Grasping as Anchoring: Expanding Urban Perching with a Hybrid Gripper for Aerial Robots

Ziyin Han[1], Bihao Mo[1], Sheng Cheng[1], Rong Wang[1], Ziyi Wang[1], Junjie Gao[1], Hung Tien Pham[2], Van Anh Ho[2], and Naira Hovakimyan[1]

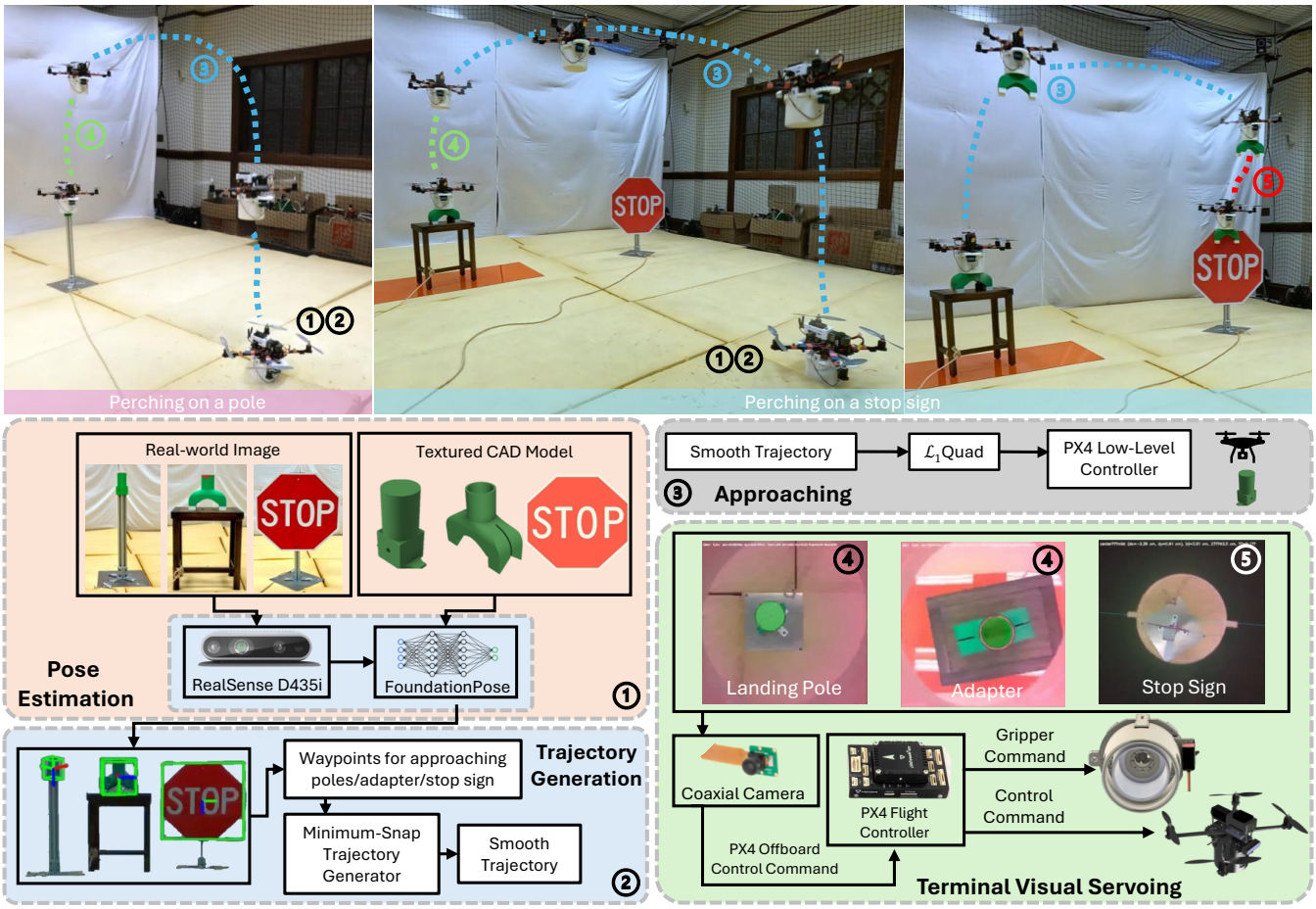

Fig. 1: Stacked images of the pole-perching and stop-sign-perching and an overview of our autonomy pipeline for perching.

***Abstract*— Aerial manipulation demands end-effectors that can support multiple functions such as grasping and perching. Yet, existing designs fall into distinct categories: mechanically compliant grippers provide adaptive grasping but lack structural stiffness for perching; stiff perching mechanisms are geometry-specific and unsuitable for object interaction. We present a hybrid aerial manipulation system that addresses this compliance-stiffness conflict by decoupling the anchoring load path from a compliant grasping interface. This system enables a novel aerial tool-use capability, where manipulable objects can serve as anchoring interfaces. By actively choosing the tools, the robot can expand the range of perchable sites to include previously non-perchable urban installations such as utility poles or traffic signs. We embed a coaxial eye-in-hand camera to guarantees zero-parallax observability for robust alignment during perching. An $\mathcal{L}_1$ adaptive controller is applied for robust flight against aerodynamic and payload introduced disturbances. Experiments demonstrate reliable grasping of various everyday objects, versatile perching on diverse landing poles, and urban installations.**

***Index Terms*— Mechanism Design, Aerial Manipulation, Perching, Visual Servoing**

I. Introduction

As aerial robots become increasingly involved in tasks requiring physical interactions such as infrastructure inspection and object manipulation, the need for versatile end-effectors has grown. Among these capabilities, grasping and perching are particularly critical: grasping enables transport and manipulation, whereas perching offers energy-efficient anchoring for long endurance.

Most prior aerial manipulation platforms specialize in only one of these functionalities. Perching-only end-effectors such as rigid talons [1], [2], vacuum cups [3], [4], microspines [5], and electroadhesion [6], can offer strong anchoring but exhibit limited adaptability for object interaction. In grasping-

only scenarios, existing solutions utilizing soft or rigid grippers [7]–[10] prioritize grasping performance but often lack the structural stiffness to support the weight of the vehicle. Fundamentally, mechanical compliance (required for conformable object engagement) and load-bearing stiffness (required for stable anchoring) are conflicting material and structural requirements. This conflict restricts the versatility of aerial manipulation systems in scenarios requiring both stable anchoring and flexible object interactions, such as infrastructure inspection and repair on utility poles or ad-hoc emergency communication relays during blackouts. In addition, the limited payload and power budgets of aerial platforms critically limit the possibility of introducing variable-stiffness mechanisms to the end-effector.

While a few aerial systems do integrate both grasping and perching functionalities, they typically demand task-specific perching interfaces subject to constrained target geometries that comply with the end-effector's shape. Avian-inspired claws are capable of both grasping and perching, but they excel only when a branch-like cylindrical support is present [1], [2]; microspine-and-hook-based designs rely on rough edges or cables [5], [11]; suction and electroadhesion approaches require smooth and conductive surfaces [3], [6]. This reliance on existing environmental features restricts the operational envelope of aerial robots in urban settings. Many common urban installations (shown in Fig. 2), such as lamp posts with spherical finials, fence tops, small maintenance ledges, or thin steel edges of road signs, lack these specific properties, making them non-perchable for aerial robots with conventional end-effectors.

In this paper, we present a hybrid aerial manipulation system for aerial robots that synthesizes grasping and perching capabilities into a unified mechanical structure. To overcome the conflict between contact compliance and anchoring stiffness, we leverage the rotational membrane buckling of the ROSE gripper [12] for geometry-agnostic grasping, and introduce a mechanically decoupled rigid exoskeleton to handle perching loads. While the soft membrane provides conformal contact forces for secure engagement, the rigid shell passes the load directly into the airframe without compromising adaptive grasping.

This synergy permits the aerial robot not only to grasp diverse objects or perch on poles, but also to execute multi-stage aerial tool use by placing and utilizing structural interfaces to adapt to non-perchable environments. We showcase this concept of "grasping as anchoring" with a modulable perching tool that allows the aerial robot to perch on the edge of a stop sign, which is infeasible for conventional landing gears. This concept of creating artificial features enables a broad range of urban installations to become viable roosts by introducing the graspable tool that can adapt to the environment.

To overcome the terminal blind-zone and occlusion issues prevalent in aerial grasping [13], we propose a perception-action coupled design. While a forward-view camera facilities coarse initial approaching, the terminal perching phase is governed by a downward coaxial vision system. We enhance the ROSE's actuation into a hollow-bore configuration, embedding an RGB camera directly along the axis of rotation. By co-locating sensing with the contact interface, we achieve a zero-parallax between the optical and manipulation axes, and maintains a well-conditioned image Jacobian up to the exact moment of touchdown. To enhance the aerial robot's robustness in flight and during physical interactions, we incorporate the $\mathcal{L}_1$Quad [14] control architecture, ensuring uniform tracking performance across heterogeneous transportation tasks.

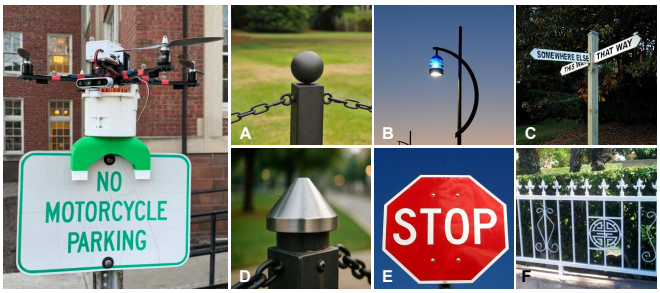

Fig. 2: Examples of perchable spots in urban environments.

The contributions of this paper are summarized as follows: From the mechanics perspective, we design a lightweight hybrid gripper that addresses the compliance-stiffness conflict via mechanically decoupled load paths, providing conformal grasping via a soft membrane and load-bearing ability via a rigid exoskeleton. From the perception and control perspectives, we introduce a zero-parallax coaxial visual servoing framework that maintains optimal target observability through the terminal descent, integrated with $\mathcal{L}_1$Quad control architecture to reject uncertainties induced from physical interactions.

II. Related Work

A. Perching mechanisms

Existing research has explored several energy-efficient perching mechanisms for aerial robots, including avian-inspired latching claws [1], [2] for tree branches, microspine arrays [5] for rough surfaces, passive hooks [11] for cables and power lines, and suction [3], [4] or electroadhesion [6] for smooth or conductive surfaces. While effective in their niches, these mechanisms are typically geometry- or material-specific: claws are suited to cylindrical edges; microspines need irregular textures with surface asperities to interlock; suction favors airtight patches; and electroadhesion requires certain electrical properties. These mechanisms often demand precise approach conditions and can hardly support general object manipulation once perched. As a result, their application in unstructured urban environments is constrained. In contrast to these geometry-dependent solutions, our work decouples the end-effector from the target's geometry constraint.

B. Grippers for aerial robots

Soft grippers are attractive for aerial manipulation because their passive compliance conforms to irregular shapes and is robust to pose uncertainty. Representative designs include

tendon-driven silicone hands [8], pneumatically actuated systems [9], and Fin-Ray grippers [15]. However, these solutions prioritize grasping at the expense of losing structural rigidity. When such soft grippers are equipped on an aerial vehicle, this drawback forces the system to continue propelling during perching, significantly reducing mission endurance due to sustained energy expenditure. In contrast, rigid grippers offer higher load capacity and precise contacts, yet require accurate alignment and adapt poorly to shape variability. Recent hybrid attempts, such as stiffening sheaths [16] or granular jamming layers [10], partially address this issue by modulating material properties. However, they either introduce mechanical and actuation complexity or cannot sustain the full weight of the aerial robot during perching.

In summary, a fundamental challenge remains as how to achieve high structural stiffness without sacrificing the compliance of soft end-effectors for aerial interaction and manipulation. Our work addresses this challenge by mechanically decoupling the rigid anchoring load path from the compliant contact mechanism. This system provides perching-grade stiffness and grasping-grade compliance simultaneously, bypassing the common trade-offs in variable-stiffness schemes [17]. Additionally, we enable an aerial robot to not only perch on existing structures but also manipulate tools to create artificial perching features, expanding the robot's urban operational envelope.

C. Visual servoing for aerial manipulation

Autonomous perching in unprepared urban environments precludes the use of cooperative fiducials (e.g., AprilTags), necessitating marker-less visual servoing. Conventional eye-to-hand setups often suffers from the loss of target at close range [13]. Moreover, the gripper can occlude the camera, requiring masking or imposing geometric constraints on the gripper's motion [18]. Furthermore, calibration drift and parallax errors scale exponentially as distance decreases, a problem exacerbated by the deformation of soft grippers [19]. Although standard eye-in-hand configurations mitigate these issues by centering the target [13], [20] and fixing the kinematic transform [21], mounting cameras externally on soft end-effectors remains challenging due to shifting load paths and unpredictable membrane buckling.

We push this paradigm further by embedding a camera directly inside the hollow actuation bore of our hybrid gripper. This coaxial colocation enforces a strict zero-parallax geometric constraint, guaranteeing continuous target observability and maintaining a well-conditioned image Jacobian right up to the moment of physical contact.

III. Hybrid Gripper Design

To synthesize grasping compliance with anchoring stiffness, we propose a mechanically decoupled hybrid architecture. The compliant contact interface leverages the principles of the ROSE gripper [12], which consists of two concentric "inner" and "outer" elastomer sleeves joined near the rim. A single rotary input twists the inner sleeve relative to the outer, inducing torsional buckling that forms inward wrinkles and draws material radially toward the axis.

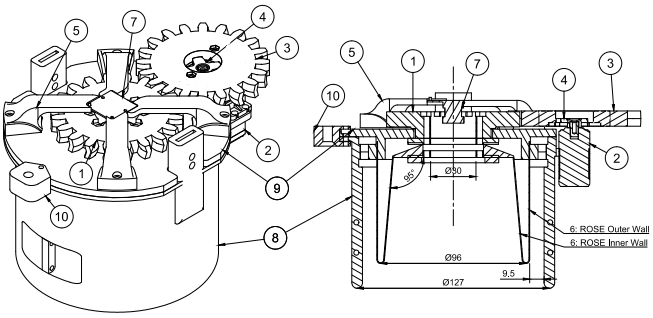

Fig. 3: Engineering drawing of coaxial hybrid gripper

TABLE I: Parts list (numbers match the engineering drawing).

#	Part name	#	Part name
1	Ring following gear	6	ROSE soft body
2	High-torque servo	7	Arducam IMX219 camera
3	Pinion driving gear	8	Rigid exoskeleton
4	Servo horn / coupler	9	Top retainer
5	Downward camera mount	10	Forward camera mount

This membrane-buckling kinematics produces clamping at the rim through high area-normal contact and caging when rotation continues, passively enveloping objects without complex finger planning. While this single-degree-of-freedom actuation simplifies control, elastomer structures deform under the static weight and dynamic moments of the vehicle during manipulation. To resolve this, our design encapsulates the soft core within a load-bearing rigid exoskeleton, which will be detailed below. We will also introduce the zero-parallax perception system that aligns with the manipulation axis.

A. Exoskeleton Design and Decoupled Load Paths

To address the stiffness-compliance conflict, we use a rigid exoskeleton that encapsulates the soft membrane, establishing decoupled load paths. The physical interactions between the vehicle and the environment can be modeled using the reaction wrench at the end-effector, $\mathbf{W}_{\text{reaction}} = [\mathbf{F}_{\text{reaction}}, \boldsymbol{\tau}_{\text{reaction}}] \in \mathbb{R}^6$. When perched, the end-effector must counteract the external wrench $\mathbf{W}_{\text{ext}}$ comprising the static weight of the quadrotor and dynamic environmental disturbances $\boldsymbol{\tau}_{\text{dist}}$. With a pure soft gripper design, the entire load is borne by the hyper-elastic membrane ($\mathbf{W}_{\text{reaction}} = \mathbf{W}_{\text{membrane}}$), leading to compliance-induced wobble and potential failure under shear loads.

In our hybrid system, the total reaction wrench is structurally distributed:

$$\mathbf{W}_{\text{reaction}} = \mathbf{W}_{\text{shell}} + \mathbf{W}_{\text{membrane}} = -\mathbf{W}_{\text{ext}}. \tag{1}$$

Let $\mathbf{K}_{\text{shell}}$ and $\mathbf{K}_{\text{membrane}}$ denote the spatial stiffness matrices of the rigid exoskeleton and the soft core, respectively. By material and geometric design, the shell exhibits orders of magnitude higher stiffness in the axial and bending directions, e.g., $||\mathbf{K}_{\text{shell}}|| \gg ||\mathbf{K}_{\text{membrane}}||$. When the shell rim contacts the perching fixture, it establishes a parallel

TABLE II: Coaxial hybrid gripper specifications.

Parameter	Value	Notes
Outer diameter	127 mm	shell OD
Clear aperture	30 mm	camera bore ID
ROSE diameter	96 mm	at rim
Soft cavity (max object ∅)	84 mm	at inner wall
Height (assembly)	127 mm	incl. servo
Mass (total)	493 g	incl. camera
Servo	DS3235 Servo	35kg high torque
Camera	Arducam IMX219	8 MP, RGB

mechanical circuit. The load distribution follows the stiffness ratio, effectively yielding:

$$\mathbf{W}_{\text{shell}} \approx -\mathbf{W}_{\text{ext}}, \quad \mathbf{W}_{\text{membrane}} \approx \mathbf{0}. \tag{2}$$

During perching, the exoskeleton's rim and internal ribs bypass the elastomer entirely, passing the load directly into the vehicle's rigid airframe. Concurrently, the internal soft membrane degrades to solely provide conformal target engagement. This structure decoupling physically isolates the hyper-elastic material from the vehicle's structural loads, eliminating the load-induced wobble characteristic of pure soft grippers. Consequently, the system achieves passive mechanical interlocking for energy-neutral perching. Figure 3 and Table I show the system layout and list of parts, and Table II summarizes the major kinematic and mechanical design parameters.

B. Zero-Parallax Perception and Peripheral Actuation

General soft grippers suffer from self-occlusion and kinematic calibration drift when outfitted with external cameras [17]. To guarantee continuous target observability during the terminal descent, we redesign the membrane-buckling mechanism to accommodate a central, unobstructed optical through-bore. By removing the central torque-transmission hub typical of prior designs, we relocate the actuation to the perimeter. A laterally mounted servo drives a spur-gear train, which engages the structural boundaries of the inner and outer elastomer walls to induce the required relative rotation.

This off-axis actuation decouples the optical line from the mechanical transmission while preserving the membrane-buckling kinematics and grasp strength. An RGB camera is mounted on vibration-isolated standoffs in side of the gripper, aligned with the axis of manipulation. This coaxial configuration provides a zero-parallax feature, ensuring the perching target remains observable right up to the moment of contact. We experimentally illustrate the necessity of coaxial eye-in-hand design via a comparison with an eye-on-body baseline. Shown in Fig. 4, eye-on-body systems suffer from occlusion during the final 20 cm of descent. This "blind zone" leads to a non-ideal open-loop contact. In contrast, our coaxial design guarantees continuous target observability and a closed-loop contact. Finally, the entire hybrid module is aligned with the vehicle's center of gravity, ensuring that contact forces and payload generate minimal unmodeled destabilizing moments.

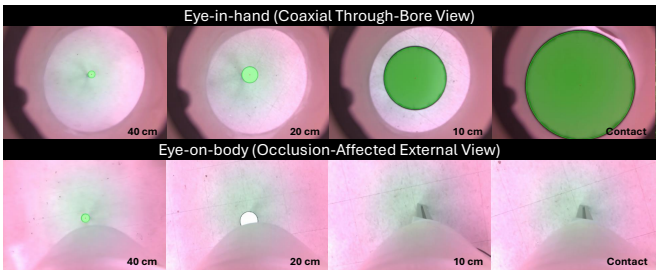

Fig. 4: Visual sequence of terminal approach using the eye-in-hand configuration (proposed) and eye-on-body (baseline) configuration. Eye-in-hand maintains the target in the center of the field of view until contact. The eye-on-body configuration suffers from structural occlusion by the gripper body when the distance to the target drops below 20 cm, resulting in a terminal blind zone.

C. Aerial Tool Design: Modulable Anchoring Interfaces

We turn the mission of perching on incompatible structures to an active aerial tool-use task; these tools can adapt to various environmental features that are previously non-perchable. To diminish the passive reliance on existing environmental geometry, we introduce modular anchoring interfaces.

The interface acts as a mechanical connector. Its upper half is geometrically standardized to conform perfectly to the hybrid gripper's inner wall, guaranteeing a secure, tilt-free grasp. The task-specific lower half acts as an adapter. For example, to perch on a 2 mm stop sign's top edge (Fig. 6B), the lower half takes the form of a clip. By switching only the lower-half's geometry, the robot can utilize a library of tools to create viable roosts on previously non-perchable urban installations. Furthermore, the interface features a central optical pathway that preserves the coaxial camera's line-of-sight, preserving the zero-parallax feature even when the robot is transporting the tool (Fig. 6C). Since the ROSE membrane induces uncontrolled torsional rotation during closure, we make a fiducial notch on the adapter and use the coaxial camera to track this notch for visually registering the tool's yaw deflection. A feed-forward yaw controller subsequently compensates for this deflection before the terminal descent (detailed in Section IV-B).

IV. Autonomy Pipeline and Perception

To enable autonomous perching in unstructured environments, we propose a perception-action pipeline (shown in Fig. 1) that seamlessly transitions from global coarse routing to local zero-parallax alignment and subsequent manipulation. The system operates across four primary stages:

1) *Target Pose Estimation:* Using images from the forward-view camera ($\mathcal{F}_{c_F}$), we utilize FoundationPose [22] to estimate the coarse 6D pose of the target (pole or tool) directly from its textured CAD model. Standard $\mathbb{SE}(3)$ hand-eye kinematics map this object-centric pose into the world frame ($\mathcal{F}_w$).
2) *Trajectory Planning:* The world-frame target position seeds a minimum-snap trajectory generator [23], yielding a smooth trajectory to a hover above the target.

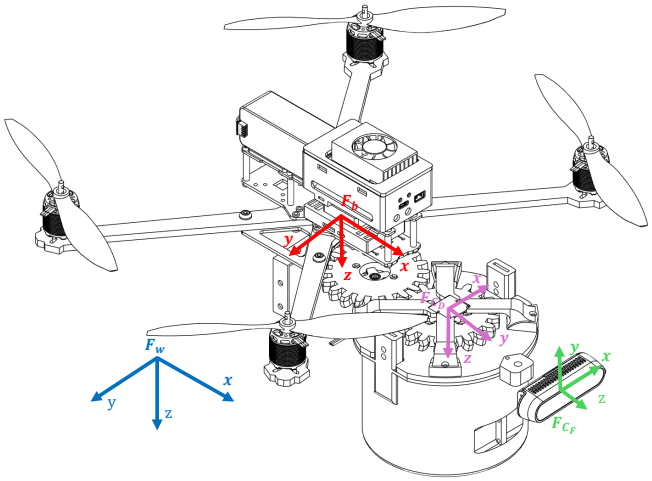

Fig. 5: The exploded view of the vehicle and associated reference frames.

3) *Approaching Flight:* We use a geometric controller [24] on $\mathbb{SE}(3)$ to track the minimum-snap trajectories that send the aerial robot to a location above a perching target. The controller send angular rate and body thrust commands to the low-level controller. To improve flight robustness against disturbances (e.g., ground effect, payload with unknown mass, and misalignment during contact), we use $\mathcal{L}_1$Quad [14] as an adaptive augmentation for the rate-loop control.
4) *Terminal Visual Servoing & Perching:* Upon reaching the hover waypoint, control transitions to a visual servoing loop driven by the downward coaxial camera ($\mathcal{F}_{c_D}$). The quadrotor descends vertically while maintaining zero-parallax lateral alignment. Upon contact, the ROSE gripper closes to secure the grasp. For the multi-stage stop-sign perching task, the quadrotor subsequently takes off with the grasped adapter, replans an approach to the stop sign, and repeats the terminal servoing to anchor the interface.

A. Target Detection for Terminal Servoing

We develop two vision front-ends to extract metric positional references to aid terminal visual servoing. Both treat the target as a single point feature in image processing and produce a metric distance between the target center and the camera, as well as yaw angles in the stop-sign perching case.

(i) Pole and Adapter Perching: A shape-based contour detector extracts the target's pixel center (u, v) and apparent radius r (in pixels). Given the camera focal length f and known physical radius R, the depth Z_{center} and metric lateral offsets $(\delta x, \delta y)$ relative to the principal point $(\bar{u}, \bar{v})$ are directly computed as:

$$Z_{\text{center}} = \frac{fR}{r}, \quad \delta x = \frac{Z_{\text{center}}}{f}(u - \bar{u}), \quad \delta y = \frac{Z_{\text{center}}}{f}(v - \bar{v}). \tag{3}$$

(ii) Stop-Sign Perching: Aligning the adapter with a thin edge requires additional yaw estimation (Fig. 6). First, template matching isolates the adapter's inner contour to

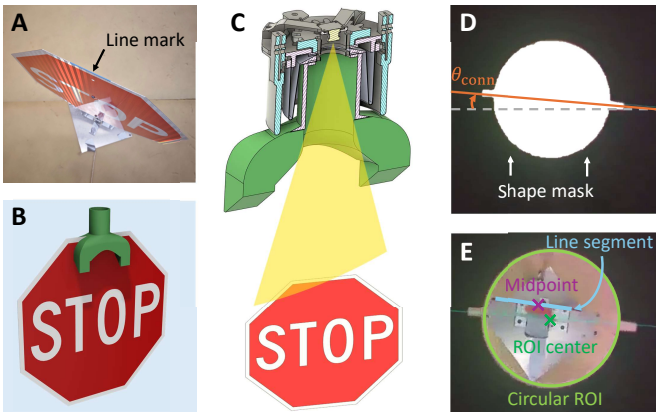

Fig. 6: Illustration of the stop-sign connector and image processing. (A) Stop sign with a line mark on top; (B) Illustration of the connector attachment to the sign; (C) Illustration of camera's view of the stop sign from above (adapter and camera-embedded gripper have been enlarged for ease of view); (D) Shape mask for image processing; (E) Extracted features for alignment.

determine its current proprioceptive orientation θ_{conn} and establish a circular region of interest (ROI). Second, within this ROI, color masking and RANSAC [25] extract a line segment corresponding to a known physical marker of length L. Let the detected pixel length be l_{px} and its midpoint be (u, v). The depth is approximated as $Z_{\text{line}} \approx fL/l_{\text{px}}$. The lateral offsets $(\delta x, \delta y)$ are computed via (3) by substituting Z_{line} for Z_{center} and the ROI center for $(\bar{u}, \bar{v})$. Finally, the line segment's angle θ provides the absolute yaw reference, enabling the flight controller to actively nullify the yaw error $e_\theta = \theta - \theta_{\text{conn}}$ prior to contact.

B. Visual servoing control

Given the frame transformations, we can map a position $\mathbf{p}_{c_D} \in \mathbb{R}^3$ in the camera frame $\{c_D\}$ to the world frame by

$$\mathbf{p}_w = R_{wb}\big(R_{bc_D}\,\mathbf{p}_{c_D} + \mathbf{t}_{bc_D}\big) + \mathbf{p}_{wb}. \tag{4}$$

The positional offset $\Delta\mathbf{p}_{c_D} = [\delta x, \delta y, \Delta z]^\top$ in the camera frame can be obtained by the front-ends in Section IV-A, where Δz is the vertical distance Z_{center} or Z_{line} gated by a hysteresis band (to avoid chatter in the terminal descent). The corresponding world-frame offset is

$$\Delta\mathbf{p}_w = R_{wb}R_{bc_D}\Delta\mathbf{p}_{c_D}, \tag{5}$$

which is used to update the position setpoint of the aerial robot by

$$\tilde{\mathbf{p}}_w = \mathbf{p}_w + \alpha\,\text{clip}\left(\Delta\mathbf{p}_w, \Delta\mathbf{p}_{\max}\right), \tag{6}$$

where $\mathbf{p}_w$ and $\tilde{\mathbf{p}}_w$ are the current position and position setpoint, respectively; $\alpha > 0$ is a tunable weight; $\Delta\mathbf{p}_{\max}$ is the positional offset cap.

For the stop sign perching, yaw control is required to align the adapter with the stop sign. The angle between the adapter's orientation θ_{conn} and the stop sign's orientation θ

is the yaw error

$$e_\theta = \text{wrap}\left(\theta - \theta_{\text{conn}}\right),$$

where $\text{wrap}(\cdot)$ maps the angle difference to the interval of $(-\frac{\pi}{2}, \frac{\pi}{2}]$. With a tolerance τ_θ and a maximum yaw rotation $\Delta\theta_{\max}$, the yaw increment is

$$\Delta\theta = \begin{cases} 0, & |e_\theta| < \tau_\theta \\ \text{clip}\left(e_\theta, \Delta\theta_{\max}\right), & \text{otherwise} \end{cases}, \tag{7}$$

which determines the yaw setpoint $\tilde{\theta}$ by $\tilde{\theta} = \theta + \Delta\theta$ for θ being the current yaw measurement.

V. Experimental Results

We use a custom-built quadrotor equipped with a Pixhawk 6X mini (running PX4 firmware) as the flight controller and an NVIDIA Jetson Orin as the companion computer. A forward-view RealSense D435i provides coarse global perception, while an Arducam IMX219 is embedded coaxially for terminal visual servoing. The hybrid end-effector is fabricated using 3D-printed PETG for the rigid exoskeleton and cast silicone for the compliant membrane. The FoundationPose module runs at 20 Hz, and the visual servoing loop runs between 10–30 Hz. We first validate the system's tolerance to eccentric loads via the off-center torque test. Subsequently, we evaluate the system across three operational scenarios: autonomous perching on heterogeneous poles, dynamic grasping of diverse payloads, and multi-stage aerial tool use (perch-while-grasping via anchoring interfaces).

A. System Validation: Decoupled Mechanics

When an aerial robot finishes perching, it disarms the actuators and then loses active attitude control. In this case, any center-of-mass offset, payload asymmetry, or environmental wind gust translates into a bending moment acting directly on the end-effector. To validate that our hybrid system can withstand these loads via its decoupled load path, we benchmark it against two baselines: the original *ROSE Gripper* and a conventional *three-finger Fin-Ray soft gripper*. To explicitly quantify the system's rotational stiffness, we designed a controlled off-center torque experiment. Each gripper is commanded to perch securely on a vertical pole. To implement the destabilizing bending moment (τ_{dist}), we attach a beam to the gripper's mounting interface on one end and suspend weights at the other end. By increasing the mass at a known lever arm distance, we apply a precise bending moment to the perched system. We quantify the system's bending stiffness by recording the tilting angle θ_{tilt} caused by the load torque τ_{app} applied to the perched airframe. The τ_{app}–θ_{tilt} curve (shown in Fig. 7) demonstrates the *mechanical interlocking* effect, with the key metric being the maximum load torque before mechanical failure or slip. Pure soft gripper (*three-finger Fin-Ray soft gripper* and *ROSE Gripper*) suffer from compliance-induced failure, exhibiting deformation and slip at small torques (< 0.3 N·m). In contrast, our Hybrid Gripper leverages its rigid exoskeleton as the decoupled load path, maintaining a minimal θ_{tilt} and sustaining maximum load torque over 4.0 N·m. With the mechanical robustness of the Hybrid Gripper established, the following sections evaluate the integrated system's performance across three paradigms: autonomous perching on heterogeneous poles, dynamic payload grasping, and multi-stage aerial tool use (perch-while-grasping).

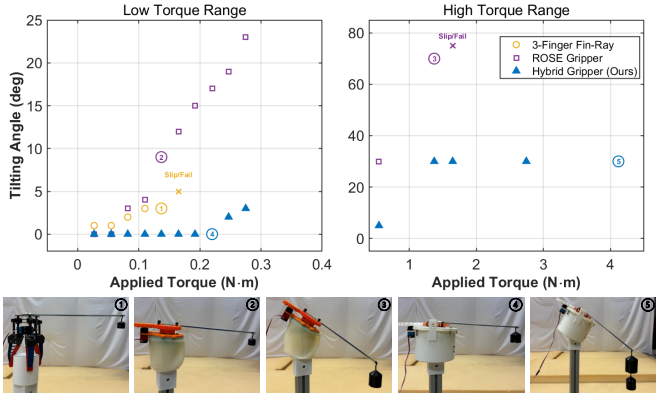

Fig. 7: Experimental validation of load-introduced disturbance rejection. (a) Pure soft grippers (Fin-Ray and ROSE) fail at negligible applied torques (< 0.3 N·m) due to inherent membrane compliance. (b) The proposed Hybrid Gripper achieves mechanical interlocking via its exoskeleton, sustaining an order-of-magnitude higher torque (> 4.0 N·m) with minimal tilting deflection.

B. Autonomous Pole Perching

We evaluate the system's perching adaptability across four types of target geometries: (i) cuboid, side length 48 mm; (ii) cylinder, diameter 55 mm; (iii) sphere, diameter 70 mm; and (iv) truncated cone, top diameter 50 mm (shown in Fig. 8). In addition, we conduct a size sweep on cylindrical heads with diameters of 35, 45, 55, 65, and 75 mm to study the influence of target scale on perching performance. A complete flight sequence is illustrated in Fig. 1. A trial is considered successful if the vehicle remains securely attached without catastrophic slip and can subsequently re-arm its rotors to lift off. Table III summarizes the nine perching experiments, reporting the post-perching lateral offset (e_{xy}) between the pole and gripper center, as well as the elapsed alignment time (t_{center}) from approaching completion to the beginning of descent.

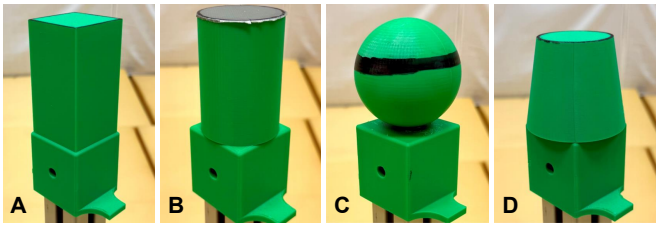

Fig. 8: Landing-pole head geometries used in perching experiments.

With visual servoing enabled, the platform perches reliably on heterogeneous pole heads. Perching performance

TABLE III: Perching performance across various pole geometries and sizes

Geometry	Success rate (%)	Total trials	e_{xy} (mm)	t_{center} (s)	β_{tilt} (deg)
Sphere	100.0	11	27.9 ± 36.3	17.49 ± 1.79	18.5
Cuboid	90.9	11	18.4 ± 22.0	17.53 ± 2.20	3.7
Trunc. Cone	100.0	11	17.6 ± 25.9	16.41 ± 1.48	5.1
Cylinder-35	100.0	11	96.1 ± 30.8	19.65 ± 1.74	47.6
Cylinder-45	100.0	11	77.9 ± 30.1	15.94 ± 1.94	32.6
Cylinder-55	90.9	11	24.2 ± 29.6	17.03 ± 1.38	15.1
Cylinder-65	72.7	11	36.6 ± 44.0	17.81 ± 2.16	0.0
Cylinder-75	54.5	11	70.2 ± 57.2	17.18 ± 1.77	0.0
No VS	10.0	20	124.8 ± 8.7	N.A	N.A

depends not only on target shape but also on the diameter D of the pole's head relative to the ROSE inner diameter (D_{in}=84 mm). We report a tilting angle β_{tilt}, defined as the angle between the quadrotor's body z-axis and the world z-axis after touchdown,

$$\beta_{\text{tilt}} = \arccos\left(\mathbf{z}_b^\top \mathbf{z}_w\right). \tag{8}$$

Tilting occurs when $D \ll D_{\text{in}}$: The shell clamps a narrow region of the pole head, and the vertical reaction line is offset from the vehicle center of gravity, creating rolling and pitching moments. The vehicle then settles to an equilibrium with a nonzero β_{tilt} when the rim of the rigid shell bears against the head edge (the moments balance). Conversely, when D approaches D_{in}, alignment tolerance becomes tight: Small lateral errors lead to early contact between the shell and pole head, which reduces success rate albeit yielding near-zero tilting angle when perched successfully. Intermediate diameters (around 55 mm) typically balance trade-off between success rate and moderate tilting angle. Furthermore, target geometry illustrates the complementary nature of our hardware-perception co-design. For instance, beveled geometries like truncated cones introduce visual ambiguity (e.g., overlapping depth edges confusing the contour detector). Their chamfered profile provides a passive mechanical lead-in. During the final millimeters of descent, this structural funneling dominates, physically self-centering the gripper to compensate for terminal visual estimation noise (Table III).

Despite the geometry-specific interactions, the distribution of t_{center} remains consistent across all successful trials, suggesting that most of the terminal time is spent in step-limited lateral corrections and the descent hysteresis. Ablating terminal visual servoing yields only a 10% success rate with a much larger lateral error e_{xy}, demonstrating the necessity of the vision-based servoing loop for centimeter-level alignment and robust perching. Lastly, we note that t_{center} concentrates in 16–18 s, reflecting conservative descent gates chosen for robustness. There is clear headroom to reduce the time via finer calibration, tuning of the visual servoing gains, or learning-based visual servoing. Mechanical compliance can likewise be improved to enlarge the capture basin on targets. Even with the current conservative settings, the results demonstrate universal landing across diverse pole heads with a simple and reliable autonomy pipeline.

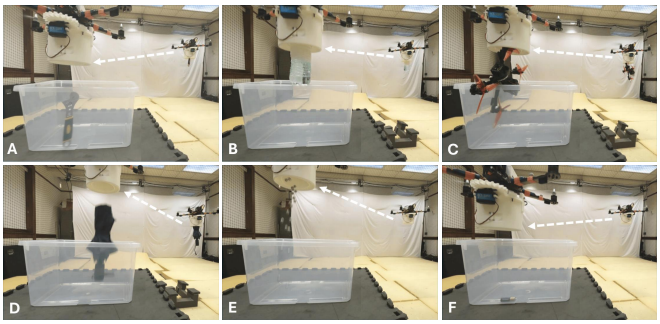

Fig. 9: Transport-and-release with diverse objects along an identical trajectory. (A) Wrench; (B) Partially filled water bottle; (C) Five-inch quadrotor; (D) Umbrella; (E) Steel marbles; (F) AA battery. The quadrotor takes off with the object grasped, follows the same trajectory, and drops each object into a box at a fixed location.

C. Grasping diverse objects

To validate the gripper's compliance to a variety of shapes and masses of objects, we let the quadrotor grasp and transport everyday objects, including an AA battery (slender, rigid), a partially filled water bottle (sloshing liquid, moving CoM), an umbrella (elongated), a screwdriver (metal, asymmetric), and a batch of steel marbles (multi-contact). In each trial, the drone grasps one object on the ground, takes off, and follows an identical minimum-snap trajectory to a fixed drop location. Despite a the uncertainties induced by the payloads with large variation in size, mass, and inertia, $\mathcal{L}_1$Quad facilitates the platform to maintain close tracking of the transit trajectory and consistent object placement. Qualitatively, these results indicate that the coaxial packaging and load-bearing rim preserve the geometry-agnostic grasping capability of the ROSE gripper [12].

D. Perch-while-grasping via adapter

We extend the pipeline to a perch-while-grasping scenario using a graspable adapter that clips onto the top edge of a stop sign. We use the FoundationPose module [22] to get a pre-flight estimate of the adapter pose and the stop-sign pose. Using these estimates as target poses, we generate two minimum-snap trajectories T_1 and T_2: Trajectory T_1 starts from a starting location and flies to hover above the adapter; Trajectory T_2 starts from the grasped adapter and flies to hover above the stop sign. The vehicle starts by tracking T_1 and consequently performs visual servoing to grasp the adapter. It then follows trajectory T_2 and executes visual servoing again to align the grasped adapter with the stop sign before descending to perch. The entire flight for stop-sign perching is shown in Fig. 1. Because the ROSE gripper grasps by torsional squeezing, the grasped adapter may rotate slightly relative to the quadrotor frame. Hence, the quadrotor shall control its yaw angle to align the adapter with the stop-sign's in-plane orientation. The quadrotor smoothly complete this two-stage task: Grasping the adapter and then perching with the grasped adapter without oscillations when stages switch. Yaw alignment between the adapter and stop sign is handled by the yaw control before lateral alignment. The

$\mathcal{L}_1$Quad maintains close trajectory tracking on T_1 and T_2 despite the extra weight from the adapter. Overall, the hybrid gripper enables reliable perching on thin edges using the same perception and control pipeline as pole perching. This procedure allows for perching on thin sign edges—fixtures that are typically infeasible for generic landing gear—thereby expanding the set of admissible perching locations in urban environments.

VI. Conclusion

We present a lightweight, dual-functional end-effector that unifies grasping and perching for aerial robots and enables a new perch-while-grasping capability. The design builds on a ROSE gripper core for compliant, geometry-agnostic grasping and adds a rigid shell that carries perching loads through a separated structural path. A coaxial, eye-in-hand camera is integrated through a central bore, aligning the sensing and manipulation axes to reduce occlusion, parallax, and calibration drift during terminal approach. These hardware choices are applied to a quadrotor carrier with an autonomy stack: forward-view 6D pose estimation for approach planning, minimum-snap trajectory tracking by a geometric controller with $\mathcal{L}_1$ adaptive augmentation, and visual servoing for centimeter-level alignment near contact. Experiments validate our hybrid girpper's versatility and robustness. The platform perches across diverse pole geometries and sizes with high success rates. It can grasp and transport everyday objects of varied mass and shape without retuning the controller. We also showcase the perch-while-grasping capability on top of a stop sign using a clip-type adapter.

Our current design has several limitations: The perching is constrained to upright poles and traffic signs only; coarse forward-view pose estimation quality can increase alignment time; and occasional failures may happen when perching to poles with relatively large diameters. Future work will address these issues with an extension to pose-aware perching for inclined and irregular fixtures using fine 6D pose estimates and adding two rotational degrees of freedom to the camera-integrated gripper while preserving the separated load path.

References

[1] C. E. Doyle, J. J. Bird, T. A. Isom, J. C. Kallman, D. F. Bareiss, D. J. Dunlop, R. J. King, J. J. Abbott, and M. A. Minor, "An avian-inspired passive mechanism for quadrotor perching," *IEEE/ASME Transactions On Mechatronics*, vol. 18, no. 2, pp. 506–517, 2012.

[2] W. R. Roderick, M. R. Cutkosky, and D. Lentink, "Bird-inspired dynamic grasping and perching in arboreal environments," *Science Robotics*, vol. 6, no. 61, p. eabj7562, 2021.

[3] H. W. Wopereis, T. Van Der Molen, T. H. Post, S. Stramigioli, and M. Fumagalli, "Mechanism for perching on smooth surfaces using aerial impacts," in *Proceedings of 2016 IEEE International Symposium on Safety, Security, and Rescue Robotics*. IEEE, 2016, pp. 154–159.

[4] J. D. Ang, L. Librado, C. J. Salaan, J. Maglasang, K. Sanchez, and M. Ang, "Drone with pneumatic-tethered suction-based perching mechanism for high payload application," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2022, pp. 12 154–12 161.

[5] A. Lussier Desbiens and M. R. Cutkosky, "Landing and perching on vertical surfaces with microspines for small unmanned air vehicles," *Journal of Intelligent and Robotic Systems*, vol. 57, pp. 313–327, 2010.

[6] S. Park, D. S. Drew, S. Follmer, and J. Rivas-Davila, "Lightweight high voltage generator for untethered electroadhesive perching of micro air vehicles," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 4485–4492, 2020.

[7] J. Fishman, S. Ubellacker, N. Hughes, and L. Carlone, "Dynamic grasping with a "soft" drone: From theory to practice," in *Procedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2021, pp. 4214–4221.

[8] A. L. Gunderman, J. A. Collins, A. L. Myers, R. T. Threlfall, and Y. Chen, "Tendon-driven soft robotic gripper for blackberry harvesting," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2652–2659, 2022.

[9] T. Sun, Y. Chen, T. Han, C. Jiao, B. Lian, and Y. Song, "A soft gripper with variable stiffness inspired by pangolin scales, toothed pneumatic actuator and autonomous controller," *Robotics and Computer-Integrated Manufacturing*, vol. 61, p. 101848, 2020.

[10] G. B. Crowley, X. Zeng, and H.-J. Su, "A 3D printed soft robotic gripper with a variable stiffness enabled by a novel positive pressure layer jamming technology," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 5477–5482, 2022.

[11] R. Kitchen, N. Bierwolf, S. Harbertson, B. Platt, D. Owen, K. Griessmann, and M. A. Minor, "Design and evaluation of a perching hexacopter drone for energy harvesting from power lines," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2020, pp. 1192–1198.

[12] K. T. Nguyen, N. H. Nguyen, and V. A. Ho, "Soft yet secure: Exploring membrane buckling for achieving a versatile grasp with a rotation-driven squeezing gripper," *The International Journal of Robotics Research*, p. 02783649241272120, 2024.

[13] J. Thomas, G. Loianno, K. Daniilidis, and V. Kumar, "Visual servoing of quadrotors for perching by hanging from cylindrical objects," *IEEE Robotics and Automation Letters*, vol. 1, no. 1, pp. 57–64, 2015.

[14] Z. Wu, S. Cheng, P. Zhao, A. Gahlawat, K. A. Ackerman, A. Lakshmanan, C. Yang, J. Yu, and N. Hovakimyan, "L1Quad: L1 adaptive augmentation of geometric control for agile quadrotors with performance guarantees," *IEEE Transactions on Control Systems Technology*, 2025.

[15] W. Crooks, G. Vukasin, M. O'Sullivan, W. Messner, and C. Rogers, "Fin ray® effect inspired soft robotic gripper: From the robosoft grand challenge toward optimization," *Frontiers in Robotics and AI*, vol. 3, p. 70, 2016.

[16] M. Langer, E. Amanov, and J. Burgner-Kahrs, "Stiffening sheaths for continuum robots," *Soft Robotics*, vol. 5, no. 3, pp. 291–303, 2018.

[17] Y. Shan, Y. Zhao, H. Wang, L. Dong, C. Pei, Z. Jin, Y. Sun, and T. Liu, "Variable stiffness soft robotic gripper: design, development, and prospects," *Bioinspiration & Biomimetics*, vol. 19, no. 1, p. 011001, 2024.

[18] S. Ubellacker, A. Ray, J. M. Bern, J. Strader, and L. Carlone, "High-speed aerial grasping using a soft drone with onboard perception," *npj Robotics*, vol. 2, no. 1, p. 5, 2024.

[19] H. Chen, F. Quan, L. Fang, and S. Zhang, "Aerial grasping with a lightweight manipulator based on multi-objective optimization and visual compensation," *Sensors*, vol. 19, no. 19, p. 4253, 2019.

[20] X. Ye, H. Cui, L. Wang, S. Xie, and H. Ni, "Vision-guided hierarchical control and autonomous positioning for aerial manipulator," *Applied Sciences*, vol. 13, no. 22, p. 12172, 2023.

[21] F. Chaumette and S. Hutchinson, "Visual servo control. ii. advanced approaches [tutorial]," *IEEE Robotics & Automation Magazine*, vol. 14, no. 1, pp. 109–118, 2007.

[22] B. Wen, W. Yang, J. Kautz, and S. Birchfield, "Foundationpose: Unified 6D pose estimation and tracking of novel objects," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 17 868–17 879.

[23] D. Mellinger and V. Kumar, "Minimum snap trajectory generation and control for quadrotors," in *Proceedings of IEEE International Conference on Robotics and Automation*. IEEE, 2011, pp. 2520–2525.

[24] T. Lee, M. Leok, and N. H. McClamroch, "Geometric tracking control of a quadrotor UAV on SE(3)," in *Proceedings of the 49th IEEE Conference on Decision and Control*, Atlanta, GA, USA, 2010, pp. 5420–5425.

[25] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.