_x, ρ_y), the camera resolution is $W \times H$, and after calibration, the camera focal length is f.

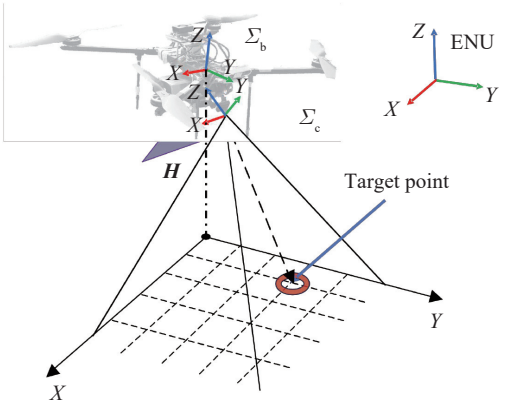

Fig. 1 Schematic diagram of the UAV platform coordinate system

In this study, a yaw-pitch two-axis gimbal is employed, allowing independent control of the two rotational degrees of freedom. As shown in the figure, the gimbal is mounted beneath the UAV, with a lightweight monocular camera fixed at its end. The two rotational joints are actuated by servo motors, which provide real-time angular feedback and allow control via angular velocity commands.

As shown in Fig. 2, the rotation of the first gimbal joint around the Z-axis is defined as yaw γ, while the rotation of the second gimbal joint around the X-axis is defined as pitch θ.

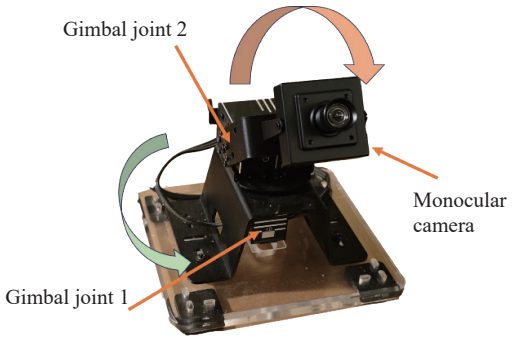

Fig. 2 Schematic diagram of gimbal joint rotations

2.2 UAV Bomb Trajectory Modeling

In ground-attack missions, the UAV typically flies at a constant velocity v_u and releases unguided munitions from the air. Assuming the bomb inherits the UAV's velocity at the moment of release and considering both air resistance and gravity, the bomb's motion in 3D space can be described by the following dynamic model

$$\begin{cases} \dot{x} = v_x \\ \dot{y} = v_y \\ \dot{z} = v_z \\ \dot{v}_x = -kv_x \\ \dot{v}_y = -kv_y \\ \dot{v}_z = -kv_z + g \end{cases} \tag{1}$$

where $p_d = [x, y, z]$ denotes the bomb's position in the geographic coordinate system, and $v_d = [v_x, v_y, v_z]$ is its velocity vector. g represents the gravitational acceleration, and k is the damping coefficient associated with air resistance.

To ensure strike accuracy, the bomb impact point should be as close as possible to the pre-

dicted position of the target at a future time t_f, which is the duration it takes for the bomb to reach the ground (i.e., $z=0$) from the release point. Let the UAV's position at the moment of release be $p_u(t_0)=[x_0,y_0,z_0]$, and assume the bomb's initial velocity $v_d(t_0)$ equals $v_u(t_0)$.

The time t_f can be numerically obtained by integrating the differential Eq. (1) above. During this interval, the target continues to move, so it is essential to accurately predict its future position before release. Assuming the target moves in a straight line at constant velocity, and the visual localization system can provide the current position $p_t(t_0)$ and velocity $v_t(t_0)$, the target' s position at time t_f can be predicted as

$$p_t(t_f)=p_t(t_0)+v_t(t_0)\,t_f \tag{2}$$

To achieve a precise strike, the bomb's predicted impact point $p_d(t_f)$ should be as close as possible to the target's predicted position at t_f. This leads to the optimization objective

$$\min\left(||p_d(t_f)-p_t(t_f)||\right) \tag{3}$$

3 Visual Target Localization and Target Fusion

3.1 Visual Target Localization

To enable effective gimbal tracking control, it is first necessary to localize the target point. After the target is identified and its center is determined through object detection, the relative position of the target can be calculated using the real-time attitude of the UAV and the current angles of the gimbal. This combined information provides the basis for accurate and dynamic target localization.

The relative orientation of the target in the camera coordinate frame Σ_c is characterized by two angular components: ϕ, representing the roll angle about the y-axis, and θ, representing the pitch angle about the x-axis. These angles can be computed through proportional transformation as follows

$$\begin{cases}\phi=-\arctan\left(\dfrac{\rho_x-\dfrac{W}{2}}{f}\right)\\ \theta=-\arctan\left(\dfrac{\rho_y-\dfrac{H}{2}}{f}\right)\end{cases} \tag{4}$$

Assuming the distance from the target to the camera plane is l_{o-c}, the target location in Σ_c is

$$P_o=\begin{bmatrix}\tan(\phi)\,l_{o-c}\\ \tan(\theta)\,l_{o-c}\\ -l_{o-c}\end{bmatrix} \tag{5}$$

Furthermore, the position of the object relative to the UAV in Σ can be expressed as

$$\Delta P=R^b\left(R_b^cP_o+P_\Delta^b\right) \tag{6}$$

where R^b is the attitude rotation matrix from Σ_b, to Σ, R_b^c is the attitude rotation matrix from Σ_c to Σ_b,P_Δ^b is the distance between the gimbal base and the UAV's center of gravity in the body in Σ_b. R_b can be derived using a rotation matrix

$$R_b=\begin{bmatrix}c_2c_1 & -c_2s_1 & s_2\\ s_1 & c_1 & 0\\ -s_2c_l & s_2s_1 & c_2\end{bmatrix} \tag{7}$$

where θ_1 is the yaw angle of the gimbal's first joint, and θ_2 is the pitch angle of the gimbal's second joint. For compactness, the shorthand notation $c_i=\cos(\theta_i)$ and $s_i=\sin(\theta_i)$ is adopted.

Since the approximate altitude of the UAV can be obtained through the flight control system, the position of the object can be further calculated using proportional transformation

$$P=P_b-\frac{H}{\Delta P_Z}\Delta P \tag{8}$$

$$\Delta P=\begin{bmatrix}\Delta P_X\\ \Delta P_Y\\ \Delta P_Z\end{bmatrix} \tag{9}$$

where P_b is the UAV's location in Σ, and P is the calculated landing point location, as shown in Fig. 1. Due to the significant error in altitude measurement, and the fact that the result becomes more accurate as the UAV approaches the target point, continuous alignment opera-

tions are performed in target localization.

3.2 Data-Driven Adaptive Covariance Kalman Filter for Target Fusion

Due to hardware limitations, false detections by recognition algorithms, and pose jitter induced by rapid gimbal movement, angle estimation may exhibit deviations even when gimbal angle data is acquired at high frequencies. This in turn leads to inaccuracies in target coordinate transformation, ultimately causing tracking jitter during the control process. Consequently, target fusion of the raw target positioning data is necessary.

The Kalman filter is a recursive estimation algorithm based on linear system state-space modeling and the assumption of Gaussian noise. Its operation can be divided into two main stages: prediction and update.

In the prediction phase, the discrete-time state transition model of the target system is first established

$$\hat{x}_{k|k-1} = A_k \hat{x}_{k-1|k-1} + B_k u_k \tag{10}$$

where $\hat{x}_{k|k-1}$ is the prior estimate of the target state at time k, $\hat{x}_{k-1|k-1}$ is the posterior estimate at time $k-1$, A_k is the state transition matrix, B_k is the control input matrix and u_k is the control input at time k. The corresponding prediction of the state covariance is given by

$$P_{k|k-1} = A_k P_{k-1|k-1} A_k^{\mathrm{T}} + Q_k \tag{11}$$

where $P_{k|k-1}$ is the predicted state covariance at time k, representing uncertainty in the state estimate, Q_k is the process noise covariance matrix, which captures internal uncertainties and modeling inaccuracies.

During the update phase, the prior estimate is fused with the current measurement using Gaussian-weighted fusion

$$\begin{cases} K_k = P_{k|k-1} H_k{}^{\mathrm{T}} (H_k P_{k|k-1} H_k{}^{\mathrm{T}} + R_k)^{-1} \\ \hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k (z_k - H_k \hat{x}_{k|k-1}) \\ P_{k|k} = (I - K_k H_k) P_{k|k-1} \end{cases} \tag{12}$$

where K_k is the Kalman gain, balancing the trust between prediction and observation, z_k is the current measurement, H_k is the observation matrix, R_k is the measurement noise covariance matrix.

In target localization scenarios, the motion input of the target is unknown, and ideal assumptions about its availability can lead to significant model mismatches. Therefore, it is necessary to estimate u_k based on previous observations. In this paper, the target recognition frequency is approximately 10 Hz, and the velocity of the moving target does not change drastically. Hence, the target velocity u_k is estimated using historical state values

$$\hat{u}_k = \frac{x_{k-1} - x_{k-n}}{t_{k-1} - t_{k-n}} | t_{k-1} - t_{k-n} < t_{\max} \tag{13}$$

where $t_{max} = 0.5$ represents the maximum acceptable time interval, and Eq. (10) can be rewritten as

$$\hat{x}_{k|k-1} = A_k \hat{x}_{k-1|k-1} + B_k \hat{u}_k \tag{14}$$

In the standard KF, both the process noise covariance Q_k and the measurement noise covariance R_k are stationary. However, in UAV strike scenarios, these noise covariances vary with the state of the platform. To address this limitation, this paper proposes a DAC-KF, based on the KF framework. DAC-KF dynamically adjusts the KF' s observation noise covariance matrix. By integrating factors including gimbal angular velocity, UAV attitude perturbation, flight altitude, and target spatial position, and collecting substantial raw data, this method statistically models the relationship between target positioning errors and these variables.

To estimate the observation error in target measurements, the k-th observation error E_k is modeled as a function of multiple system states. Specifically, the angular measurement error is primarily affected by the UAV's pitch angular velocity ω_p and the gimbal's rotational velocity ω_g, while the distance measurement error is mainly influenced by the UAV altitude h and the relative angle α. Additionally, the UAV flight

speed v is associated with GPS positioning bias. Accordingly, the k-th estimated observation error $\hat{E}_k$ is expressed as

$$\hat{E}_k = A_3 \frac{(A_1\omega_p + A_2\omega_g)}{\tan(\alpha)} h + A_4 v + A_5 \tag{15}$$

where A_1, A_2, A_3, A_4 are model parameters to be determined. The meanings of these parameters are summarized in Tab. 1.

Tab. 1 Parameter descriptions

Parameter	Description
A_1	Weighting factor for the influence of UAV pitch angular velocity error
A_2	Weighting factor for the influence of gimbal rotational velocity on angular measurement error
A_3	Weighting factor for the combined amplification effect of angular velocity and altitude on the total error
A_4	Weighting factor for the impact of UAV flight speed

Since the given fitting model is a nonlinear model, it is rewritten into a linear form:

$$\hat{E}_k = A_1' \frac{\omega_p}{\tan(\alpha)} h + A_2' \frac{\omega_g}{\tan(\alpha)} + A_4 v + A_5 \tag{16}$$

$$A_1' = A_1 A_3 \tag{17}$$

$$A_2' = A_2 A_3 \tag{18}$$

To estimate the model parameters A_1 to A_5, the least squares method is employed to minimize the squared difference between the predicted observation error and the measured error. Suppose N sets of measurement samples are available, the least squares objective is to minimize the following cost function

$$A_1 \sim A_5 = \operatorname{argmin} \sum_{k=1}^{N} \left[\hat{E}_k - \mathrm{E}_k\right]^2 \tag{19}$$

Based on this, the new observation noise covariance $R_k{}'$ and the new process noise covariance matrix $Q_k{}'$ can be defined as

$$\begin{cases} R_k' = \lambda_1 \left(1 - \mathrm{e}^{-\alpha \frac{\hat{\mathrm{E}}_k}{\mathrm{E}_{\max}}}\right) + \lambda_2 \hat{E}_k + \varepsilon_1 \\ Q_k' = \dfrac{\lambda_3}{\hat{E}_k + \lambda_4 \sqrt{\hat{E}_k} + \varepsilon_2} \end{cases} \tag{20}$$

where λ is a scaling factor and ε is a small constant ensuring positive definiteness, $E_{\max} = 5$ is the predefined maximum error. Furthermore, Eq. (11) can be rewritten as

$$P_{k|k-1} = A_k P_{k-1|k-1} A_k{}^{\mathrm{T}} + Q_k' \tag{21}$$

$$\begin{cases} K_k = P_{k|k-1} H_k{}^{\mathrm{T}} (H_k P_{k|k-1} H_k{}^{\mathrm{T}} + R_k')^{-1} \\ \hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k (z_k - H_k \hat{x}_{k|k-1}) \\ P_{k|k} = (I - K_k H_k) P_{k|k-1} \end{cases} \tag{22}$$

The pseudocode implementation and operational workflow of the proposed DAC-KF algorithm for target fusion are presented below.

Algorithm 1 DAC-KF

Initialize: Target candidates $C = \varnothing$, maximum distance $D_{\max} = 5$, collected error data $T = \{\text{state}, \text{error}\}$

Use T to calculate A_1 to A_5 using Eq. (19)

for each incoming detection p do

  Find p_{near} in C

  if $\mathrm{H}(p, p_{\text{near}}) \le D_{\max}$ then

    Predict the current state using Eq. (20)

    Get the state: $\omega_p, h, \alpha, \omega_g, v$, Compute R_k using Eq. (16)

    Combine the calculated R_k with the actual observations to correct the prediction using Eq. (22)

    Update p_{near}.

  else

    Add p into C

Return the fused target point

4 Vision-Based Gimbal Control

During gimbal control, a "target tracking plane" is established, which is parallel to the base plane of the gimbal at distance L beneath. The intersection point between the extension line from the center of the gimbal's end and this plane is considered the current position of the gimbal on the "target tracking plane." Similarly, the detected position of the object also intersects this plane and is regarded as the desired position for the gimbal.

The coordinates of the gimbal center on the tracking plane can be expressed as

$$P_{\text{gimbal}} = \begin{bmatrix} -\sin(\gamma)\tan(\theta)\, L \\ \cos(\gamma)\tan(\theta)\, L \end{bmatrix} \tag{23}$$

By computing the Jacobian matrix of this expression, the relationship between the linear velocity of the gimbal center on the tracking plane V_{gimbal} and the gimbal's angular velocity $\mathrm{d}(\theta_{\text{gimbal}})$ can be obtained as follows

$$J = \begin{bmatrix} -\cos(\gamma)\tan(\theta)L & \dfrac{-\sin(\gamma)}{\cos^2(\theta)}L \\ -\sin(\gamma)\tan(\theta)L & \dfrac{\cos(\gamma)}{\cos^2(\theta)}L \end{bmatrix} \tag{24}$$

$$V_{\text{gimbal}} = J\begin{bmatrix}\mathrm{d}\gamma \\ \mathrm{d}\theta\end{bmatrix} = J\mathrm{d}(\theta_{\text{gimbal}}) \tag{25}$$

When calculating the desired velocity of the gimbal V_{cmd} on the tracking plane, a proportional–feedforward controller with velocity limiting is used.

The desired velocity vector is generated by combining a proportional term K_p based on the target position error with the feedforward velocity V_{target} of the target

$$V_{\text{desire}} = K_p(P_{\text{target}} - P_{\text{gimbal}}) + V_{\text{target}} \tag{26}$$

$$V_{cmd} = \text{sat}\left(\frac{V_{\max}}{||V_{\text{desire}}||}\right)V_{\text{desire}} \tag{27}$$

where $V_{\max}$ represents the velocity limit on the tracking plane and sat() denotes the saturation function. Based on V_{cmd}, the angular velocity control input for the gimbal can be calculated using Eqs. (26) (27). To avoid division by zero during the inversion of the Jacobian matrix, a damping factor λ is introduced

$$\omega_{\text{desire}} = J^{\mathrm{T}}(JJ^{\mathrm{T}} + \lambda^2 I)^{-1}V_{\text{cmd}} \tag{28}$$

$$\omega_{\text{cmd}} = \text{sat}\left(\frac{\omega_{\max}}{||\omega_{\text{desire}}||}\right)\omega_{\text{desire}} \tag{29}$$

In Eq. (28), directly computing the pseudoinverse of J may lead to numerical divergence near singular configurations. With a small damping factor λ, the tracking accuracy is slightly reduced, but oscillations around singularities are effectively mitigated. Furthermore, when the system operates far from singularities, it holds that $J^{\mathrm{T}}(JJ^{\mathrm{T}} + \lambda^2 I)^{-1} \approx J^{-1}$, which guarantees convergence.

5 Experimental Results

5.1 Validation of Target Fusion Algorithm

To evaluate the effectiveness of the proposed target fusion algorithm, the following experimental procedure was designed. Firstly, a large number of visual localization samples of a static target were collected to fit an error estimation function. Secondly, based on the fitted results, the proposed method was compared with four other filtering approaches. During the experiments, the parameter configurations were set as shown in Tab. 2.

Tab. 2 Experimental parameter settings

Parameter Name	Range	Unit
UAV altitude	80	m
Camera parameters	640×480	pix
Flight speed	0–5	m/s
Gimbal angular speed	0–0.5	rad/s
Recognition rate	10–20	Hz

These data are used to fit the observation error model introduced in Section 3.2, and the fitting result is shown in Fig. 3.

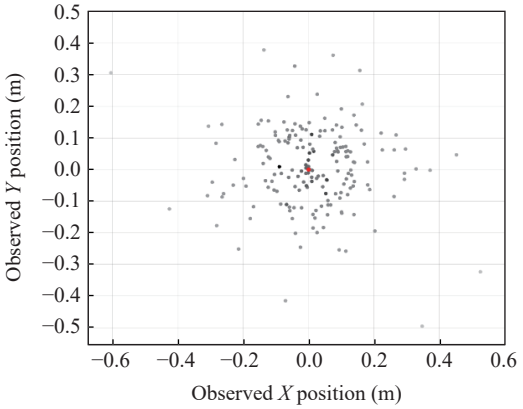

Fig. 3 Fitting results of the vision-detect error

In the figure, the red points represent the true target positions, while the gray points denote the observed target positions. The color gradient of each point corresponds to the magnitude of the estimated error, with lighter shades indicating larger errors. It can be observed that points farther from the origin tend to exhibit lighter colors. The results indicate that the aver-

age difference between the estimated error and the true error is 9.47%.

Once the error fitting model was obtained, the filtering algorithm can be applied for comparative fusion experiments in dynamic conditions.

Fig. 4 illustrates the fusion results of five different fusion filters in one of the motion scenarios. The blue dashed line represents the trajectory provided by an inertial sensor rigidly attached to the target. Due to the limited range of motion, this trajectory is considered the ground truth. The red circles denote raw visual observations from the UAV. The figure also shows the fusion results obtained from five different filters: KF, the proposed DAC-KF, the Extended Kalman Filter (EKF) used in [9], the improved Kalman filter in [17] and adaptive unscented Kalman filter (AUKF) in [16].

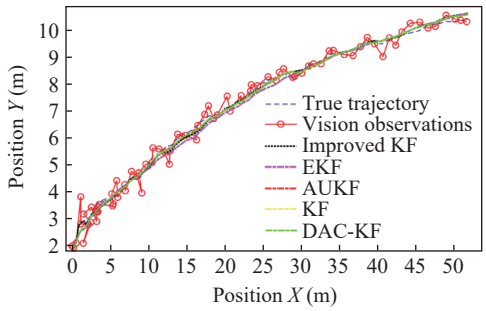

Fig. 4 Comparison of filtering algorithms for moving target fusion

As shown in Tab. 3, across the three comparative experiments, the proposed DAC-KF demonstrates significant advantages in terms of accuracy, robustness, and stability. Compared with EKF and the conventional KF, DAC-KF achieves a substantial reduction in mean error and a notable suppression of maximum error, indicating a stronger adaptability to dynamic and uncertain scenarios. Moreover, the error variance remains consistently low, suggesting that DAC-KF maintains stable estimation performance throughout the entire sequence.

More importantly, when compared with other data-driven filtering methods such as AUKF and improved-KF, DAC-KF achieves a comparable level of accuracy while exhibiting superior performance in maximum error suppression. This characteristic is particularly critical for UAV precision strike missions, where occasional large deviations may lead to severe targeting errors. By dynamically adjusting the observation noise covariance according to system states, DAC-KF effectively balances estimation accuracy and robustness, ensuring reliable localization even under rapid gimbal maneuvers or UAV attitude variations.

5.2 Verification of the Gimbal Tracking Control Algorithm

To evaluate the performance of the gimbal tracking control algorithm, a trajectory tracking experiment was designed. In this experiment, the maximum angular velocity of the gimbal was constrained to 0.5 rad/s, and the control time step was set to 0.05 s to comply with hardware limitations.

Two typical tracking scenarios were implemented to assess the algorithm's effectiveness.

As shown in Fig. 5, the target point starts

Tab. 3 Comparison of localization errors

Scenario	Metrics	Original	KF	AUKF	EKF	Improved-KF	DAC-KF
Test1	Mean error	0.3806	0.1705	0.1580	0.1792	0.1537	0.1552
	Max error	0.9791	0.6415	0.4878	0.4744	0.5224	0.4209
	Var error	0.0465	0.0132	0.0084	0.0084	0.0091	0.0087
Test2	Mean error	0.3437	0.1994	0.2027	0.2103	0.1805	0.1657
	Max error	0.9012	0.4885	0.3688	0.5343	0.4078	0.3688
	Var error	0.0373	0.0107	0.0093	0.0127	0.0096	0.0087
Test3	Mean error	0.3712	0.1845	0.1603	0.2091	0.1493	0.1441
	Max error	0.9827	0.8412	0.5838	0.5250	0.5838	0.5331
	Var error	0.0552	0.0157	0.0092	0.0143	0.01	0.0095

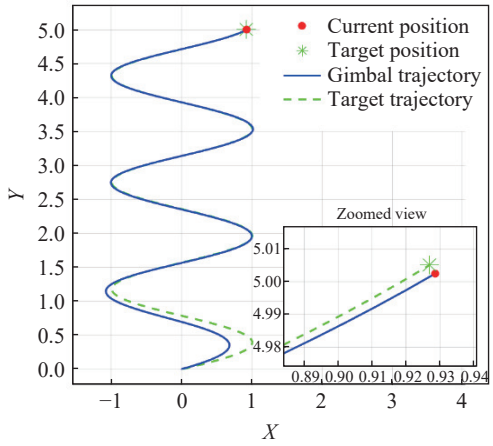

Fig. 5 Gimbal tracking of a curved trajectory

from the origin and moves on the tracking plane with a velocity of $v_x = 2\cos(2\,t)$, $v_y = 0.5$, forming a trajectory similar to a sine wave. The gimbal is tasked with tracking this moving target. The tracking accuracy improves as the target moves away from the origin and no oscillatory behavior is observed near the singularity, further demonstrating the effectiveness of the proposed method.

In the second scenario, the target point moves along a straight line on the tracking plane. As shown in Fig. 6, the tracking remains highly stable during the linear motion, and the control inputs exhibit smooth variations.

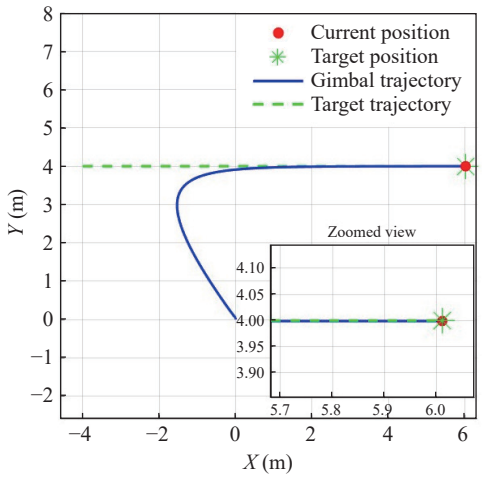

Fig. 6 Gimbal tracking of a linear trajectory

5.3 Outdoor Flight Experiments

To verify the effectiveness of the proposed method, an outdoor flight experiment was carried out. The UAV searched for the target in a designated area. Once the target was detected, the gimbal began tracking, and the UAV performed a strike by approaching the target and releasing a sandbag, aiming to hit as close to the target as possible using visual guidance.

As shown in Fig. 7, the flight verification platform incorporates a flight control system, which is the CUAV V5+ using the open-source APM autopilot framework. The onboard computing platform, the Jetson Orin NX, generates control commands for the UAV and communicates with the flight control system. Its high-performance computing capabilities are essential for running target recognition algorithms and other computationally intensive tasks simultaneously. A two-axis gimbal is installed under the front fuselage to stabilize the camera system, expanding the visual coverage for target point detection and preventing sudden visual loss caused by abrupt attitude changes during flight. At the gimbal's distal end, a distortion-free monocular camera captures real-time imagery for landing point identification using the YOLOv5 algorithm.

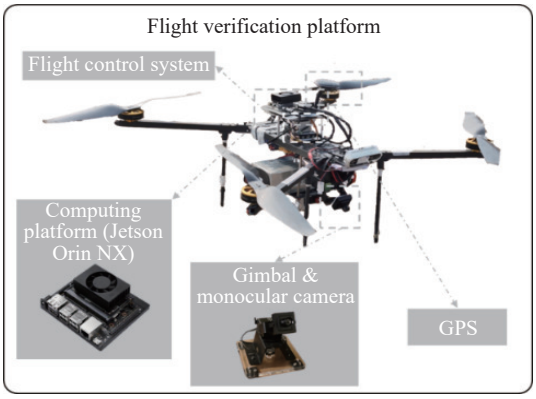

Fig. 7 Components of the flight verification platform

As shown in Fig. 8, the real-flight results for guiding the UAV to strike a moving target verify the effectiveness of the visual localization algorithm proposed in Section 3.1 and the target fusion algorithm proposed in Section 3.2. From the camera view, it can be observed that the target remains stably centered in the camera frame during the experiment, demonstrating the robustness of the gimbal tracking algorithm proposed in Section 4. In the third-person view, the process of

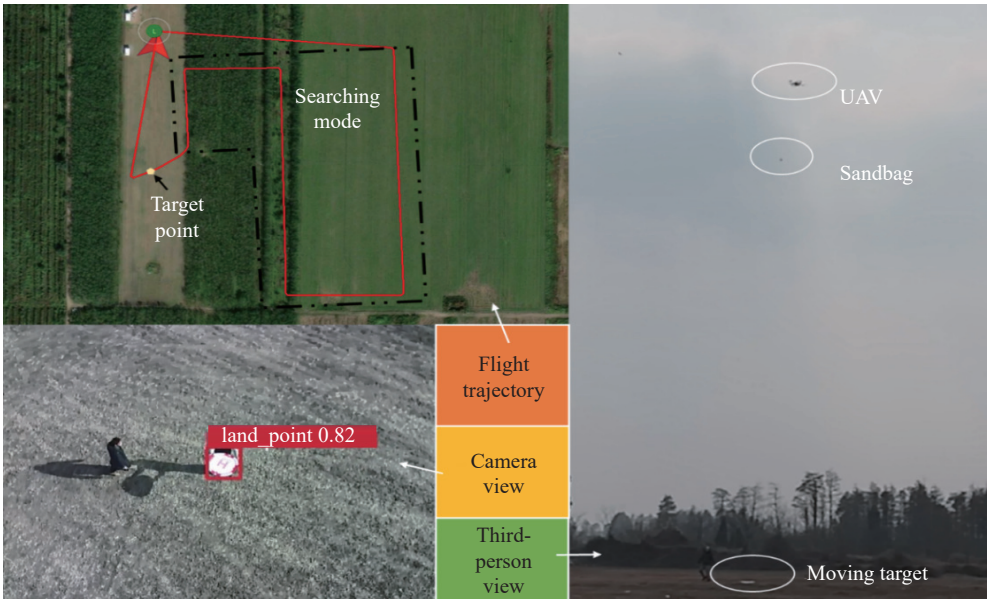

Fig. 8 Outdoor flight experiment

the sandbag falling toward the moving target is captured. The target was uniformly dragged by a staff member, and the UAV flight altitude was set to 20 m. In the process of striking the moving target, the UAV perform tracks the position and velocity of the target, which ensures that when the target moves at a constant velocity, the sandbag remains close to the target, in accordance with Eq. (3) in Section 2.2. The final impact point landed within one meter of the target.

6 Conclusions

This paper presents a vision-based tracking control framework for gimbal-equipped UAV strike systems. The proposed method addresses several practical challenges in real-time moving target engagement, including single-frame detection instability under UAV attitude perturbation, kinematic singularity-induced oscillations in gimbal control, and dynamic error accumulation in geodetic coordinate conversion. To enhance operational robustness, a multi-frame transformation model integrating UAV pose, gimbal angles, and camera parameters is established, while a spatial projection strategy simplifies target tracking through virtual plane mapping.

The core contributions of this work lie in two aspects. First, a DAC-KF is proposed to dynamically adjust the observation noise covariance matrix according to system state features such as gimbal angular velocity and UAV altitude. This improves the accuracy and robustness of target localization, especially in dynamic and uncertain environments. Second, a feedforward-proportional velocity controller based on damped Jacobian inversion is designed to ensure stable and responsive gimbal movement. The proposed method effectively suppresses control oscillations near singular configurations and maintains smooth tracking trajectories.

A series of simulation experiments were conducted to validate the performance of the filtering and gimbal control algorithms independently. Finally, these modules were integrated into a complete UAV strike system and tested through an outdoor flight experiment. The results demonstrate that the proposed framework can accurately track and strike moving ground targets, achieving a final landing error within one meter.

In future work, the proposed system will be extended to support multi-target tracking and engagement, further improving strike precision and enhancing the diversity of UAV strike modes beyond pure tracking.

References:

[1] X. Li, J. Tan, A. Liu, P. Vijayakumar, N. Kumar and M. Alazab,"A novel UAV-enabled data collection scheme for intelligent transportation system through UAV speed control," *in IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 4, pp. 2100-2110, 2021.

[2] P. K. Reddy Maddikunta, S. Hakak, M. Alazab, S. Bhattacharya, T. R. Gadedallu, and W. Z. Khan,"Unmanned aerial vehicles in smart agriculture: Applications, requirements, and challenges," *IEEE Sensors Journal*, vol. 21, no. 16, pp. 17608-17619, 2021.

[3] Y. Xu, Z. Yang, and H. Song,"Current situation and development trend of battlefield target information acquisition and damage assessment technology based on unmanned aerial vehicles," *Transactions of Beijing Institute of Technology*, vol. 45, no. 7, pp. 657-671, 2025.

[4] E. Chatzikalymnios and K. Moustakas, "Autonomous vision-based landing of UAV's on unstructured terrains, " in *2021 IEEE International Conference on Autonomous Systems (ICAS)*, Montreal, QC, Canada, pp. 1-5, 2021.

[5] Y. He, Z. Zeng, Z. Li, and T. Deng,"A new vision-based method of autonomous landing for UAVs," *in 2023 9th International Conference on Electrical Engineering*, no. EECR, pp. 1-6, 2023.

[6] D. Y. Wu, S. F. Lo, Y. W. Liu, and W. M. Liu, "Applying small object detection for bombing UAV in battle field environment, " in *2024 International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan)*, Taichung, Taiwan, China, pp. 309-310, 2024.

[7] G. Niu, Q. Cao, and C. S. Chen, "Vision-based target localization with cooperative UAVs towards indoor surveillance, " in *2023 IEEE 98th Vehicular Technology Conference (VTC2023-Fall)*, Hong Kong, pp. 1-6, 2023.

[8] R. Chang, A. Pan, K. Yu, C. Zhou, and Y. Yang,"A dual UAV cooperative positioning system with advanced target detection and localization," *in IEEE Access*, vol. 12, pp. 43235-43244, 2024.

[9] H. Lee, S. Jung and D. H. Shim, "Vision-based UAV landing on the moving vehicle, " in *2016 International Conference on Unmanned Aircraft Systems (ICUAS)*, Arlington, VA, USA, pp. 1-7, 2016.

[10] P. Kumar, S. Sonkar, A. K. Ghosh, and D. Philip, "Real-time vision-based tracking of a moving terrain target from light weight fixed wing UAV using gimbal control, " in *2020 7th International Conference on Control, Decision and Information Technologies (CoDIT)*, Prague, Czech Republic, pp. 154-159, 2020.

[11] 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)*, Saint Petersburg, Russian Federation, pp. 455-458, 2024.

[12] M. H. Ahmad, K. Osman, M. F. M. Zakeri, and S. I. Samsudin, "Mathematical modelling and PID controller design for two DOF gimbal system, " in *2021 IEEE 17th International Colloquium on Signal Processing & Its Applications (CSPA)*, Langkawi, Malaysia, pp. 138-143, 2021.

[13] T. Wu and X. Fan,"A novel gesture recognition model under sports scenarios based on Kalman filtering and YOLOv5 algorithm," *in IEEE Access*, vol. 12, pp. 64886-64896, 2024.

[14] S. Kanakaraj, M. S. Nair and S. Kalady, "Adaptive Importance Sampling Unscented Kalman Filter With Kernel Regression for SAR Image Super-Resolution, " in *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1-5, 2022.

[15] K. L. Sowmya and P. P, "Real Time Recognition and Monitoring of Moving Targets in Video using Kalman Filter, " *2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, Bangalore, India, pp. 1587-1590, 2018.

[16] A. Dutta and M. Das, "ECG signal denoising using adaptive unscented Kalman filter, " in *2022 IEEE Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, Gwalior, India, 2022.

[17] X. L. Wang, X. Y. Deng, and X. T. Guo,"Research on mobile robot target recognition and location based by improved Kalman filter," *Machine Tool& Hydraulics*, vol. 52, no. 16, pp. 26-31, 2024.

Lihaoqi Zheng received the B.S. degree in automation science in 2023 from the School of Automation Science and Electrical Engineering, Beihang University, Beijing, China, where he is currently working toward the master's degree in automation science. His research inter-

ests include aerial manipulator autonomous control and bioinspired control.

Yichen Tao received the B.S. degree in automation science in 2025 from the School of Software, Beihang University, Beijing, China, where he is currently working toward the master's degree in automation science. His research interests include multi-UAV control and bioinspired control.

Chen Wei received the Ph. D. degree in 1997 from Institute of Systems Science, Chinese Academy of Sciences, Beijing, China, She is currently an associate professor of Beihang University. Her research interests include multi-UAV cooperation and swarm intelligence.