

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

Threat-Aware UAV Dodging of Human-Thrown Projectiles with an RGB-D Camera

Yuying Zhang¹, Na Fan², Haowen Zheng¹, Junning Liang¹, Zongliang Pan³, Qifeng Chen², and Ximin Lyu^{1,4}

Abstract—Uncrewed aerial vehicles (UAVs) performing tasks such as transportation and aerial photography are vulnerable to intentional projectile attacks from humans. Dodging such a sudden and fast projectile poses a significant challenge for UAVs, requiring ultra-low latency responses and agile maneuvers. Drawing inspiration from baseball, in which pitchers’ body movements are analyzed to predict the ball’s trajectory, we propose a novel real-time dodging system that leverages an RGB-D camera. Our approach integrates human pose estimation with depth information to predict the attacker’s motion trajectory and the subsequent projectile trajectory. Additionally, we introduce an uncertainty-aware dodging strategy to enable the UAV to dodge incoming projectiles efficiently. Our perception system achieves high prediction accuracy and outperforms the baseline in effective distance and latency. The dodging strategy addresses temporal and spatial uncertainties to ensure UAV safety. Extensive real-world experiments demonstrate the framework’s reliable dodging capabilities against sudden attacks and its outstanding robustness across diverse scenarios.

Index Terms— Aerial systems; Perception and Autonomy, UAV safety, Human pose estimation, Projectile dodging

I. INTRODUCTION

THE rapid advancement of uncrewed aerial vehicles (UAVs) and their supporting infrastructure has significantly expanded the UAV market, enabling diverse applications such as aerial imaging, last-mile delivery, and air traffic management [1], [2]. To meet the demands of these complex tasks, modern UAVs are increasingly equipped with autonomous modules for environmental perception, navigation, and obstacle avoidance. Despite these advances, UAVs often fail to cope with sudden human-initiated attacks. Recent reports have documented cases where crowds at public events throw projectiles to disrupt UAV operations [3], [4], posing significant threats to their safety and public security. Conse-

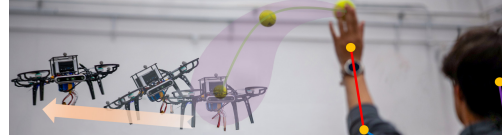


Fig. 1. Upon detecting the human’s intention to attack with a tennis ball, the UAV executes a rapid dodging maneuver to avoid the sudden, fast projectile.

quently, there is an urgent need for robust strategies to counter human-initiated attacks involving fast-moving projectiles.

Developing robust UAV systems capable of rapid responses to sudden human-initiated attacks remains a critical and unresolved research problem. Dodging such projectile threats involves overcoming several challenges: (1) *Perception Latency*: Projectiles often emerge suddenly at close range, leaving a narrow time window for detection and dodging. Therefore, minimizing the delay between sensing and control is crucial while maintaining high prediction accuracy to ensure effective avoidance. (2) *Inaccurately Predictable Threats*: The unpredictable timing of human-initiated attacks, combined with the complex dynamics of thrown projectiles influenced by environmental factors such as wind, presents significant challenges in accurately modeling projectile trajectories. (3) *Computational Constraints and Cost-effectiveness*: UAV hardware must support real-time computing capabilities while remaining cost-effective to facilitate widespread deployment. Furthermore, the integration of affordable, high-performance sensors and processors is crucial for enabling scalable and reliable threat-aware projectile dodging.

To address these challenges, we draw inspiration from baseball, where batters facing high-speed pitches struggle to hit the ball if they react only after the pitcher’s release. Instead, they anticipate the pitch by observing the pitcher’s posture. Motivated by this analogy, we propose an onboard real-time predictive dodging framework using an RGB-D camera. Unlike conventional approaches that track projectiles post-launch [5], our strategy predicts trajectories by analyzing human poses, which enables early anticipation even before the projectile becomes visible or is released. This capability significantly enhances early motion planning. Moreover, since human skeletal features are substantially larger than the projectile, our method supports longer-range predictions. To mitigate uncertainties in attack timing and trajectory, we introduce an ivory-shaped uncertainty model that expands the hazard zone of the projectile, allowing the UAV to execute dodging maneuvers under hazardous conditions. By leveraging a large-scale pretrained human pose estimation (HPE) model [6], our approach achieves robust generalization to diverse attackers and projectiles without the need for retraining. Extensive real-world experiments demonstrate that our method, when deployed on hardware with limited computational resources,

Manuscript received: August 8, 2025; Revised September 26, 2025; Accepted November 20, 2025. This paper was recommended for publication by Editor G. Loianno upon evaluation of the Associate Editor and Reviewers’ comments. This work is supported by the National Natural Science Foundation of China (Grant No. 62303495), the Young Talent Support Project of Guangzhou Association for Science and Technology (Grant No. QT-2025-004), and the Research Grants Council of HKSAR (Grant No. AoE/E-601/24-N). (Yuying Zhang and Na Fan are co-first authors.) (Corresponding author: Ximin Lyu.) (Project page: <https://threat-aware-dodging.github.io/>)

¹Yuying Zhang, Haowen Zheng, Junning Liang, and Ximin Lyu are with the School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou, China (e-mail: {zhangyy393, zhenghw23, liangjn33}@mail2.sysu.edu.cn, lvxm6@mail.sysu.edu.cn).

²Na Fan and Qifeng Chen are with the Visual Intelligence Lab, Cheng Kar-Shun Robotics Institute, The Hong Kong University of Science and Technology, Hong Kong SAR, China (e-mail: {nfan, cqf}@cse.ust.hk).

³Zongliang Pan is with Shenzhen ePropulsion Technology Limited, Shenzhen, China.

⁴Ximin Lyu is also with Differential Robotics Technology Co., Ltd., Hangzhou, China.

Digital Object Identifier (DOI): see top of this page.

©2026 IEEE

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

achieves significant improvements in detection latency, prediction accuracy, and dodging success rate (SR).

Our contributions can be summarized as follows:

- 1) **Threat-Aware Dodging System:** We propose a novel threat-aware framework for UAVs to dodge human attacks using an RGB-D camera. Our approach predicts projectile trajectories by recognizing human poses and executes dodging maneuvers based on an uncertainty-aware dodging strategy to adjust the UAV trajectory.
 - 1A. **Pose-Aware Projectile Trajectory Prediction (PAPT):** We develop a pose-aware framework for real-time perception and prediction of projectile trajectories, integrating human pose estimation with motion modeling to achieve a 6-meter effective detection range and an average detection latency of 26.4 ms on CPU-only hardware.
 - 1B. **Uncertainty-Aware Dodging Strategy (UAD):** To address the challenges in accurately predicting attack timing and projectile trajectory, we propose an uncertainty-aware dodging strategy. Our approach constructs a joint uncertainty region that encompasses multiple surviving projectile trajectories and incorporates a penalty mechanism to optimize the UAV's current trajectory.
- 2) **Extensively Tested in Real-World Environments:** We conduct extensive real-world evaluations under diverse conditions, including varying speeds, launch angles, distances, throwers, lighting and occlusion settings, projectile types, and multiple simultaneous threats, providing compelling evidence of the framework's effectiveness and robustness.

II. RELATED WORK

Thrown projectiles represent a type of sudden and fast dynamic obstacle. Protecting UAVs from human-initiated attacks involving such projectiles requires two essential capabilities: early detection of human threats and rapid dodging of the projectiles. Early dynamic obstacle avoidance methods, designed primarily for slower-moving pedestrians [7], [8], [9], [10], [11], are inadequate for such intentional attacks due to the critical demands of low-latency perception and robust generalization. Consequently, recent research has shifted focus to dodging sudden, fast projectiles, often leveraging advanced sensing and intention prediction techniques.

A. UAV Dodging of Sudden Fast Projectiles

There is a growing interest in UAVs dodging the sudden, fast projectile, where prediction-response latency becomes critical. This research trend emerges as event cameras gain popularity, with many approaches utilizing their low-latency, asynchronous detection of per-pixel brightness changes over fixed-rate capture. Mueggler et al. [12] pioneered the use of probabilistic models and extended Kalman filters for tracking the projectile using event-based stereo vision. However, their analysis was primarily limited to drone hovering under ideal conditions. Subsequent studies, such as [1], [13], [14] and [15], addressed challenges like noise and motion

compensation in event cameras during dynamic motion. He et al. [16] proposed an asynchronous fusion of event and depth camera data, which improves the accuracy of 3D trajectory prediction for moving objects. Nevertheless, event cameras are typically bulky, noisy, expensive, and sensitive only to moving objects, which limits their ability to perceive environments fully and restricts their practical applications. Consequently, researchers have explored alternative solutions. Kong et al. [17] demonstrated a method that utilizes 3D LiDAR sensors combined with onboard processing to achieve high-frequency, safe navigation in cluttered and unknown environments, effectively avoiding small dynamic obstacles. Despite these strengths, the approach fundamentally treats dynamic obstacles as static entities, and LiDAR sensors are also heavy and expensive. Lu et al. [5] adopted an RGB-D camera with YOLO-FastestV2 [18] for ball detection. However, their approach requires training on 2000 tennis ball images and fixed attack distances due to camera resolution limitations, hindering its generalization to new scenarios. Recent work [19] uses a transformer architecture to predict object motion from multi-frame LiDAR sequences with high 3D tracking accuracy. However, it is not developed for UAV dodging because it requires high computational resources, making real-time onboard perception infeasible. Therefore, our work aims to develop a solution that enables long-range detection, exhibits strong generalization capabilities, and effectively avoids the projectile using low-cost devices.

B. Human Pose Estimation in Intention-Aware Robotics

With recent advancements in HPE, including enhanced accuracy, multi-person tracking, higher frame rates, and improved generalization [20], [21], [22], pretrained HPE models are increasingly integrated into downstream robotics applications. As a result, HPE has attracted considerable attention for improving autonomous navigation in scenes involving human movement. It provides a vital cue for understanding motion and predicting intentions, particularly in improving SLAM in dynamic scenes [23], forecasting 2D human trajectories in crowded settings using a transform-based framework [24], avoiding pedestrian dynamics [25], and enabling safe human tracking with UAV [26]. Although these methods primarily focus on slow-moving pedestrian scenarios, the idea of integrating human intent into robotic systems has inspired our work. Meanwhile, [27] learns human motion representations and enables semantic-level intent understanding, but it cannot operate in real time. Building on this, we propose leveraging HPE to enable UAVs to timely dodge human aggressive actions, enhancing the robustness against sudden fast projectiles.

III. METHODOLOGY

This work presents a novel framework that enables the UAV to dodge intentional human-initiated attacks. The proposed approach comprises two core modules: PAPT and UAD.

A. Pose-Aware Projectile Trajectory Prediction (PAPT)

The PAPT framework is a real-time system that integrates HPE with projectile motion modeling to track human joint movements and predict the trajectory of the thrown projectile.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

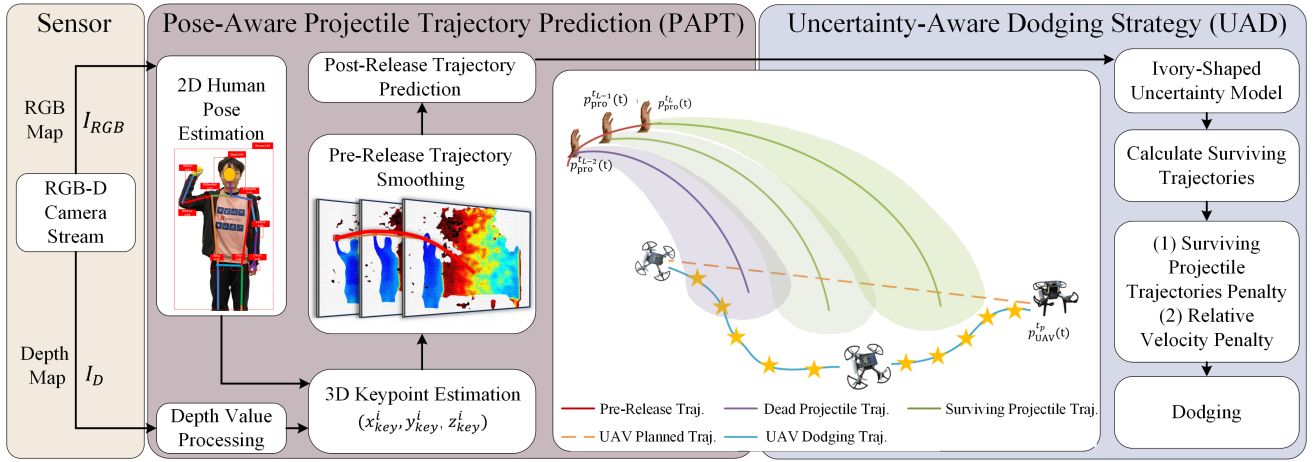


Fig. 2. **System overview.** The system architecture for UAV dodging of human-initiated projectile attacks consists of two main modules: PAPT and UAD. PAPT takes RGB-D camera streams as input, processes RGB-D camera streams through 2D HPE and depth value processing to track 3D body keypoints associated with projectile throwing, and calculates pre- and post-release trajectory. Using initial projectile trajectory predictions, the UAD employs an ivory-shaped uncertainty model and identifies surviving trajectories and regions for assessing UAV collision risk. It then incorporates penalties from the proximity and relative velocity of surviving projectile trajectories to generate a safe dodging trajectory.

Specifically, it combines modules for 3D Keypoint Estimation and Projectile Trajectory Prediction, enabling the UAV to assess and respond to potential threats effectively.

1) *3D Keypoint Estimation:* PAPT leverages synchronized RGB-depth image pairs (I_{RGB}, I_D) , where I_{RGB} denotes the RGB map and I_D the depth map, to estimate 3D human joint positions in real time to anticipate projectile trajectory. The system first extracts 2D pose positions (u_{key}^i, v_{key}^i) for each critical joint (e.g., wrists) in each frame i from I_{RGB} using RTMPose [6], then maps these positions to the corresponding locations in the depth map I_D to obtain the depth values d_{key}^i . To mitigate noise inherent in depth sensing, spatial and temporal consistency are applied to ensure the reliability of d_{key}^i . Spatially, we compute the average depth within a small window centered at (u_{key}^i, v_{key}^i) after removing outliers, discarding the frame if the result remains unreliable. Temporally, we compare the current frame's depth with the previous valid frame, discarding frames where the difference exceeds a predefined threshold.

Thus, the 3D coordinates $(x_{key}^i, y_{key}^i, z_{key}^i)$ in the world frame are obtained using the camera intrinsic matrix \mathbf{K} and the transformation matrix \mathbf{T}_{WC} from camera to world frame:

$$\begin{bmatrix} x_{key}^i \\ y_{key}^i \\ z_{key}^i \\ 1 \end{bmatrix} = \mathbf{T}_{WC} \cdot \begin{bmatrix} d_{key}^i \cdot \mathbf{K}^{-1} \begin{bmatrix} u_{key}^i \\ v_{key}^i \\ 1 \end{bmatrix} \\ 1 \end{bmatrix} \quad (1)$$

2) *Projectile Trajectory Prediction:* The PAPT framework predicts projectile trajectories by modeling the dynamics of human keypoints in two phases: Pre-Release Trajectory Smoothing and Post-Release Trajectory Prediction.

a) *Pre-Release Trajectory Smoothing:* We utilize cubic spline regression to fit the 3D keypoint positions $(x_{key}^i, y_{key}^i, z_{key}^i)$, thereby obtaining the pre-release trajectory. To reduce computational overhead, only the most recent N valid frames are considered. To avoid zig-zag trajectory, we fit cubic splines independently along each axis: (t_{key}^i, x_{key}^i) , (t_{key}^i, y_{key}^i) , and (t_{key}^i, z_{key}^i) . The smoothing factor is empiri-

cally selected. Here, t^i denotes the timestamp of frame i for $i = n - N + 1, \dots, n$, where n is the current frame number. Consequently, the smoothed keypoint trajectories along each axis are obtained as $x_{key}(t)$, $y_{key}(t)$, and $z_{key}(t)$. The position of the keypoint at time t along the pre-release trajectory is then represented as $\mathbf{p}_{key}(t) = (x_{key}(t), y_{key}(t), z_{key}(t))^T$. The corresponding velocity $\mathbf{v}_{key}(t)$ and acceleration $\mathbf{a}_{key}(t)$ are derived from the spline-interpolated position $\mathbf{p}_{key}(t)$ as its first- and second-order time derivatives, respectively.

b) *Post-Release Trajectory Prediction:* Given the inherent unpredictability of human decision-making, which results in stochastic timing of throwing, a conservative approach is adopted to predict the post-release trajectory. Specifically, we consider any time instant t_L as a potential release time if it satisfies the conditions that the magnitude of the pre-release keypoint acceleration $\|\mathbf{a}_{key}(t_L)\|$ exceeds a predefined threshold θ_a , and directions of the pre-release velocity and acceleration are consistent (verified by checking if the dot product $\mathbf{a}_{key}(t_L) \cdot \mathbf{v}_{key}(t_L)$ is positive). We use the corresponding position $\mathbf{p}_{key}(t_L)$ and velocity $\mathbf{v}_{key}(t_L)$ as initial conditions to predict the projectile's velocity $\mathbf{v}_{pro}^{t_L}(t)$ and position $\mathbf{p}_{pro}^{t_L}(t)$ for a potential release at t_L , modeled as:

$$\begin{aligned} \mathbf{v}_{pro}^{t_L}(t) &= \mathbf{v}_{key}(t_L) + \mathbf{g}t, \\ \mathbf{p}_{pro}^{t_L}(t) &= \mathbf{p}_{key}(t_L) + \mathbf{v}_{key}(t_L)t + \frac{1}{2}\mathbf{g}t^2. \end{aligned} \quad (2)$$

where $\mathbf{g} = [0, 0, -g]^T$ is the gravitational acceleration.

B. Uncertainty-Aware Dodging Strategy (UAD)

Beyond the uncertainty in the release timing, the predicted trajectory also exhibits uncertainty. We propose a UAD module that enables the UAV to dodge the approaching projectile by simultaneously accounting for both factors.

Accurate prediction of projectile trajectory is essential for UAVs to dodge threats in human-initiated attacks. However, estimation errors in initial conditions, nonlinear aerodynamic effects, and environmental noise hinder high-precision prediction using ideal parabolic models [12]. To address this challenge, we introduce an ivory-shaped uncertainty model

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

that augments an ideal parabolic trajectory with a temporally expanding uncertainty envelope.

1) *Ivory-Shaped Uncertainty Model*: Multiple sources of uncertainty lead to cumulative errors in trajectory prediction over time. These errors manifest as an expanding uncertainty region around the predicted trajectory. To model this uncertainty, we represent the possible positions at each time t as a three-dimensional ball with uncertainty radius $R(t)$. Nonlinear aerodynamics contribute a quadratic term αt^2 to the uncertainty of position, while initial velocity and position errors add linear βt and constant γ terms. Thus, the uncertainty radius is modeled as:

$$R(t) = \alpha t^2 + \beta t + \gamma, \quad t \in [0, T_s^{t_L}]. \quad (3)$$

where α , β , and γ are empirically determined parameters, and $T_s^{t_L}$ denotes the survival duration of the post-release trajectory, which is the time from the projectile's release instant to its landing. A post-release trajectory is classified as surviving if $t < t_L + T_s^{t_L}$ and as dead (i.e., non-threatening) otherwise. For a post-release trajectory released at time t_L , the possible position distribution at future times $t \in [0, T_s^{t_L}]$ is given by:

$$\mathcal{P}^{t_L}(t) = \left\{ \mathbf{p} \in \mathbb{R}^3 \mid \|\mathbf{p} - \mathbf{p}_{\text{pro}}^{t_L}(t)\|^2 \leq R^2(t) \right\}. \quad (4)$$

This formulation defines an ivory-shaped uncertainty region, signaling collision risk upon intersection with the UAV's position or planned trajectory.

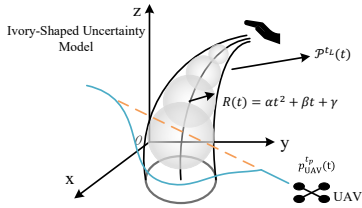


Fig. 3. **Ivory-shaped uncertainty model**: The prediction error in the projectile's trajectory increases over time. Based on this, we construct an ivory-shaped uncertainty region around the predicted trajectory. A collision risk is deemed present when the UAV's current position or planned trajectory intersects with this region.

2) *UAV Trajectory Optimization*: The UAV's trajectory is parameterized using the MINCO polynomial class [28]:

$$\begin{aligned} \mathfrak{T}_{\text{MINCO}} = \{ & \mathbf{p}_{\text{UAV}}^{t_p}(t) : [0, T_u] \mapsto \mathbb{R}^m \mid \mathbf{c} = \mathcal{M}(\mathbf{q}, \mathbf{T}), \\ & \mathbf{q} \in \mathbb{R}^{m(M-1)}, \mathbf{T} = [T_1, \dots, T_M]^T \in \mathbb{R}_{>0}^M \} \end{aligned} \quad (5)$$

where $\mathbf{p}_{\text{UAV}}^{t_p}(t)$ denotes an m -dimensional UAV trajectory planned at time t_p , composed of M quintic polynomial segments, and T_u is the total trajectory duration. The trajectory is compactly parameterized by the intermediate waypoints \mathbf{q} and segment durations \mathbf{T} , while the polynomial coefficients $\mathbf{c} = (\mathbf{c}_1^T, \dots, \mathbf{c}_M^T)^T$ are computed via a linear mapping $\mathcal{M}(\mathbf{q}, \mathbf{T})$. Given this parameterization, any trajectory cost $\mathcal{J}(\mathbf{q}, \mathbf{T}) = \mathcal{F}(\mathbf{c}, \mathbf{T}) = \mathcal{F}(\mathcal{M}(\mathbf{q}, \mathbf{T}), \mathbf{T})$ can be evaluated and optimized over \mathbf{q} and \mathbf{T} , where \mathcal{F} is any user-defined objective. Thus, the gradients $\frac{\partial \mathcal{J}}{\partial \mathbf{q}}$ and $\frac{\partial \mathcal{J}}{\partial \mathbf{T}}$ are propagated from the gradients $\frac{\partial \mathcal{F}}{\partial \mathbf{c}}$ and $\frac{\partial \mathcal{F}}{\partial \mathbf{T}}$ respectively, to optimize the trajectory cost \mathcal{J} . We formulate the trajectory generation as an unconstrained optimization problem. The optimization objective is given by:

$$\begin{aligned} \min_{\mathbf{c}, \mathbf{T}} & (\omega_s \mathcal{J}_s + \omega_o \mathcal{J}_o + \omega_t \mathcal{J}_t + \omega_f \mathcal{J}_f) + \omega_d \mathcal{J}_d + \omega_v \mathcal{J}_v, \\ \omega_s = \omega_o = \omega_f = & 10.0, \quad \omega_t = 0.5, \quad \omega_d = 20.0, \quad \omega_v = 5.8, \quad T = 4.0 \end{aligned} \quad (6)$$

where \mathcal{J}_s , \mathcal{J}_o , \mathcal{J}_t , and \mathcal{J}_f respectively denote the smoothness, static obstacle, time, and feasibility costs. The constraints and gradient computations follow the GCOPTER framework [28]. To handle projectile threats, we introduce two novel penalty terms: \mathcal{J}_d , which discourages the UAV trajectory from approaching predicted projectile trajectories, and \mathcal{J}_v , which promotes velocity directions favorable for dodging. The corresponding weights $\omega_d, \omega_v > 0$ control their relative importance. The proposed cost formulation is general and platform-independent, allowing the same set of weights to be transferred across different UAV platforms with only minor re-tuning for variations in agility and dynamic limits.

a) *Surviving Projectile Trajectories Penalty*: Dodging is triggered if the UAV's position or planned trajectory intersects an ivory-shaped uncertainty region $\mathcal{P}^{t_l}(t)$ associated with a projectile released at t_l . Accurately predicting projectile release timing poses a key challenge in trajectory planning. Relying solely on the current observed trajectory of a projectile can lead to significant limitations. Specifically, planning algorithms may erroneously interpret the current trajectory as the actual trajectory, thus overlooking latent risks embedded in historical trajectory data. Such misjudgments can cause the planner to optimize the trajectory toward regions perceived as collision-free, which, in reality, may coincide with the projectile's actual trajectory, thereby incurring collision hazards. To address this issue, we consider all surviving post-release trajectories at the current planning time t_p , represented as:

$$\mathbf{p}_{\text{pro}}^{t_l}(t), \quad l = 1, 2, \dots, L, \quad t \in [0, T_u]. \quad (7)$$

subject to $t_p > t_l$ and $t_l + T_s^{t_l} > t_p$ for all l , where each such trajectory is a potential candidate for the actual trajectory, and L is the number of surviving projectile trajectories. To address this uncertainty, we penalize the proximity between the UAV's future sampling points and predicted projectile positions:

$$\begin{aligned} \mathcal{J}_d = \sum_{m,k,l} & \left[\max \left(0, R(\tau_{m,k} + \Delta t_{p,l}) + R_s - \|\mathbf{d}_{m,k,l}\| \right) \right]^2, \\ \mathbf{d}_{m,k,l} = & \mathbf{p}_{\text{UAV}}^{t_p}(\tau_{m,k}) - \mathbf{p}_{\text{pro}}^{t_l}(\tau_{m,k} + \Delta t_{p,l}), \Delta t_{p,l} = t_p - t_l, \\ \tau_{m,k} = & \sum_{j=0}^{m-1} T_j + \frac{k}{K} T_m, m = 1, \dots, M, k = 0, \dots, K-1, \end{aligned} \quad (8)$$

where $\sum_{m,k,l} = \sum_{m=1}^M \sum_{k=0}^{K-1} \sum_{l=1}^L$, K represents the number of sampling points on each piece of the UAV's trajectory, R_s denotes the UAV's safety margin radius, and $\Delta t_{p,l}$ represents the time elapsed since the l -th projectile is released. The gradients of \mathcal{J}_d w.r.t. \mathbf{c}_i and T_i can be derived by the chain rule:

$$\begin{aligned} \frac{\partial \mathcal{J}_d}{\partial \mathbf{q}} &= \sum_{m,k,l} \frac{\partial \mathcal{J}_d}{\partial \mathbf{p}_{\text{UAV}}^{t_p}} \frac{\partial \mathbf{p}_{\text{UAV}}^{t_p}}{\partial \mathbf{c}_m}, \\ \frac{\partial \mathcal{J}_d}{\partial T_i} &= \sum_{m,k,l} \frac{\partial \mathcal{J}_d}{\partial \mathbf{p}_{\text{UAV}}^{t_p}} \left(\frac{\partial \mathbf{p}_{\text{UAV}}^{t_p}}{\partial T_i} + \frac{\partial \mathbf{p}_{\text{UAV}}^{t_p}}{\partial \mathbf{c}_m} \frac{\partial \mathbf{c}_m}{\partial T_i} \right) \end{aligned} \quad (9)$$

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

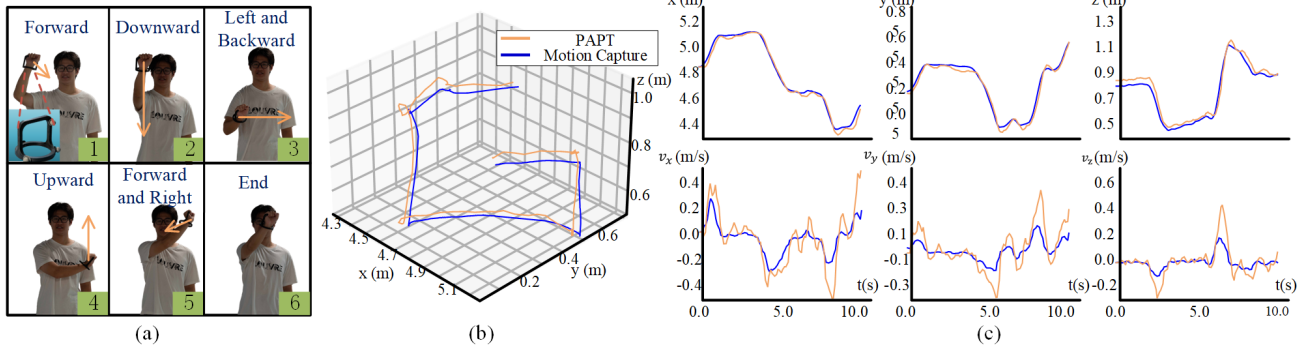


Fig. 4. (a) A participant performs arm motions while wearing a 3D-printed device equipped with mocap markers. (b) Comparison of PAPT-estimated 3D trajectory with ground truth from the mocap. (c) Positional and velocity differences between the estimated and ground truth along the x -, y -, and z -axes.

b) Relative Velocity Penalty: To mitigate collision risk, we penalize the component of the relative velocity between the UAV and projectile onto the relative position vector, encouraging adjustments in speed magnitude or lateral maneuvers.

$$\mathcal{J}_v = \sum_{m,k,l} \left(\frac{\mathbf{v}_{\text{rel},m,k,l} \cdot \mathbf{d}_{m,k,l}}{\|\mathbf{d}_{m,k,l}\| + \epsilon} \right)^2, \quad (10)$$

$$\mathbf{v}_{\text{rel},m,k,l} = \mathbf{v}_{\text{UAV}}^{t_p}(\tau_{m,k}) - \mathbf{v}_{\text{pro}}^{t_l}(\tau_{m,k} + \Delta t_{p,l})$$

where $\mathbf{v}_{\text{pro}}^{t_l}(t)$ and $\mathbf{v}_{\text{UAV}}^{t_p}(t)$ are the velocities at time t of the projectile released at t_l and the UAV whose trajectory is planned at t_p , and $\epsilon > 0$ ensures numerical stability. The gradient computation follows the method outlined in (9):

$$\begin{aligned} \frac{\partial \mathcal{J}_v}{\partial \mathbf{q}} &= \sum_{m,k,l} \left(\frac{\partial \mathcal{J}_v}{\partial \mathbf{p}_{\text{UAV}}^{t_p}} + \frac{\partial \mathcal{J}_v}{\partial \mathbf{v}_{\text{UAV}}^{t_p}} \right) \frac{\partial \mathbf{p}_{\text{UAV}}^{t_p}}{\partial \mathbf{c}_m}, \\ \frac{\partial \mathcal{J}_v}{\partial T_i} &= \sum_{m,k,l} \left(\frac{\partial \mathcal{J}_v}{\partial \mathbf{p}_{\text{UAV}}^{t_p}} + \frac{\partial \mathcal{J}_v}{\partial \mathbf{v}_{\text{UAV}}^{t_p}} \right) \left(\frac{\partial \mathbf{p}_{\text{UAV}}^{t_p}}{\partial T_i} + \frac{\partial \mathbf{p}_{\text{UAV}}^{t_p}}{\partial \mathbf{c}_m} \frac{\partial \mathbf{c}_m}{\partial T_i} \right) \end{aligned} \quad (11)$$

IV. EXPERIMENT

A. Hardware Platform

We evaluate our method on a 1.35 kg custom quadrotor (240×240×105 mm) integrating an Intel RealSense D455 RGB-D camera, a Minisforum EM780 (AMD Ryzen 7 7840U CPU) computer, and an NxtPX4v2 flight controller.

B. Evaluation of the PAPT Module

1) 3D Pose Keypoint Tracking Accuracy: To evaluate the accuracy of our PAPT algorithm, we compare its keypoint position estimates against ground truth (GT) data obtained from a Nokov motion capture (mocap) system. A human participant performs wrist movements with a reflective marker rigidly attached via a custom 3D-printed fixture, which serves as the GT reference (Fig. 4 (a)). The subject executes unconstrained wrist motions, and we quantify the tracking errors between PAPT estimates and the mocap GT.

The 3D keypoint tracking performance of PAPT closely aligns with mocap GT across all spatial dimensions. For position estimates, PAPT achieves a root mean squared error (RMSE) of 0.037 m and a maximum error (MaxE) of 0.0876 m. For velocity estimates, it yields an RMSE of 0.154 m/s and a MaxE of 0.5257 m/s. Specifically, the mean positional errors along the x - and y -axes achieve sub-centimeter accuracy ($< 0.01\text{m}$), while the z -axis exhibits a slightly higher bias

of approximately 0.021 m. Mean velocity errors are slightly elevated at instances of abrupt motion changes, approximately 0.192 m/s. Despite these deviations, errors remain within acceptable limits for real-time applications.

2) Detection Distance and Time Consumption: To evaluate the PAPT's perception capability, we reproduce the algorithm from [5] on our platform and conduct 20 trials. Two key indicators are statistically assessed: (1) the maximum effective detection distance, defined as the farthest range at which the 3D position of the target can be perceived (human keypoints in our case, vs. tennis ball in [5]), and (2) the perception latency, defined as the time from target detection to its trajectory generation. The results are summarized in Table I.

TABLE I
COMPARISON OF DETECTION DISTANCE AND LATENCY

	Lu et al. [5]	Ours
Effective Distance (m)	4.2 ¹ / 3.9*	6.0¹
Average Time Cost (ms)	320 ¹ / 29*	26.4¹

¹ CPU results reproduced on our platform; * GPU results reported in [5].

Our method achieves an effective detection distance of 6 m and an average latency of 26.4 ms, outperforming the approach in [5] in terms of detection range and responsiveness. Unlike direct object detection, PAPT infers projectile positions by recognizing the entire human skeletal structure, thereby eliminating the need for the projectile to be visible or to meet a minimum size threshold. This approach significantly extends the effective detection range. Our algorithm accurately identifies the 2D positions of keypoints at 6 m, a limit imposed by the hardware constraints of the camera's depth sensing capabilities, suggesting that 6 m is not the true upper bound of our method's capability. Notably, our PAPT algorithm is implemented on a CPU platform, achieving an average latency lower than that of [5], which fails to enable real-time detection when tested on a CPU. This highlights the versatility of our method, enabling effective detection on lower-cost platforms.

3) Evaluation under diverse lighting: The PAPT algorithm is further tested under various lighting conditions, including dim-light, planar illumination, and different spotlight settings. Across all conditions, the wrist position RMSE remains below 0.09 m, indicating stable perception performance. Only under extremely bright illumination (40W spotlight) do short-term detection failures occur due to overexposure and the resulting loss of visual features, after which the tracking quickly recov-

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

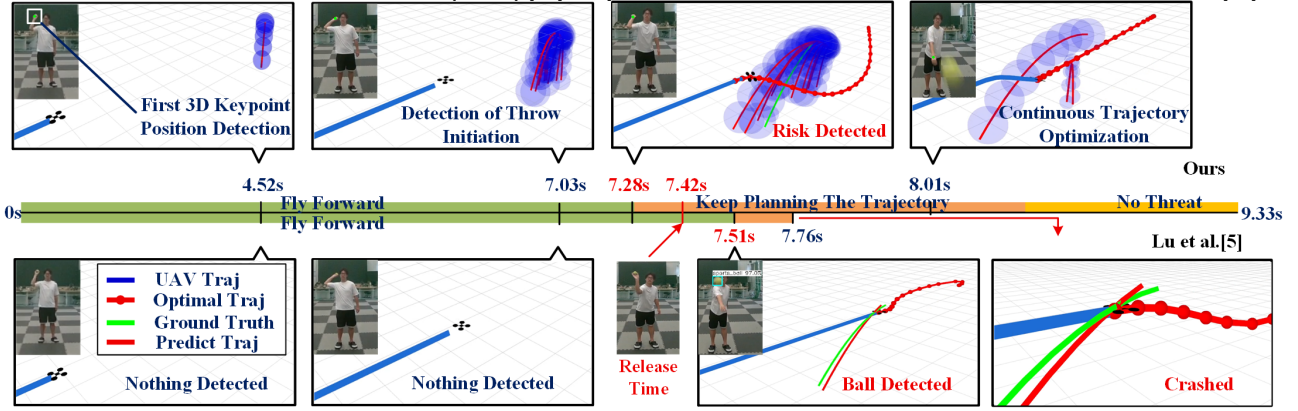


Fig. 5. Timeline analysis illustrating system performances for attack dodging within the same simulated environment. The figure presents a horizontal time axis with processing operations at distinct time nodes for two systems.

ered, demonstrating the robustness of the perception module.

C. Evaluation of the UAD Module

To validate the effectiveness of our UAD module, we conduct an ablation study in simulation. A collision is deemed to occur when the minimum distance d_{\min} between the UAV and the actual projectile is less than 0.4 m. The experimental procedure is as follows: (1) The UAV is controlled to fly in a real environment while collecting RGB-D images and trajectory data; a tennis ball is thrown toward the UAV at the final moment. (2) The recorded data are then imported into the simulation environment to test variations of our method. To ensure accurate synchronization between RGB-D information and positional data, the UAV follows the pre-recorded trajectory while processing the RGB-D images for perception. Once a potential threat is detected, the system immediately switches to its projectile dodging strategy and plans a trajectory to reach the predefined goal. Ours-NoTemporal method solely considers the single projectile trajectory predicted at the current timestep. Ours-NoSpatial incorporates temporal evolution but assumes perfect prediction accuracy, ignoring the spatial uncertainty. We perform 30 trials across diverse scenarios, recording the mean minimum distance and SR.

TABLE II
PERFORMANCE COMPARISON OF PROJECTILE DODGING STRATEGIES

	Ours-NoTemporal	Ours-NoSpatial	Ours
Average d_{\min} (m)	0.38	0.51	0.85
SR(%)	43.33	86.67	100

As shown in Table II, our method achieves the best performance in terms of both average minimum distance d_{\min} and dodging SR. The Ours-NoTemporal variant suffers the most, as it dodges non-existent projectile trajectories for most of the time, resulting in collisions with the actual projectile trajectory. This approach relies heavily on accurately predicting the projectile's release time—a task that is exceedingly challenging in practice. Ours-NoSpatial accounts for temporal uncertainty but neglects spatial uncertainty, assuming that a set of predicted trajectories includes the actual projectile trajectory. This leads to a degraded SR because prediction errors cause inadequate dodging maneuvers, resulting in collisions. In contrast, our method considers both temporal and spatial uncertainties,

enabling the UAV to anticipate and avoid threats by reacting to the broader, dynamically evolving risk region.

D. Evaluation of the Overall System

To demonstrate the superiority of our overall framework, we conduct comparative experiments against the baseline in [5] within a simulated environment. We utilize the experimental setup described in Section IV-C. Both systems adopt their own perception modules for perception and switch to their respective dodging strategies upon detecting a potential attack.

Fig. 5 illustrates a timeline for a representative dodging scenario, where UAVs initially follow the same pre-recorded trajectory in both systems. At 4.52s, our method first detects the 3D positions of human keypoints, 2.99s earlier than the baseline's initial object detection [5]. Our method detects the throw initiation at 7.03s, and then predicts a potential collision risk at 7.28s via HPE and commences trajectory planning. The projectile is released at 7.42s, by which time planning has already been underway for 0.14s. Consequently, the UAV dodges the uncertainty region that encompassing the actual projectile trajectory. As the set of surviving projectile trajectories decreases, the UAV continuously refines its trajectory to reach the goal efficiently. In contrast, the baseline [5] detects the ball only at 7.51s, when it is in mid-air, affording insufficient planning time and resulting in a collision at 7.76s. This 0.23s detection advantage enables our successful dodging. Overall, we perform 30 trials across diverse scenarios and compute the corresponding SRs. Our approach yields a dodging SR of 96.67%, markedly outperforming the baseline's 60%, underscoring the efficacy against human-initiated threats.

E. Real-World Deployment and Evaluation

We evaluate the dodging capabilities under two main real-world scenarios. In the first, a hovering UAV faces a projectile thrown by an attacker, as shown in Fig. 1. In the second, the UAV navigates toward a goal while subjected to sudden human-initiated attacks, as shown in Fig. 6, where it performs an emergency lateral maneuver to dodge the projectile, reaches the target, and avoids collisions with static obstacles.

1) *Evaluation of Performance Limits:* This subsection evaluates the extreme dodging capabilities of our framework in real-world conditions. Because manually launched projectile speeds vary, directly measuring a single maximum dodge

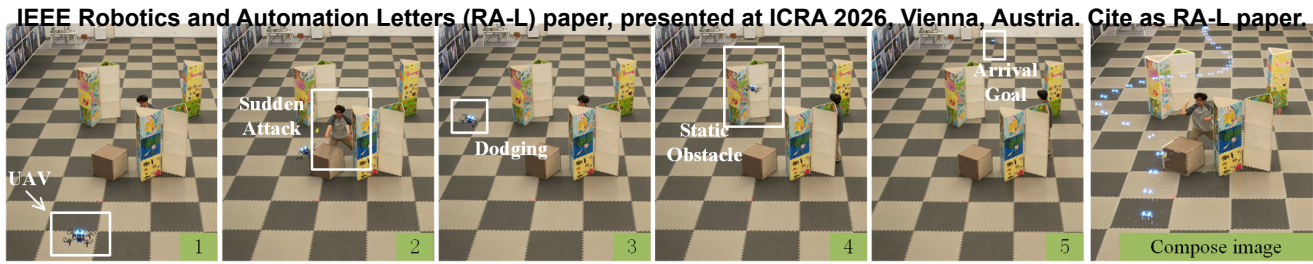


Fig. 6. Sequence of a UAV detecting a sudden attack, avoiding the attack and static obstacles, and reaching the target during flight.

TABLE III
DODGING SR ACROSS VARIOUS CONDITIONS WHILE HOVERING

Attacker-to-UAV Dist.(m)		Success Rate (%)															
		3				4				5				6			
Proj. Speed Group	Avg. Proj. Speed (m/s)	Low	Med	High	Ext	Low	Med	High	Ext	Low	Med	High	Ext	Low	Med	High	Ext
	30°	100	100	85.7	71.4	100	100	100	85.7	100	85.7	100	100	100	100	71.4	71.4
	0°	100	100	85.7	100	100	100	85.7	71.4	100	100	100	85.7	85.7	100	85.7	85.7
	-30°	100	100	100	85.7	100	100	85.7	85.7	100	100	71.4	85.7	100	85.7	85.7	71.4
Overall		100	100	90.5	81.0	100	100	90.5	81.0	100	95.2	90.5	90.5	95.2	95.2	81.0	76.2

speed is infeasible. We therefore infer the system limit indirectly from dodging SRs. Specifically, for each attacker-to-UAV distance (3 m, 4 m, 5 m and 6 m) and attack angle (30°, 0°, and -30°), we perform 21 trials with a spread of throwing intensities and group trials into four speed bands (Low, Medium, High, and Extreme, with each group represented by its respective average speed). We adopt the operational criterion that a speed band is considered to approach the UAV’s performance limit when the band exhibits two or more failures.

As shown in Table III, the results reveal that the UAV achieves high SRs in projectile dodging under low- and medium-speed conditions across all distances and angles, with an overall SR consistently at or above 95.2%. At 3 m, the UAV reached a performance limit of approximately 5.27 m/s, as two failures occurred in the high-speed group. Analogously, the limits are inferred as 7.3 m/s at 4 m and 9.01 m/s at 5 m. At 6 m, the overall SR decreased, likely due to amplified prediction errors in projectile position and velocity as distance and speed escalate, leading to erroneous dodging decisions by the UAV. We further test higher projectile speeds in the Extreme group, where SR shows a slight decline. Nevertheless, even under the most challenging 6 m and 13.29 m/s condition, the UAV maintains an SR above 76.2%, and achieves an overall SR of 82.4% in the Extreme group, demonstrating robust performance for most human-initiated attack scenarios.

2) *Evaluation of System Robustness*: To validate the robustness of our system, we conduct three sets of experiments. (1) A single participant performs throws using varied joint configurations (Right-hand Visible throw(RV), Right-hand Hidden throw(RH), Left-hand Visible throw(LV), Left-hand Hidden throw(LH)) and throwing motions (Downward, Upward, Right-to-Left, Left-to-Right) at a distance of 4 m from the hovering UAV, with throwing speeds around 6 m/s. Each unique combination of joint configuration and throwing motion is repeated 21 times. (2) Six participants (A–F), exhibiting diverse throwing styles and physical builds, throw projectiles of varying sizes and shapes (a tennis ball, a 3D-printed part, a bottle, a loopy toy, a football and a slipper) from the same distance and with speeds similar to those in the

first experiment. Each configuration is also repeated 21 times. (3) A single participant stands approximately 4 m from the UAV and performs throwing motions with release speeds of about 6 m/s. Three occlusion conditions are tested: partial-arm occlusion, half-body occlusion, and full-body occlusion. Each condition is repeated 21 times to ensure statistical reliability.

As shown in Fig. 7(a), the system exhibits consistent performance across diverse movement configurations, evidenced by a standard deviation (SD) of 5.45% and a coefficient of variation (CV) of 5.8% in the dodging SR for various joint positions and throwing motions. By leveraging human pose detection, the system effectively infers the projectile trajectory from global skeletal data, even in the presence of occlusions. Notably, hidden throws achieve higher SR compared to visible throws, attributable to the extended preparatory motions in concealed throws, which elongate the wind-up phase and provide the UAV with approximately 20 ms of additional reaction time. Throws using the non-dominant hand (e.g., left-handed right-to-left motion) exhibit reduced SR owing to wrist instability and excessive force, leading to inaccuracies in trajectory prediction. As illustrated in Fig. 7(b), the second experimental set presents an SD of 4.60% and a CV of 5.01%, indicating minimal influence from variations in participants and projectiles. We also test dual-attacker simultaneous throws toward a hovering UAV, the system successfully identifies and tracks both attack trajectories and executes coordinated avoidance maneuvers, as shown in Fig. 7(c).

TABLE IV
DODGING SR UNDER DIFFERENT OCCLUSION CONDITIONS.

Condition	Partial-Arm	Half-Body	Full-Body
Success Rate (%)	95.2	95.2	90.4

In occlusion experiments, the UAV achieves an average dodging SR of 93.6%, demonstrating reliable performance even with incomplete visual information. The minor differences among occlusion conditions arise because human attackers generally follow a similar motion pattern, emerging from behind obstacles before releasing the projectile, which enables the perception module to extract sufficient cues for

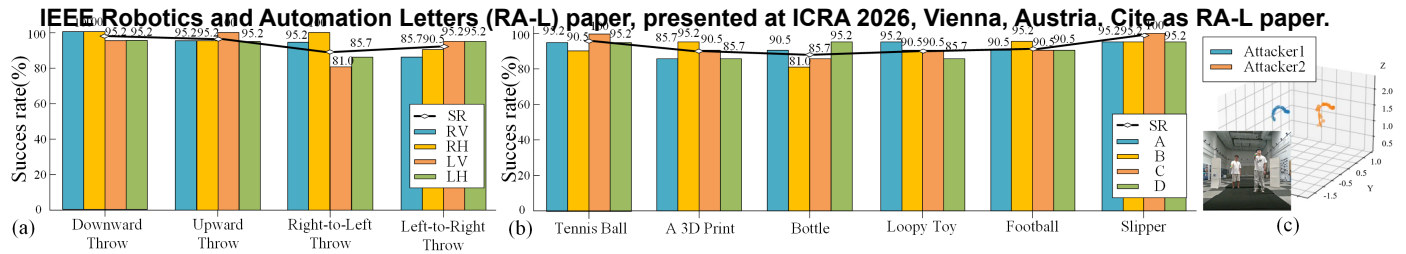


Fig. 7. (a) Success rates under varied throwing styles and joint configurations by a single participant. (b) Success rates with different projectiles thrown by four participants. (c) Detection and prediction of two simultaneous attacks.

threat prediction. Overall, all these findings underscore the system’s robustness and adaptability in varied scenarios.

V. CONCLUSION AND FUTURE WORK

In this letter, we propose a novel real-time framework that enables UAVs to dodge small and fast projectiles using an RGB-D camera. The system integrates a Pose-Aware Projectile Trajectory Prediction module and an Uncertainty-Aware Dodging module, allowing the UAV to anticipate threats from human motion cues and execute agile maneuvers under uncertainty, and can be easily transferred across different UAV platforms due to its modular perception–planning architecture. Beyond this specific case, the framework contributes to enhancing human–UAV interaction safety, where UAVs interpret human motion cues to operate safely in dynamic and uncertain environments. Potential applications include sports broadcasting and aerial filming, where UAVs avoid incoming balls; security and patrolling missions, where they cope with sudden human-initiated threats; and crowded public events, where they predict local motion trends to navigate safely. Future work will pursue semantic understanding of human posture and intention, enabling UAVs to perform more context-aware and purposeful actions beyond simple avoidance.

REFERENCES

- [1] D. Falanga, K. Kleber, and D. Scaramuzza, “Dynamic obstacle avoidance for quadrotors with event cameras,” *Sci. Robotics*, vol. 5, no. 40, p. 9712, 2020.
- [2] Y. Wang, J. Wang, X. Zhou, T. Yang, C. Xu, and F. Gao, “Safe and agile transportation of cable-suspended payload via multiple aerial robots,” *arXiv:2501.15272*, 2025.
- [3] “Drone attacked by rock throwing person,” 2023, accessed: 2025-04-04. [Online]. Available: https://www.reddit.com/r/drones/comments/16t6xo2/drone_attacked_by_rock_throwing_person/
- [4] “Woman accused of damaging dps drone during protest at state capitol in phoenix,” 2025, accessed: 2025-04-04. [Online]. Available: <https://www.azfamily.com/2025/02/04/woman-accused-damaging-dps-drone-during-protest-state-capitol-phoenix/>
- [5] M. Lu, H. Chen, and P. Lu, “Perception and avoidance of multiple small fast moving objects for quadrotors with only low-cost RGBD camera,” *IEEE Robotics Autom. Lett.*, vol. 7, no. 4, pp. 11 657–11 664, 2022.
- [6] T. Jiang, P. Lu, L. Zhang, N. Ma, R. Han, C. Lyu, Y. Li, and K. Chen, “Rtcompose: Real-time multi-person pose estimation based on mmpose,” *arXiv:2303.07399*, 2023.
- [7] T. Nageli, J. Alonso-Mora, A. Domahidi, D. Rus, and O. Hilliges, “Real-time motion planning for aerial videography with real-time with dynamic obstacle avoidance and viewpoint optimization,” *IEEE Robotics Autom. Lett.*, vol. 2, no. 3, pp. 1696–1703, 2017.
- [8] H. Zhu and J. Alonso-Mora, “Chance-constrained collision avoidance for mavs in dynamic environments,” *IEEE Robotics Autom. Lett.*, vol. 4, no. 2, pp. 776–783, 2019.
- [9] J. Lin, H. Zhu, and J. Alonso-Mora, “Robust vision-based obstacle avoidance for micro aerial vehicles in dynamic environments,” in *Proc. of the IEEE Intl. Conf. on Robot. and Autom.*, 2020, pp. 2682–2688.
- [10] H. Chen and P. Lu, “Real-time identification and avoidance of simultaneous static and dynamic obstacles on point cloud for uavs navigation,” *Robotics and Autonomous Systems*, vol. 154, p. 104124, 2022.
- [11] Z. Xu, H. Jin, X. Han, H. Shen, and K. Shimada, “Intent prediction-driven model predictive control for UAV planning and navigation in dynamic environments,” *IEEE Robotics Autom. Lett.*, vol. 10, no. 5, pp. 4946–4953, 2025.
- [12] E. Mueggler, N. Baumli, F. Fontana, and D. Scaramuzza, “Towards evasive maneuvers with quadrotors using dynamic vision sensors,” in *Proc. of the Euro. Conf. on Mobile Robot.*, 2015, pp. 1–8.
- [13] A. Mitrokhin, C. Fermüller, C. Parameshwara, and Y. Aloimonos, “Event-based moving object detection and tracking,” in *Proc. of the IEEE/RSSJ Intl. Conf. on Intell. Robots and Syst.*, 2018, pp. 1–9.
- [14] N. J. Sanket, C. M. Parameshwara, C. D. Singh, A. V. Kuruttukulam, C. Fermüller, D. Scaramuzza, and Y. Aloimonos, “Evdodgenet: Deep dynamic obstacle dodging with event cameras,” in *Proc. of the IEEE Intl. Conf. on Robot. and Autom.*, 2020, pp. 10 651–10 657.
- [15] D. Falanga, S. Kim, and D. Scaramuzza, “How fast is too fast? the role of perception latency in high-speed sense and avoid,” *IEEE Robotics Autom. Lett.*, vol. 4, no. 2, pp. 1884–1891, 2019.
- [16] B. He, H. Li, S. Wu, D. Wang, Z. Zhang, Q. Dong, C. Xu, and F. Gao, “Fast-dynamic-vision: Detection and tracking dynamic objects with event and depth sensing,” in *Proc. of the IEEE/RSSJ Intl. Conf. on Intell. Robots and Syst.*, 2021, pp. 3071–3078.
- [17] F. Kong, W. Xu, Y. Cai, and F. Zhang, “Avoiding dynamic small obstacles with onboard sensing and computation on aerial robots,” *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7869–7876, 2021.
- [18] “Yolo-fastestv2,” 2021, developed by GitHub user dog-qiuqiu, Accessed: 2024-08-06. [Online]. Available: <https://github.com/dog-qiuqiu/Yolo-FastestV2>
- [19] X. Chen, S. Shi, C. Zhang, B. Zhu, Q. Wang, K. C. Cheung, S. See, and H. Li, “Trajectoryformer: 3d object tracking transformer with predictive trajectory hypotheses,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 18 527–18 536.
- [20] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, “Blazepose: On-device real-time body pose tracking,” *arXiv:2006.10204*, 2020.
- [21] Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y. Sheikh, “Openpose: Realtime multi-person 2d pose estimation using part affinity fields,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 172–186, 2021.
- [22] H. Fang, J. Li, H. Tang, C. Xu, H. Zhu, Y. Xiu, Y. Li, and C. Lu, “Alphapose: Whole-body regional multi-person pose estimation and tracking in real-time,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 6, pp. 7157–7173, 2023.
- [23] Y. Qiu, C. Wang, W. Wang, M. Henein, and S. A. Scherer, “Airdos: Dynamic SLAM benefits from articulated objects,” in *Proc. of the IEEE Intl. Conf. on Robot. and Autom.*, 2022, pp. 8047–8053.
- [24] T. Salzmann, H. L. Chiang, M. Ryll, D. Sadigh, C. Parada, and A. Bewley, “Robots that can see: Leveraging human pose for trajectory prediction,” *IEEE Robotics Autom. Lett.*, vol. 8, no. 11, pp. 7090–7097, 2023.
- [25] Z. Xu, X. Han, H. Shen, H. Jin, and K. Shimada, “Navrl: Learning safe flight in dynamic environments,” *IEEE Robotics Autom. Lett.*, vol. 10, no. 4, pp. 3668–3675, 2025.
- [26] Q. Ren, H. Yu, J. Dai, Z. Zheng, J. Meng, L. Xu, C. Xu, F. Gao, and Y. Cao, “Intention-aware planner for robust and safe aerial tracking,” in *Proc. of the IEEE/RSSJ Intl. Conf. on Intell. Robots and Syst.*, 2024.
- [27] W. Zhu, X. Ma, Z. Liu, L. Liu, W. Wu, and Y. Wang, “Motionbert: A unified perspective on learning human motion representations,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 15 085–15 099.
- [28] Z. Wang, X. Zhou, C. Xu, and F. Gao, “Geometrically constrained trajectory optimization for multicopters,” *IEEE Trans. Robotics*, vol. 38, no. 5, pp. 3259–3278, 2022.