ISB-NMPC: Interception-Oriented Spherical-Bearing Nonlinear Model Predictive Control for Quadrotors

Li Hong[1], Linkai Liu[2], Xiangyu Ren[2], Shuli Lv[2], Chen Min[1], Chunjiang Wu[3], Yujun Zhang[1], Dingchen Zeng[2], and Quan Quan[2,*]

arXiv:2610.34789v1 [cs.RO] 26 Oct 2026

***Abstract*—Quadrotor interceptors equipped with strapdown cameras offer a low-cost solution for physically countering unauthorized UAVs. However, this configuration inherently couples vehicle maneuvering and camera pointing, creating a conflict between rapid closing and maintaining target visibility. We present interception-oriented spherical-bearing nonlinear model predictive control (ISB-NMPC), which couples spherical-bearing prediction with nonlinear quadrotor dynamics to jointly optimize target-directed closing motion and visibility. An attitude-compensated extended Kalman filter on $\mathbb{S}^2$ estimates target bearing motion, compensates for perception delay, and predicts target bearings to guide scale-adaptive detection. Across 30 paired simulation trials, ISB-NMPC achieves an interception success rate of 86.7%, the highest among the evaluated controllers, with a mean per-trial target-visible time of 99.11%.**

Outdoor flights demonstrate acquisition of a $30\,\mathrm{cm}$ target at approximately $150\,\mathrm{m}$, with interceptor speed reaching $17\,\mathrm{m\,s^{-1}}$ in the long-range flight and approximately $14\,\mathrm{m\,s^{-1}}$ while pursuing a quadrotor-towed maneuvering target. The software and hardware are open source at https://github.com/HOLYNOTBAD/ISB-NMPC and https://github.com/HOLYNOTBAD/holydrone, respectively.

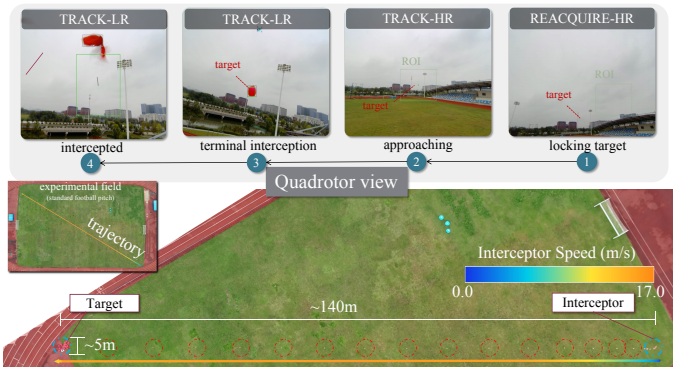
***Index Terms*—Aerial interception, nonlinear model predictive control, quadrotor, visual servoing, delay compensation.**

I. Introduction

The growth of low-altitude aerial applications makes shared-airspace safety increasingly important. Non-cooperative small targets create a practical need for affordable onboard interception systems [1], [2]. A strapdown monocular camera offers lightweight, mechanically simple sensing, but the interceptor must acquire the target, retain visual feedback, and close to a capture region without cooperative target-state measurements. Figure 1 illustrates the long-range and maneuvering-target scenarios considered here.

This mounting arrangement creates a fundamental control conflict. Target-directed acceleration requires tilting the thrust vector, which rotates the camera and can push the target toward the field-of-view (FOV) boundary. Recentering the target can, in turn, restrict closing acceleration. Thus, the same thrust and attitude motion govern both closing efficiency and visibility [3], requiring coordination under nonlinear dynamics and actuator limits (Fig. 2). Camera rotation, detection delay, and target-scale variation further affect the supporting visual feedback.

Existing approaches follow three complementary routes: target-state estimation and trajectory planning, visual servoing and line-of-sight (LOS) guidance, and perception-aware model predictive control (MPC). The first plans and tracks interception trajectories. Optimization-based planning explicitly han-



(a) Long-range interception of a 30 cm quasi-static target over an approximately 150 m approach.



(b) Pursuit–evasion interception of a 30 cm-diameter quadrotor-towed target from an initial separation of approximately 50 m.

Fig. 1. Outdoor flight experiments with ISB-NMPC. Trajectories are colored by interceptor speed; insets show onboard views.

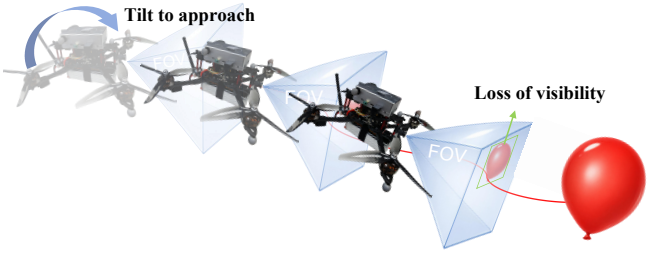


Fig. 2. Closing–visibility trade-off: target-directed acceleration tilts the strapdown camera and may reduce the target's FOV margin.

*Corresponding author: Quan Quan (e-mail: qq_buaa@buaa.edu.cn).

[1]Li Hong, Chen Min, and Yujun Zhang are with Shen Yuan Honors College, Beihang University, Beijing, China (e-mail: hongli@buaa.edu.cn for Li Hong).

[2]Linkai Liu, Xiangyu Ren, Shuli Lv, Dingchen Zeng, and Quan Quan are with the School of Automation Science and Electrical Engineering, Beihang University, Beijing, China.

[3]Chunjiang Wu is with the Hangzhou International Innovation Institute, Beihang University, Hangzhou, China.

dles engagement geometry, but its execution depends on the available target-state information [2]. A second route generates motion directly from visual feedback or relative geometry. Image-based visual servoing (IBVS) regulates image features, whereas position-based visual servoing (PBVS) uses estimated relative pose [4], [5]; barrier-function constraints can preserve visibility [6].

For interception, LOS guidance connects visual feedback to closing. Yang et al. developed high-speed image-based control [7] and analyzed LOS-constrained multirotor interceptability [8]. Related studies address planar-sector guidance [9], pursuit–evasion interceptability [10], and persistent interception [11]. Bearing-only following also uses helical motion to improve observability while maintaining a prescribed geometry [1]. These methods establish useful feedback structures; rapid capture additionally requires coordinating closing demand with visibility-preserving vehicle motion.

The third coordinates motion and sensing over a prediction horizon. Falanga et al. [12] incorporate perception into quadrotor motion optimization. Related formulations address reactive collision and occlusion avoidance [13] and combine visibility with actuator constraints [14], while robust visual predictive controllers address tracking and visibility constraints [15]. Qin et al. [16] combine quaternion-based spherical features and visibility constraints for aggressive flight, and Li et al. [17] integrate estimation and control in a robocentric model. MPC has also been applied to non-cooperative aerial interception and impact-angle-constrained visual engagement [18], [19].

For strapdown interception, the remaining question is how to choose closing direction and speed under visibility and actuation constraints. Uncoordinated closing can lose visual feedback, whereas recentering alone can delay capture. We connect LOS guidance to spherical prediction through an optimized closing-speed state; paired simulations and flight observations examine closing and visual retention together.

The supporting perception front end draws on depth-independent bearing estimation [20], [21], small-object slicing [22], and delayed-feedback estimation [23] to address camera rotation, target-scale variation, and latency.

We propose interception-oriented spherical-bearing nonlinear model predictive control (ISB-NMPC), predicting $SO(3)$-lifted bearings jointly with quadrotor dynamics. LOS velocity tracking, bearing-dependent speed demand, and visual costs and constraints coordinate approach and visibility. Compensated bearings and predicted regions of interest (ROIs) support detection.

The main contributions are as follows:

- An interception-oriented spherical-bearing NMPC that jointly optimizes desired closing speed and vehicle motion. LOS-aligned relative velocity objectives, visual costs, and soft FOV constraints coordinate rapid closing and target visibility under quadrotor dynamics and actuator limits.
- A spherical-bearing estimator with attitude and delay compensation that supplies current-time bearing feedback and absolute-time ROI predictions to support onboard operation during large attitude changes and target-scale variation.
- Paired controller comparisons over 30 simulation trials and outdoor flights involving long-range acquisition of a quasi-static small target and pursuit of a maneuvering target, evaluating controller performance and the operation of the integrated perception–control system.

Section II defines the system and interception problem. Sections III and IV present the perception front end and ISB-NMPC, respectively. Section V reports simulation and flight validation, followed by the discussion and conclusion.

II. System Architecture and Problem Formulation

A. System Architecture

The onboard system comprises a PX4 flight controller whose extended Kalman filter (EKF) supplies ego-state, a strapdown monocular USB camera, and an NVIDIA Jetson Orin NX companion computer. As shown in Fig. 3, 30 Hz images feed a scale-adaptive YOLOv5n detector,[1] and a delay-compensated $\mathbb{S}^2$ EKF (DC-$\mathbb{S}^2$EKF) fuses the detections with this ego-state to provide current-time bearings and ROI predictions, where $\mathbb{S}^2 := \{\mathbf{n} \in \mathbb{R}^3 : \|\mathbf{n}\|_2 = 1\}$ denotes the unit two-sphere. ISB-NMPC runs at 50 Hz and sends collective-thrust and body-rate commands to PX4.

B. Coordinate Frames and Quadrotor Dynamics

Fig. 4 defines the world, body, camera, and target frames $\mathcal{F}^w$, $\mathcal{F}^b$, $\mathcal{F}^c$, and $\mathcal{F}^t$. The vector ${}^a\mathbf{p}_b$ points from the origin of $\mathcal{F}^a$ to that of $\mathcal{F}^b$ and is expressed in $\mathcal{F}^a$; the unit quaternion $\mathbf{q}_{ab} \in \mathbb{S}^3$ maps vectors from $\mathcal{F}^b$ to $\mathcal{F}^a$ through $R(\mathbf{q}_{ab}) \in SO(3)$. Quaternions are ordered scalar-first, composition is denoted by $\otimes$, and $\mathbf{e}_3 = [0, 0, 1]^\top$. The fixed camera extrinsic $\mathbf{q}_{bc}$ and vehicle attitude $\mathbf{q}_{wb}$ give $\mathbf{q}_{wc} = \mathbf{q}_{wb} \otimes \mathbf{q}_{bc}$ and $\mathbf{q}_{cw} = \mathbf{q}_{wc}^{-1}$. Let ${}^b\mathbf{p}_c$ denote the camera position in the body frame.

Given the UAV and target positions ${}^w\mathbf{p}_b$ and ${}^w\mathbf{p}_t$, the target position and unit bearing in the camera frame are

$$\begin{aligned} {}^c\mathbf{p}_t &= R(\mathbf{q}_{bc})^\top \left[R(\mathbf{q}_{wb})^\top \left({}^w\mathbf{p}_t - {}^w\mathbf{p}_b \right) - {}^b\mathbf{p}_c \right], \\ \mathbf{n}_t &= \frac{{}^c\mathbf{p}_t}{\|{}^c\mathbf{p}_t\|} \in \mathbb{S}^2. \end{aligned} \tag{1}$$

The controller uses the mass-normalized collective thrust T and body angular velocity ${}^b\boldsymbol{\omega}$ as inputs. The quadrotor model is

$$ {}^w\dot{\mathbf{p}}_b = {}^w\mathbf{v}_b, \qquad {}^w\dot{\mathbf{v}}_b = TR(\mathbf{q}_{wb})\mathbf{e}_3 - g\mathbf{e}_3, \tag{2}$$

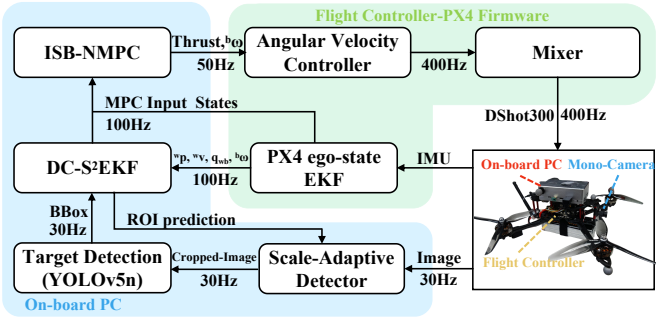


Fig. 3. Onboard architecture integrating scale-adaptive detection, DC-$\mathbb{S}^2$EKF bearing estimation, ISB-NMPC, and PX4 actuation.

[1] YOLOv5 source code: https://github.com/ultralytics/yolov5.

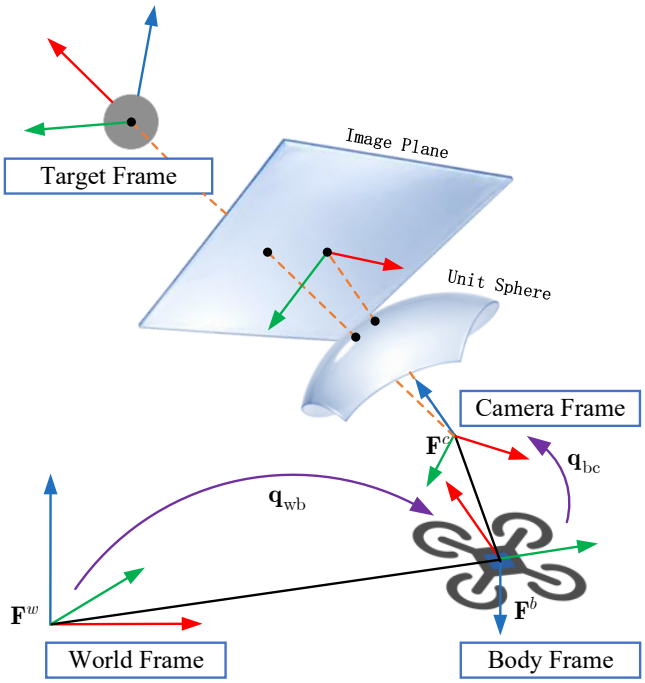


Fig. 4. World, body, camera, and target frames, with spherical bearing $\mathbf{n}_t \in \mathbb{S}^2$ and image-plane projection. Axes x, y, and z are red, green, and blue, respectively.

with attitude kinematics

$$\dot{\mathbf{q}}_{wb} = \frac{1}{2}\mathbf{q}_{wb} \otimes \begin{bmatrix} 0, & {}^b\boldsymbol{\omega}^\top \end{bmatrix}^\top. \tag{3}$$

The camera is rigidly attached to the body. Because the PX4 body-rate loop is substantially faster than the NMPC update, the commanded body rate is treated as a directly tracked input; rotational rigid-body, motor, and propeller dynamics are not included in the prediction model.

C. Visual Interception Problem

We consider a quadrotor intercepting a non-cooperative moving target using a rigidly mounted monocular camera, as illustrated in Fig. 2. The target does not communicate its state to the interceptor. The onboard front end always supplies a compensated bearing; when an onboard target-state estimate is available, its position and velocity are also used by the controller. In bearing-only operation, the controller initializes the prediction with a nominal range and zero target velocity. Let

$${}^w\mathbf{r}_t = {}^w\mathbf{p}_t - {}^w\mathbf{p}_b, \qquad r = \|{}^w\mathbf{r}_t\|, \qquad {}^w\mathbf{n}_t = \frac{{}^w\mathbf{r}_t}{r}, \tag{4}$$

denote the relative displacement, range, and line-of-sight (LOS) direction from the quadrotor to the target. With ${}^w\mathbf{v}_r = {}^w\mathbf{v}_b - {}^w\mathbf{v}_t$, the physical instantaneous closing speed is

$$s_c = -\dot{r} = {}^w\mathbf{n}_t^\top {}^w\mathbf{v}_r, \tag{5}$$

where $s_c > 0$ indicates that the range is decreasing.

Visibility is expressed using the camera-frame bearing $\mathbf{n}_t = [n_x, n_y, n_z]^\top$ defined in (1). Its normalized image coordinate is $\boldsymbol{\pi}(\mathbf{n}_t) = [n_x/n_z,\ n_y/n_z]^\top$ for $n_z > 0$. The ideal visibility requirement is $\boldsymbol{\pi}_{\min} \le \boldsymbol{\pi}(\mathbf{n}_t) \le \boldsymbol{\pi}_{\max}$, where the bounds are determined by the usable camera field of view.

LOS-constrained interceptability couples target-directed motion with visual contact [8], [24]. At fixed geometry and target velocity, Cauchy–Schwarz gives

$$\begin{aligned} s_c &= {}^w\mathbf{n}_t^\top {}^w\mathbf{v}_b - {}^w\mathbf{n}_t^\top {}^w\mathbf{v}_t \\ &\le s - {}^w\mathbf{n}_t^\top {}^w\mathbf{v}_t, \qquad \|{}^w\mathbf{v}_b\|_2 = s. \end{aligned} \tag{6}$$

Equality holds for ${}^w\mathbf{v}_b = s\,{}^w\mathbf{n}_t$: LOS alignment maximizes instantaneous closing, not minimum-time capture. With estimated target velocity, the nominal relative reference is ${}^w\mathbf{v}_{r,d} = s\,{}^w\mathbf{n}_t$, giving ${}^w\mathbf{v}_{b,d} = \hat{\mathbf{v}}_t + s\,{}^w\mathbf{n}_t$; bearing-only operation sets $\hat{\mathbf{v}}_t = \mathbf{0}$.

Dynamic coupling. Differentiating (5) using (2), for differentiable target velocity and $r > 0$, gives

$$\begin{aligned} \dot{s}_c &= {}^w\mathbf{n}_t^\top (T\mathbf{b}_3 - g\mathbf{e}_3 - {}^w\dot{\mathbf{v}}_t) - \frac{\|\mathbf{v}_\perp\|_2^2}{r}, \\ \mathbf{b}_3 &= R(\mathbf{q}_{wb})\mathbf{e}_3, \qquad \mathbf{v}_\perp = (\mathbf{I}_3 - {}^w\mathbf{n}_t\,{}^w\mathbf{n}_t^\top)\,{}^w\mathbf{v}_r. \end{aligned} \tag{7}$$

The negative transverse-motion term follows from ${}^w\dot{\mathbf{n}}_t = -\mathbf{v}_\perp/r$. Increasing thrust improves closing acceleration only when ${}^w\mathbf{n}_t^\top \mathbf{b}_3 > 0$; changing this alignment rotates the strapdown camera. The same body rate enters the rotational term of (13). Thus thrust and attitude cannot independently set closing and visibility. This motivates predicting both and optimizing the speed demand rather than imposing a fixed maximal speed.

The soft relaxation used by the controller is introduced in Sec. IV-B. Meanwhile, the control input must remain in the feasible set

$$\mathcal{U} = \left\{ (T, {}^b\boldsymbol{\omega}) \;\middle|\; T_{\min} \le T \le T_{\max},\ |\omega_i| \le \bar{\omega}_i,\ i \in \{x, y, z\} \right\}. \tag{8}$$

Interception problem. Given bearing feedback and available target-state estimates, choose $(T, {}^b\boldsymbol{\omega}) \in \mathcal{U}$ to reach $r(t_f) \le r_{\text{int}}$ rapidly while retaining the target within the FOV, without a preplanned target trajectory. Performance is assessed jointly by success rate, interception time among successful trials, and target-visible time: fast closing alone is insufficient. Section IV-B constructs a finite-horizon surrogate, not a guaranteed minimum-time interception solution. In particular, the non-cooperative target's future motion is unknown a priori.

III. Onboard Visual Perception Front End

The visual front end converts timestamped images and ego-motion into native-image bounding boxes, current-time unit bearings, residual bearing motion, and future ROI predictions for detection and control.

A. Bearing Estimation, Temporal Filtering, and Delay Compensation

The estimator preserves unit bearings during camera rotation and uses two-dimensional tangent innovations (Fig. 5). Superscripts r, $-$, and $+$ denote rotation-transported, prior, and posterior quantities; k indexes filter samples, a hat denotes an estimate, and t identifies the target.

At image time t_{img}, the bounding-box center $(u_{\text{img}}, v_{\text{img}})$ and camera intrinsics $\mathbf{K}$ give the unit bearing

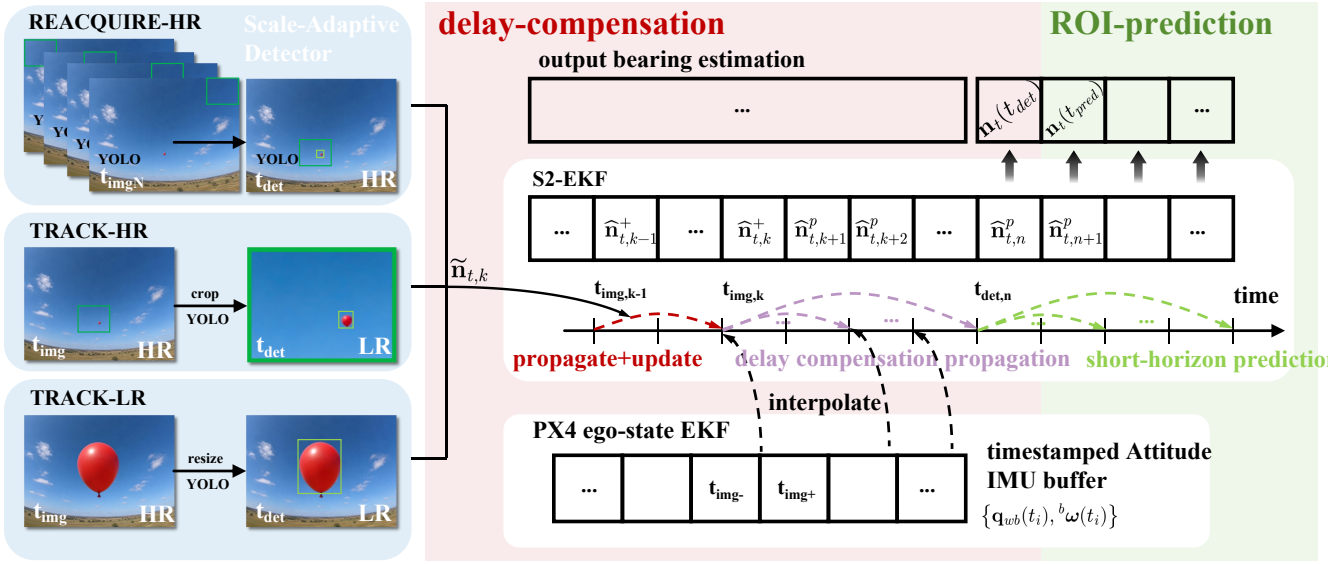


Fig. 6. Detection modes and timing of delay compensation and ROI prediction. Superscript p denotes forward predictions from the latest posterior, without modifying the EKF state.

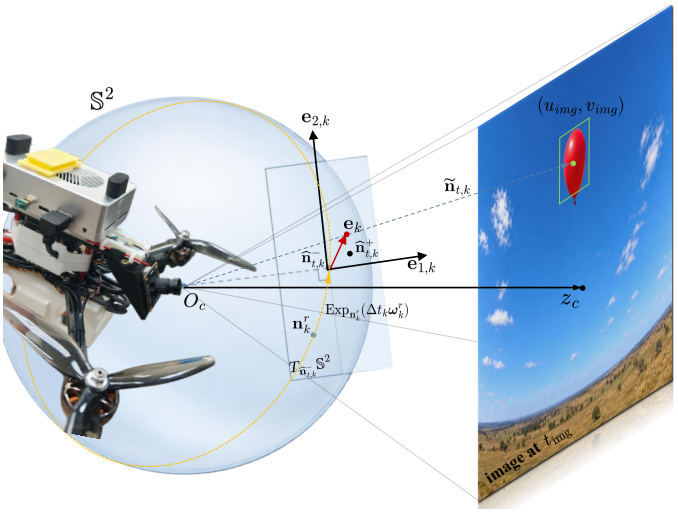


Fig. 5. DC-$\mathbb{S}^2$EKF geometry: spherical back-projection, rotation compensation, propagation, and tangent-space correction. Bearings and tangent bases are shown schematically in the camera frame at measurement time.

$$\widetilde{\mathbf{n}}_t(t_{\mathrm{img}}) = \frac{\mathbf{K}^{-1}[u_{\mathrm{img}}, v_{\mathrm{img}}, 1]^\top}{\|\mathbf{K}^{-1}[u_{\mathrm{img}}, v_{\mathrm{img}}, 1]^\top\|} \in \mathbb{S}^2. \tag{9}$$

The DC-$\mathbb{S}^2$EKF separates camera rotation from residual LOS motion using the bearing $\widehat{\mathbf{n}}_t$ and tangent velocity $\widehat{\boldsymbol{\omega}}_n \in T_{\widehat{\mathbf{n}}_t}\mathbb{S}^2$. Local tangent coordinates follow the manifold-estimation principle in [25]. For two times t_a and t_b, define the camera-frame relative rotation $\mathbf{R}_{c(t_b)c(t_a)} = R(\mathbf{q}_{wc}(t_b))^\top R(\mathbf{q}_{wc}(t_a))$. For estimator times t_k and $\Delta t_k = t_k - t_{k-1}$, let $\mathbf{R}_k = \mathbf{R}_{c(t_k)c(t_{k-1})}$. Propagation compensates camera rotation before integrating residual motion:

$$\begin{aligned} \mathbf{n}^r_k &= \mathbf{R}_k \widehat{\mathbf{n}}_{t,k-1}, & \boldsymbol{\omega}^r_k &= \Pi_{\mathbf{n}^r_k} \mathbf{R}_k \widehat{\boldsymbol{\omega}}_{n,k-1}, \\ \widehat{\mathbf{n}}^-_{t,k} &= \mathrm{Exp}_{\mathbf{n}^r_k}(\Delta t_k \boldsymbol{\omega}^r_k), & \Pi_{\mathbf{n}} &= \mathbf{I}_3 - \mathbf{n}\mathbf{n}^\top. \end{aligned} \tag{10}$$

The predicted residual velocity is projected into the new tangent plane, $\widehat{\boldsymbol{\omega}}^-_{n,k} = \Pi_{\widehat{\mathbf{n}}^-_{t,k}} \boldsymbol{\omega}^r_k$. This residual tangent velocity differs from the total spin-free angular velocity $\mathbf{n}_t \times \dot{\mathbf{n}}_t$ used in the control cost.

With tangent basis $\mathbf{B}_k = [\mathbf{e}_{1,k}, \mathbf{e}_{2,k}]$, the local error comprises two bearing and two residual-velocity coordinates. A constant-velocity model propagates its covariance using Δt_k, residual-acceleration variance σ_a^2, and IMU rotation uncertainty σ_g^2. The update is

$$\begin{aligned} \mathbf{y}_k &= \mathbf{B}_k^\top \mathrm{Log}_{\widehat{\mathbf{n}}^-_{t,k}}(\widetilde{\mathbf{n}}_{t,k}), \\ \mathbf{K}_k &= \mathbf{P}^-_k \mathbf{H}^\top (\mathbf{H}\mathbf{P}^-_k \mathbf{H}^\top + \mathbf{R}^n_k)^{-1}, \\ \begin{bmatrix} \delta\boldsymbol{\theta}_k \\ \delta\boldsymbol{\omega}_k \end{bmatrix} &= \mathbf{K}_k \mathbf{y}_k, \\ \widehat{\mathbf{n}}^+_{t,k} &= \mathrm{Exp}_{\widehat{\mathbf{n}}^-_{t,k}}(\mathbf{B}_k \delta\boldsymbol{\theta}_k), \\ \widehat{\boldsymbol{\omega}}^+_{n,k} &= \Pi_{\widehat{\mathbf{n}}^+_{t,k}}(\widehat{\boldsymbol{\omega}}^-_{n,k} + \mathbf{B}_k \delta\boldsymbol{\omega}_k), \qquad \mathbf{H} = [\mathbf{I}_2 \ \mathbf{0}]. \end{aligned} \tag{11}$$

The angular covariance $\mathbf{R}^n_k$ follows from pixel covariance and the back-projection Jacobian, with a noise floor set by confidence, box area, boundary proximity, and temporal consistency through $\rho_k \in [\rho_{\min}, 1]$. Joseph-form covariance update and basis transport maintain consistency; the exponential map preserves bearing norm.

For a detection received at t_{det}, the measurement update is applied at t_{img} and propagated to the current time using buffered attitude and IMU data, with the relative rotation $\mathbf{R}_{c(t_{\mathrm{det}})c(t_{\mathrm{img}})}$ defined above. Propagation to controller queries supplies current-time bearings and to image acquisition times supplies ROI predictions. Missing timestamps use a calibrated delay; stale IMU data or prolonged detection loss triggers reacquisition.

B. Multi-Resolution Detection and ROI-Based Tracking

A TensorRT-accelerated YOLO backend switches among tiled high-resolution reacquisition, prediction-centered high-

resolution tracking, and downsampled full-frame tracking using hysteretic size and confidence thresholds. Valid predictions bridge brief dropouts; repeated loss triggers reacquisition. Figure 6 details these modes (left) and the timestamp-based delay compensation and ROI prediction process (right).

IV. Interception-Oriented Spherical-Bearing NMPC

The controller receives the current ego-state, compensated camera-frame bearing, and the available target-motion estimate. At every update it builds a short-horizon prediction, schedules the closing-speed reference from the current LOS error, solves the constrained optimal-control problem, and applies only the first thrust and body-rate command.

A. Spherical-Bearing Model and Prediction Dynamics

The rotation-equivariant unit bearing separates direction from range [20]. Its $SO(3)$ lift enables joint visual and attitude prediction [16].

1) SO(3) Lifting of the Filtered Bearing: The front end supplies $\mathbf{n}_t \in \mathbb{S}^2$ through the measurement, filtering, and delay-compensation process of Sec. III. For forward bearings $n_z = \mathbf{e}_3^\top \mathbf{n}_t > 0$, the minimal rotation from $\mathbf{e}_3$ to $\mathbf{n}_t$ has the nonsingular unit quaternion

$$ {}^c\mathbf{f} = \frac{[1+n_z, (\mathbf{e}_3 \times \mathbf{n}_t)^\top]^\top}{\sqrt{2(1+n_z)}} \in \mathbb{S}^3. \tag{12} $$

It recovers the bearing as $\mathbf{n}_t = R({}^c\mathbf{f})\mathbf{e}_3$; the camera-frame target position is ${}^c\mathbf{p}_t = r\mathbf{n}_t$, with $r = \|{}^c\mathbf{p}_t\|_2$.

2) Spherical-Bearing Kinematics: To propagate ${}^c\mathbf{f}$, account for relative target translation and camera rotation. Let ${}^c\boldsymbol{\omega}_{wc}$ denote the angular velocity of the camera frame with respect to the world frame, expressed in $\mathcal{F}^c$. Differentiating the normalized bearing gives

$$ \dot{\mathbf{n}}_t = -\left[{}^c\boldsymbol{\omega}_{wc}\right]_\times \mathbf{n}_t + \frac{1}{r}\left(\mathbf{I}_3 - \mathbf{n}_t\mathbf{n}_t^\top\right) {}^c\mathbf{v}_{t/b}. \tag{13} $$

Only transverse relative velocity changes bearing. Since $\dot{\mathbf{n}}_t \in T_{\mathbf{n}_t}\mathbb{S}^2$, define the spin-free angular velocity

$$ {}^c\boldsymbol{\omega}_n = \mathbf{n}_t \times \dot{\mathbf{n}}_t, \qquad \dot{\mathbf{n}}_t = \left[{}^c\boldsymbol{\omega}_n\right]_\times \mathbf{n}_t. \tag{14} $$

With $\mathbf{n}_t = R({}^c\mathbf{f})\mathbf{e}_3$, the quaternion kinematics used in the prediction model are

$$ {}^c\dot{\mathbf{f}} = \frac{1}{2}\begin{bmatrix} 0 \\ {}^c\boldsymbol{\omega}_n \end{bmatrix} \otimes {}^c\mathbf{f}. \tag{15} $$

This directly propagates the SO(3)-lifted visual feature while preserving the unit-norm constraint of the quaternion representation.

3) NMPC State and Control Input: The prediction state combines vehicle motion, the visual feature, and interception variables:

$$ \mathbf{x} = [{}^w\mathbf{p}_b^\top, \mathbf{q}_{wb}^\top, {}^w\mathbf{v}_b^\top, {}^c\mathbf{f}^\top, r, s, z_u, z_v]^\top \in \mathbb{R}^{18}. \tag{16} $$

Vehicle variables follow Sec. II-B, with the current position as the local prediction origin. The desired closing-speed state s is distinct from physical closing speed $-\dot{r}$. Auxiliary accumulators z_u, z_v retain the FOV slacks as decision inputs in the generated Automatic Control and Dynamic Optimization (ACADO) [26] solver and are not physical states.

The control input is defined as

$$ \mathbf{u} = [T, {}^b\omega_x, {}^b\omega_y, {}^b\omega_z, u_s, \epsilon_u, \epsilon_v]^\top \in \mathbb{R}^7. \tag{17} $$

Here T and ${}^b\boldsymbol{\omega}$ are the thrust and body-rate inputs defined in Sec. II-B; u_s drives desired speed, and ϵ_u, ϵ_v relax horizontal and vertical FOV bounds, with $\dot{z}_u = \epsilon_u$ and $\dot{z}_v = \epsilon_v$.

4) Prediction Dynamics: The calibrated camera rotation gives the angular and relative velocities used in (13)–(15):

$$ {}^c\boldsymbol{\omega}_{wc} = R(\mathbf{q}_{bc})^\top {}^b\boldsymbol{\omega}, \tag{18} $$

$$ {}^w\mathbf{v}_r = {}^w\mathbf{v}_b - {}^w\hat{\mathbf{v}}_t, \tag{19} $$

$$ {}^c\mathbf{v}_{t/b} = -R(\mathbf{q}_{cw}) {}^w\mathbf{v}_r. \tag{20} $$

These differential kinematics neglect the small camera–body translational lever arm.

The range state obeys $\dot{r} = -s_c$ from (5). The desired closing-speed state is driven by the optimization input as $\dot{s} = u_s$. At prediction node i, the world-frame bearing is ${}^w\mathbf{n}_{t,i} = R(\mathbf{q}_{wb,i})R(\mathbf{q}_{bc})R({}^c\mathbf{f}_i)\mathbf{e}_3$. It evolves with the predicted attitude and visual feature; only the reference ${}^w\mathbf{n}_{t,0}$ in the objective is frozen. The velocity-tracking cost below encourages the physical closing speed to approach s.

Together, the vehicle, attitude, bearing, range, and desired-speed dynamics define the continuous-time prediction model.

B. NMPC Objective and Constraints

1) Objective: Equation (7) motivates the objective: LOS velocity tracking encourages approach and discourages off-axis motion, while visual costs account for camera rotation during acceleration. Optimizing s permits adjustment of the approach demand under vehicle constraints. Each solve holds the estimated target velocity, measured LOS ${}^w\mathbf{n}_{t,0}$, and bearing-based reference s_d constant. The interception residual promotes LOS-aligned closing motion [7] through $\mathbf{e}_{\text{int}} = {}^w\mathbf{v}_r - s\,{}^w\mathbf{n}_{t,0}$. To moderate the speed demand near the FOV boundary, s_d is scheduled using the horizontal and vertical LOS errors. For $\mathbf{n}_t = [n_x, n_y, n_z]^\top$, the implementation uses

$$ \begin{aligned} e_h &= \frac{n_x}{\sqrt{n_x^2 + n_z^2 + \varepsilon^2}}, \\ e_v &= \frac{n_y}{\sqrt{n_y^2 + n_z^2 + \varepsilon^2}}, \\ \alpha_{\text{los}} &= \exp[-(k_h e_h^2 + k_v e_v^2)], \\ s_d &= s_{\min} + \alpha_{\text{los}}(s_{\max} - s_{\min}), \end{aligned} \tag{21} $$

where $k_h, k_v \geq 0$ determine the horizontal and vertical scheduling sensitivity. Using the current bearing prevents predicted errors from reducing the optimizer's own speed reference. The running interception cost is

$$ \mathcal{L}_{\text{int}} = \|\mathbf{e}_{\text{int}}\|^2_{\mathbf{Q}_{\text{int}}} + q_s(s - s_d)^2. \tag{22} $$

Tracking s_d encourages rapid approach near the optical axis; optimizing s allows compromise with visual and actuation requirements.

Consistent with the field-of-view constraint, we formulate the visual objective directly in the normalized image plane. Let $\boldsymbol{\pi}_\varepsilon(\mathbf{n}_t) = [n_x/(n_z+\varepsilon), n_y/(n_z+\varepsilon)]^\top$ denote the regularized

normalized image projection. To measure the distance from the target projection to the image center, we define

$$\begin{aligned} \boldsymbol{\pi}_c &= \frac{\boldsymbol{\pi}_{\max} + \boldsymbol{\pi}_{\min}}{2}, \qquad \boldsymbol{\rho} = \frac{\boldsymbol{\pi}_{\max} - \boldsymbol{\pi}_{\min}}{2}, \\ \mathbf{e}_{\text{vis}} &= \operatorname{diag}(\boldsymbol{\rho})^{-1} \left(\boldsymbol{\pi}_\varepsilon(\mathbf{n}_t) - \boldsymbol{\pi}_c \right). \end{aligned} \tag{23}$$

Since velocity tracking alone cannot prevent camera-induced target displacement, the visual cost penalizes image offset and bearing motion:

$$\mathcal{L}_{\text{vis}} = \|\mathbf{e}_{\text{vis}}\|^2_{\mathbf{Q}_{\text{vis}}} + \|{}^c\boldsymbol{\omega}_n\|^2_{\mathbf{Q}_{\text{jit}}}, \tag{24}$$

where ${}^c\boldsymbol{\omega}_n$ is defined in (14).

The regularization term $\mathcal{L}_{\text{reg}} = \|\mathbf{u} - \mathbf{u}_{\text{ref}}\|^2_{\mathbf{R}_u}$ penalizes excessive actuation and aggressive closing-speed changes, where $\mathbf{u}_{\text{ref}} = [g, 0, 0, 0, 0, 0, 0]^\top$ is the nominal hover input. In particular, this term penalizes thrust deviation from gravity, body rates, closing-speed rate, and both FOV slacks. The terminal cost and its cross-track residual are

$$\begin{aligned} \mathcal{L}_N &= \|\mathbf{e}_{\text{vis},N}\|^2_{\mathbf{Q}_{\text{vis},N}} + q_{\ell,N} e^2_{\ell,N}, \\ e_{\ell,N} &= \|({}^w\mathbf{p}_N - {}^w\mathbf{p}_{\ell,0}) \times {}^w\mathbf{n}_{t,0}\|_2 . \end{aligned} \tag{25}$$

The residual $e_{\ell,N}$ penalizes the perpendicular deviation of the predicted terminal position from the line passing through the current local prediction origin ${}^w\mathbf{p}_{\ell,0}$ along the frozen LOS direction. This soft penalty discourages off-axis motion while the underactuated vehicle accelerates toward the desired closing speed.

2) Constraints: The input set and visibility requirements of Sec. II-C become actuator bounds and soft FOV constraints, supplemented by range and speed bounds:

$$\begin{aligned} \mathbf{u}_{\min} \le \mathbf{u} \le \mathbf{u}_{\max}, \\ 0 \le s \le v_{\ell,\max}, \\ r \ge r_{\min}, \qquad n_z \ge n_{z,\min}, \\ \boldsymbol{\pi}_{\min} - \boldsymbol{\epsilon} \le \boldsymbol{\pi}_\varepsilon(\mathbf{n}_t) \le \boldsymbol{\pi}_{\max} + \boldsymbol{\epsilon}, \\ \mathbf{0} \le \boldsymbol{\epsilon} = [\epsilon_u, \epsilon_v]^\top \le \boldsymbol{\epsilon}_{\max}. \end{aligned} \tag{26}$$

Here $\boldsymbol{\pi}_{\min}$ and $\boldsymbol{\pi}_{\max}$ denote the lower and upper normalized image bounds induced by the camera field of view. The bounded, nonnegative slack vector $\boldsymbol{\epsilon}$ makes the FOV constraint soft, and its large running penalty discourages violations. The additional lower bound on n_z keeps the predicted target in front of the camera. Since $\dot{s} = u_s$, the corresponding entries of $\mathbf{u}_{\min}$ and $\mathbf{u}_{\max}$ impose $|u_s| \le a_{\ell,\max}$.

3) Discrete Optimal Control Problem: Let Δt denote the prediction discretization interval, and let $\mathbf{F}(\cdot)$ be the discretized continuous-time prediction model. Define $\mathbf{X} = \{\mathbf{x}_i\}_{i=0}^{N}$ and $\mathbf{U} = \{\mathbf{u}_i\}_{i=0}^{N-1}$. At each control update, the following finite-horizon problem is solved:

$$\begin{aligned} \min_{\mathbf{X},\mathbf{U}} \quad & \sum_{i=0}^{N-1} \left(\mathcal{L}_{\text{int},i} + \mathcal{L}_{\text{vis},i} + \mathcal{L}_{\text{reg},i} \right) + \mathcal{L}_N \\ \text{s.t.} \quad & \mathbf{x}_0 = \hat{\mathbf{x}}, \\ & \mathbf{x}_{i+1} = \mathbf{F}(\mathbf{x}_i, \mathbf{u}_i, \hat{\mathbf{v}}_t), \\ & \mathbf{h}_c(\mathbf{x}_i, \mathbf{u}_i) \le \mathbf{0}, \qquad i = 0, \dots, N-1. \end{aligned} \tag{27}$$

Existence and scope of optimality. For fixed state and target-velocity estimates, assume a nonempty bounded input set, a projection denominator separated from zero, and at least one feasible control sequence. Continuity of $\mathbf{F}$ and the costs, together with the closed constraints, makes the feasible image of the compact input set compact; the Weierstrass theorem therefore guarantees a global minimizer of (27). This is an existence result only: the nonconvex dynamics, projection, and constraints mean that the real-time solver generally returns a local solution. Bounded FOV slack improves practical feasibility but does not guarantee recursive feasibility, which would require a compatible invariant terminal set and controller.

C. Real-Time Implementation

ISB-NMPC runs at 50 Hz in Robot Operating System (ROS) using an ACADO-generated real-time-iteration solver with qpOASES [27]. Ten 0.05 s intervals form the 0.5 s horizon; Gauss–Newton approximation, implicit Gauss–Legendre integration, full condensing, and warm starts are used. With seven inputs and $N = 10$, the condensed subproblem has at most 70 control increments and dense linear algebra scales as $\mathcal{O}((7N)^3)$. Typical onboard solve time on the Jetson Orin NX is about 10 ms versus the 20 ms control period, not a worst-case guarantee. A failed solve briefly reuses the last valid command; persistent failure resets the warm start and commands level hover.

V. Simulation and Experimental Validation

Simulation isolates controller performance with ground-truth target inputs; outdoor flights evaluate the integrated bearing-only system.

A. Simulation Validation and Comparison

Thirty paired trials use ROS Noetic, Gazebo Classic 11, and PX4 software-in-the-loop at 50 Hz with identical platform settings and random seeds. Trials start near 30 m altitude and 100 m separation. The 1 m-radius target follows a three-dimensional random walk bounded by 5 m/s and 1.5 m/s^2, with direction updated every 3 s. At 30 Hz, ground-truth target inputs isolate controller performance. Simulations run on an AMD Ryzen 9 9950X3D central processing unit and an NVIDIA RTX 5080 graphics processing unit.

TABLE I
Simulation comparison of interception controllers over 30 paired random-walk target-motion trials.

Metric	ISB-NMPC	PS-LOS	LOS-IBVS	Back-stepping
Interception performance				
Success rate (%) ↑	**86.7**	63.3	6.7	33.3
Interception time (s) ↓	25.75 ± 7.10	35.40 ± 15.28	15.51 ± 0.57	37.37 ± 5.55
Minimum target range (m) ↓	0.55 ± 0.55	1.40 ± 0.33	13.26 ± 11.67	13.59 ± 29.76
Mean closing speed (m s^{-1}) ↑	4.25 ± 1.04	3.02 ± 1.21	5.81 ± 1.09	2.11 ± 1.66
Target visibility				
Target-visible time (%) ↑	99.11 ± 1.21	56.10 ± 9.97	58.81 ± 16.10	63.87 ± 14.84
RMS normalized image error ↓	0.82 ± 0.76	0.99 ± 0.05	1.24 ± 0.41	1.06 ± 0.23
Control execution				
RMS body-rate variation (rad s^{-2}) ↓	4.06 ± 2.88	9.58 ± 2.48	0.75 ± 0.22	2.87 ± 1.80

Mean ± standard deviation (SD) over all trials; interception time uses successful trials only. Image error is normalized by safe FOV radii. ↑/↓: higher/lower is preferred; closing speed must be assessed jointly with success and visibility.

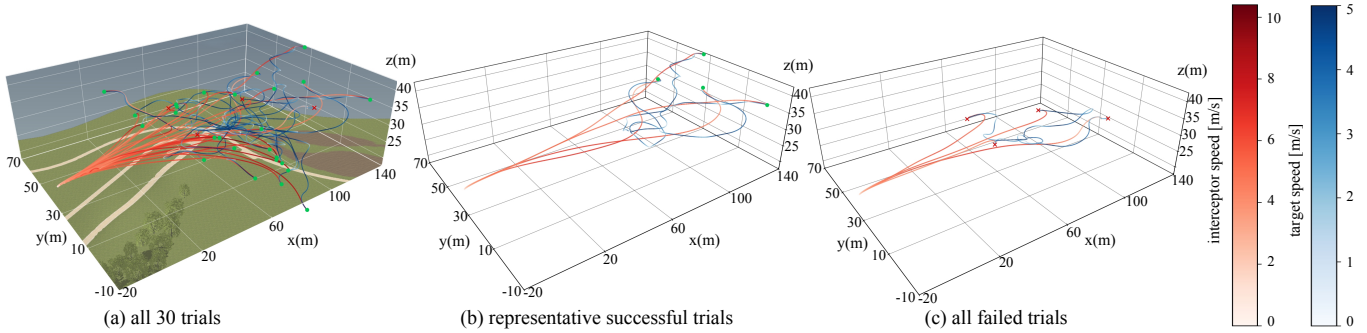


Fig. 7. ISB-NMPC simulation trajectories: (a) all 30 trials, (b) selected successful trials, and (c) all failed trials. Colors indicate interceptor and target speeds; green markers denote interception and red crosses denote failure.

Interception Algorithm Comparison. Baselines are planar-sector LOS (PS-LOS) guidance [9], LOS image-based visual servoing (LOS-IBVS), and backstepping [7]. LOS-IBVS combines proportional bearing feedback with forward-speed and yaw-rate commands [4]. All methods share target states, trajectories, bearings, and initial conditions (Table I and Fig. 8).

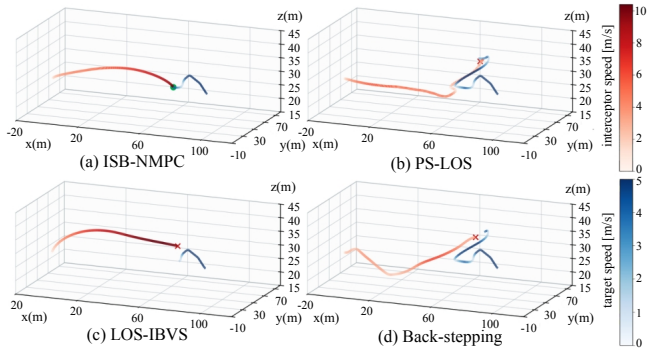


Fig. 8. Paired simulation trajectories: (a) ISB-NMPC, (b) PS-LOS, (c) LOS-IBVS, and (d) backstepping. All methods share target motion and initial geometry. Colors indicate speed; green markers denote interception and red crosses denote failure.

Figure 7 summarizes the ISB-NMPC simulation trajectories. The joint criteria in Sec. II-C are reported in Table I: ISB-NMPC succeeds in 26/30 trials (86.7%) with 99.11% mean target-visible time, and is faster than PS-LOS and backstepping among successful trials. LOS-IBVS closes faster but succeeds only twice and frequently loses visibility, illustrating why closing speed alone is insufficient. For the paired seed in Fig. 8, ISB-NMPC is the only successful controller while retaining aggressive closing. Its four failures coincide with sudden lateral target maneuvers, but the logs do not isolate the effects of prediction, control response, and active constraints.

B. Experimental Setup and Flight Validation

1) Experimental Platform: The 1.3 kg, 7-inch interceptor uses an IMX577 camera (2592×1944, 30 Hz, 2.8 mm focal length, $82^\circ \times 66^\circ$ calibrated FOV) and a 50 Hz bearing filter. PX4 EKF supplies ego-state; no target position or velocity is provided. Bearing-only prediction uses zero target velocity and nominal range equal to the configured maximum interception speed times the horizon duration. A separate quadrotor tows the moving target (Fig. 1b).

2) Experiment 1: Long-Range Interception of a Quasi-Static Small Target: Detection progressed from REACQUIRE-HR through TRACK-HR to TRACK-LR as the apparent target size increased, while interceptor speed reached 17 m/s (Fig. 1a).

3) Experiment 2: Pursuit–Evasion Interception of a Dynamic Target: The target followed an undulating path at approximately $5\,\mathrm{m\,s^{-1}}$ (Fig. 1b). Figure 9 compares online detections and DC-$\mathbb{S}^2$EKF estimates with offline detections at acquisition timestamps. The filter follows the offline trend, particularly in the vibration-sensitive v direction, and supplies bearings between detections. This comparison is qualitative because offline detections are not independent ground truth.

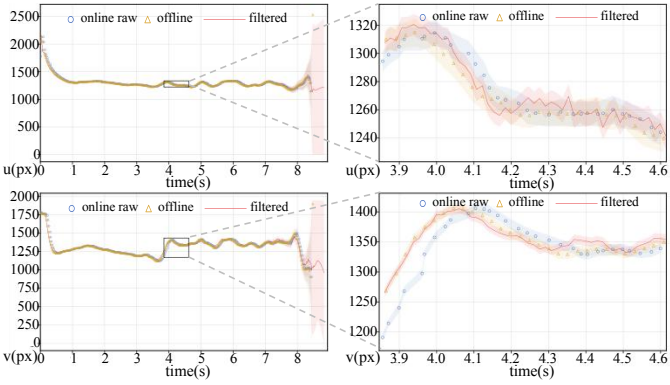


Fig. 9. Target image coordinates from online detections, offline results, and filtered estimates: u (top) and v (bottom). Right panels enlarge the highlighted intervals.

Figure 10 complements the joint performance criteria with onboard observations: visual cost dominates near the image boundary, while LOS velocity-alignment cost grows near encounter. Speed reaches approximately 14 m/s within 4 s and is largely sustained. These observations are consistent with the modeled coupling, not a causal test or proof of constraint satisfaction.

VI. Discussion and Conclusion

The paired simulations show that closing speed should be coordinated with visibility rather than maximized alone: ISB-NMPC achieved **86.7% success** and **99.11% mean target-visible time**, whereas LOS-IBVS was faster when successful

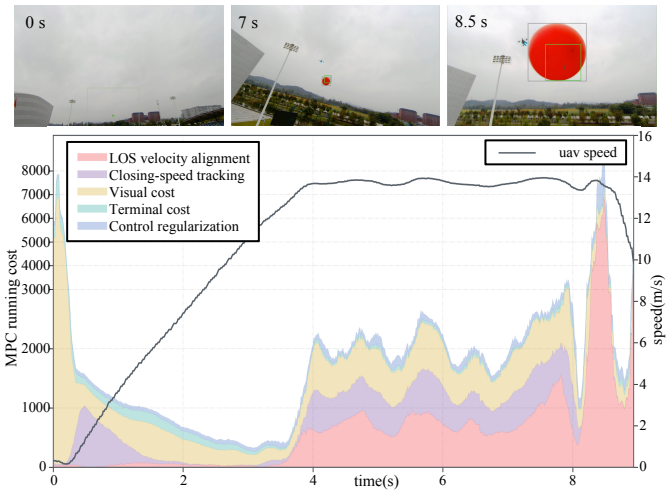


Fig. 10. Flight interception: onboard images at 0, 7, and 8.5 s (top); stacked MPC cost components and interceptor speed (right axis) below.

but succeeded only twice. Jointly optimizing desired closing speed and vehicle motion therefore provides a mechanism for allocating shared attitude authority between approach and visual retention, although the comparison does not isolate individual cost terms.

Simulations with ground-truth target inputs isolate controller behavior, whereas bearing-only outdoor flights demonstrate the integrated perception–control pipeline. Long-range acquisition and maneuvering-target pursuit complement rather than replace the paired comparison; agreement with offline detections is qualitative, not independent bearing-accuracy evidence.

Flight prediction assumes nominal range and zero target velocity, which can limit prediction during strong maneuvers. The four simulation failures and long occlusions motivate better lateral-motion prediction and active-constraint analysis. Neither recursive feasibility nor capture against arbitrary target motion is guaranteed, and the typical 10 ms solve time is not a worst-case bound. Future work should address range sensing, maneuver prediction, uncertainty-aware reacquisition, and component ablations.

In summary, ISB-NMPC connects LOS-aligned closing objectives to spherical visual prediction under quadrotor dynamics. The evaluated scenarios support its utility, while prediction uncertainty and sustained visibility remain key deployment boundaries.

REFERENCES

[1] J. Li, Z. Ning, S. He, C.-H. Lee, and S. Zhao, "Three-dimensional bearing-only target following via observability-enhanced helical guidance," *IEEE Transactions on Robotics*, vol. 39, no. 2, pp. 1509–1526, 2023.

[2] M. Pliska, M. Vrba, T. Báča, and M. Saska, "Towards safe mid-air drone interception: Strategies for tracking & capture," *IEEE Robotics and Automation Letters*, vol. 9, no. 10, pp. 8810–8817, Oct. 2024.

[3] O. Bourquardez, R. Mahony, N. Guenard, F. Chaumette, T. Hamel, and L. Eck, "Image-based visual servo control of the translation kinematics of a quadrotor aerial vehicle," *IEEE Transactions on Robotics*, vol. 25, no. 3, pp. 743–749, 2009.

[4] F. Chaumette and S. Hutchinson, "Visual servo control. I. Basic approaches," *IEEE Robotics and Automation Magazine*, vol. 13, no. 4, pp. 82–90, Dec. 2006.

[5] F. Xu, H. Wang, and H. Gao, "What matters in constructing a visual servoing scheme: A review of key issues and solutions," *IEEE/ASME Transactions on Mechatronics*, vol. 31, no. 2, pp. 1509–1523, 2026.

[6] D. Zheng, H. Wang, J. Wang, X. Zhang, and W. Chen, "Toward visibility guaranteed visual servoing control of quadrotor UAVs," *IEEE/ASME Transactions on Mechatronics*, vol. 24, no. 3, pp. 1087–1095, 2019.

[7] K. Yang, C. Bai, Z. She, and Q. Quan, "High-Speed Interception Multicopter Control by Image-based Visual Servoing," *IEEE Transactions on Control Systems Technology*, vol. 33, no. 1, pp. 119–135, Jan. 2025.

[8] K. Yang, C. Bai, and Q. Quan, "Line-of-Sight-Constrained Multicopter Interceptability," *Journal of Guidance, Control, and Dynamics*, vol. 48, no. 4, pp. 951–960, Apr. 2025.

[9] L. Liu, K. Yang, H. Zou, C. Min, S. Lv, S. Wang, and Q. Quan, "Planar-sector LOS guidance for interception of agile targets with lifting-wing quadcopters," in *2026 IEEE International Conference on Robotics and Automation (ICRA)*, 2026, pp. 15 780–15 787.

[10] Y. Zhang, K. Yang, Y. He, and Q. Quan, "Interceptability of vision-guided UAVs: A minimax approach to LOS-constrained pursuit-evasion dynamics," *IFAC-PapersOnLine*, vol. 59, no. 20, pp. 998–1003, 2025.

[11] K. Yang, L. Liu, C. Bai, and Q. Quan, "A unified and persistent interception control of multicopters with strapdown monocular camera," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 62, pp. 1244–1253, 2026.

[12] D. Falanga, P. Foehn, P. Lu, and D. Scaramuzza, "PAMPC: Perception-aware model predictive control for quadrotors," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Madrid: IEEE, Oct. 2018, pp. 1–8.

[13] B. Penin, P. Robuffo Giordano, and F. Chaumette, "Vision-based reactive planning for aggressive target tracking while avoiding collisions and occlusions," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3725–3732, 2018.

[14] A. Dmytruk, G. Silano, D. Bicego, D. Bonilla Licea, and M. Saska, "A perception-aware NMPC for vision-based target tracking and collision avoidance with a multi-rotor UAV," in *2022 International Conference on Unmanned Aircraft Systems (ICUAS)*, 2022, pp. 1668–1673.

[15] K. Zhang, Y. Shi, and H. Sheng, "Robust nonlinear model predictive control based visual servoing of quadrotor UAVs," *IEEE/ASME Transactions on Mechatronics*, vol. 26, no. 2, pp. 700–708, 2021.

[16] C. Qin, Q. Yu, H. S. H. Go, and H. H.-T. Liu, "Perception-aware image-based visual servoing of aggressive quadrotor UAVs," *IEEE/ASME Transactions on Mechatronics*, vol. 28, no. 4, pp. 2020–2028, Aug. 2023.

[17] Y. Li, G. Lu, D. He, and F. Zhang, "Robocentric model-based visual servoing for quadrotor flights," *IEEE/ASME Transactions on Mechatronics*, vol. 28, no. 4, pp. 2155–2166, 2023.

[18] R. Srivastava, R. Lima, and K. Das, "Aerial interception of non-cooperative intruder using model predictive control," in *2022 American Control Conference (ACC)*, 2022, pp. 494–499.

[19] A. T. Çetin and E. Koyuncu, "Model predictive control-based guidance with impact angle constraints for visual quadrotor interception," in *2023 9th International Conference on Control, Decision and Information Technologies (CoDIT)*, 2023, pp. 1–6.

[20] T. Hamel and R. Mahony, "Visual servoing of an under-actuated dynamic rigid-body system: An image-based approach," *IEEE Transactions on Robotics and Automation*, vol. 18, no. 2, pp. 187–198, Apr. 2002.

[21] R. Mebarki, V. Lippiello, and B. Siciliano, "Nonlinear visual control of unmanned aerial vehicles in GPS-denied environments," *IEEE Transactions on Robotics*, vol. 31, no. 4, pp. 1004–1017, Aug. 2015.

[22] F. C. Akyon, S. O. Altinuc, and A. Temizel, "Slicing aided hyper inference and fine-tuning for small object detection," in *2022 IEEE International Conference on Image Processing (ICIP)*, 2022, pp. 966–970.

[23] Z. Guo, J. Yang, S. Li, and Z. Wang, "Robust visual landing control of quadrotor on a moving platform: A sampled-data approach with delayed output and disturbances," *IEEE Transactions on Control Systems Technology*, vol. 32, no. 6, pp. 2283–2297, 2024.

[24] Y. Zhang, K. Yang, L. Liu, Y. He, and Q. Quan, "Interceptability analysis under line-of-sight constraints for a class of hybrid UAVs," *Aerospace Science and Technology*, vol. 176, p. 112602, Sep. 2026.

[25] C. Hertzberg, R. Wagner, U. Frese, and L. Schröder, "Integrating generic sensor fusion algorithms with sound state representations through encapsulation of manifolds," *Information Fusion*, vol. 14, no. 1, pp. 57–77, 2013.

[26] B. Houska, H. J. Ferreau, and M. Diehl, "ACADO toolkit—An open-source framework for automatic control and dynamic optimization," *Optimal Control Applications and Methods*, vol. 32, no. 3, pp. 298–312, 2011.

[27] H. J. Ferreau, C. Kirches, A. Potschka, H. G. Bock, and M. Diehl, "qpOASES: a parametric active-set algorithm for quadratic programming," *Mathematical Programming Computation*, vol. 6, no. 4, pp. 327–363, 2014.



Li Hong received the B.S. degree in control science and engineering from Shen Yuan Honors College, Beihang University, Beijing, China, in 2026, where he is currently pursuing the Ph.D. degree in control science and engineering. His main research interests include UAV navigation, planning, and control.



Linkai Liu is currently pursuing the Ph.D. degree in Navigation, Guidance and Control at the School of Automation Science and Electrical Engineering, Beihang University (formerly Beijing University of Aeronautics and Astronautics), Beijing, China. He received the B.S. degree in Intelligent Flight Vehicle Technology from Beihang University in 2025. His main research interests include vision-based guidance and control, and autonomous aerial robotics.



Xiangyu Ren is currently pursuing the master's degree in control engineering with the School of Automation Science and Electrical Engineering, Beihang University, Beijing, China. His research interests include aerial robotics.



Shuli Lv received the B.S. degree in control science and engineering from Shen Yuan Honors College, Beihang University, Beijing, China, in 2022. He is currently pursuing the Ph.D. degree in control science and engineering with the School of Automation Science and Electrical Engineering, Beihang University. His main research interests include UAV navigation, planning, and control.



Chen Min received the B.S. degree in control science and engineering from Shen Yuan Honors College, Beihang University, Beijing, China, in 2025. He is currently working toward the Ph.D. degree in control science and engineering with Shen Yuan Honors College, Beihang University, Beijing, China. His main research interests include UAV navigation, planning, and control.



Chunjiang Wu received the B.S. degree in automation from North China Electric Power University, China, in 2023, and the M.S. degree in control science and engineering from the School of Automation Science and Electrical Engineering, Beihang University, Beijing, China, in 2026. He is currently pursuing the Ph.D. degree in low-altitude intelligent navigation technology at the Hangzhou International Innovation Institute of Beihang University, Hangzhou, China. His research interests include visual recognition and embodied intelligence.



Yujun Zhang is currently pursuing the Ph.D. degree in Control Science and Engineering at Shen Yuan Honors College, Beihang University, Beijing, China. She received the B.S. degree in Automation from Beihang University in 2026. Her research interests include pursuit-evasion games, multi-UAV cooperative planning, and autonomous aerial robotics.



Dingchen Zeng is currently pursuing a bachelor's degree in automation at the School of Automation Science and Electrical Engineering, Beihang University, Beijing, China. Current research interests include autonomous aerial robotics and embodied intelligence.



Quan Quan (Senior Member, IEEE) received the B.S. and Ph.D. degrees in control science and engineering from Beihang University, Beijing, China, in 2004 and 2010, respectively. Since 2022, he has been a Professor with Beihang University, where he is currently with the School of Automation Science and Electrical Engineering. His research interests include reliable flight control, swarm intelligence, vision-based navigation, and health assessment.