Autonomous Wingtip Docking of Fixed-Wing UAVs: Relative Navigation, Closed-Loop Control, and System-Level Challenges

Yulong Zhang, Wei Zhou*, Jing Zhou, Peiyang Ma, and Qiuyan Wang

Department of Rocket Force University of Engineering, Xi'an, China

*Corresponding author: zhouweilunwen@163.com

Abstract

Autonomous wingtip docking of fixed-wing unmanned aerial vehicles offers a potential approach to long-endurance flight, modular mission reconfiguration, and airborne assembly. Unlike conventional formation flight, this task requires two forward-flying aircraft to reduce the relative position, attitude, and closing-velocity errors between their actual docking points while satisfying fixed-wing flight-envelope, visibility, aerodynamic-interaction, and collision-avoidance constraints. This article organizes relevant research around two closely coupled themes: relative navigation and closed-loop docking control. Vision-based relative measurement methods are examined, including cooperative fiducial markers, structure-based non-cooperative perception, monocular and stereo pose estimation, and learning–geometry hybrid approaches; multi-sensor relative-state estimation is then discussed with emphasis on visual–inertial fusion, GNSS/RTK and UWB assistance, dual-moving-platform modeling, asynchronous measurements, lever-arm compensation, observability, and uncertainty representation. The control literature is further analyzed from the perspectives of position-based and image-based visual servoing, robust and prescribed-performance control, model predictive and safety-critical control, and learning-based disturbance compensation. The central technical challenge lies not in improving pose-estimation accuracy or tracking convergence in isolation, but in constructing an uncertainty-aware, constraint-compliant, and contact-transition-capable autonomous closed-loop system. Future research should therefore focus on dual-platform relative observability, confidence-aware closure control, visibility-preserving trajectory planning, real-time near-field aerodynamic modeling, hybrid contact dynamics, navigation–control–mechanism co-design, and progressively structured experimental validation.

Keywords

Fixed-wing UAV; Wingtip Docking; Relative Navigation; Visual–inertial Fusion; Visual Servoing.

1. Introduction

1.1 Background and Motivation

Fixed-wing unmanned aerial vehicles (UAVs) offer high aerodynamic efficiency, long endurance, and relatively large payload capacity, and have consequently been widely used in surveillance, communication relay, environmental monitoring, disaster response, and long-range inspection. However, the energy supply, payload configuration, and aerodynamic layout of an individual UAV are generally determined before takeoff and cannot be substantially modified during flight. As unmanned aviation evolves from single-vehicle autonomy toward cooperative, heterogeneous, and reconfigurable systems, airborne refueling, aerial recovery, modular assembly, and in-flight vehicle combination have attracted increasing attention as possible means of extending mission endurance and overcoming the capability limitations of a single platform [1, 2].

Among these concepts, wingtip docking enables two or more fixed-wing vehicles to form an airborne multibody configuration through mechanical connections located near their wingtips. Such a configuration may provide opportunities for aerodynamic efficiency improvement, modular mission reconfiguration, and distributed airborne systems. Unlike conventional formation flight, however, docking requires the relative separation between two moving aircraft to be continuously reduced until their mechanical interfaces enter a limited capture region. During this process, the relative position, attitude, and closing velocity of the actual docking points must satisfy the tolerances of the docking mechanism while the aircraft remain within their flight envelopes and avoid unintended contact.

Several studies have examined individual aspects of wingtip connection. Meng et al. investigated the conceptual design, dynamics, and flight control of a wingtip-connected multibody aircraft and demonstrated that the connected configuration could maintain flight under dedicated control, although its stability characteristics differed substantially from those of an isolated aircraft [3]. Zhou et al. experimentally studied the aerodynamic characteristics of two fixed-wing UAVs during close wingtip approach and showed that the rolling moment acting on the following aircraft varied significantly with the relative lateral, vertical, and longitudinal positions [4]. Huo et al. integrated a wingtip docking mechanism, monocular relative localization, and docking control on compound vertical-takeoff-and-landing UAVs and conducted preliminary outdoor docking experiments under conditions in which the target vehicle remained in hover [5]. These studies confirm the technical relevance of wingtip docking, but they address connected flight, near-field aerodynamic interaction, or docking with a hovering target separately. A complete autonomous process involving target acquisition, close-range approach, precise alignment, mechanical capture, and post-contact stabilization between two continuously forward-flying fixed-wing UAVs has not yet been demonstrated in the open literature.

1.2 Distinctive Characteristics of Fixed-Wing Wingtip Docking

Fixed-wing wingtip docking differs fundamentally from both conventional formation flight and multirotor docking. A fixed-wing aircraft must maintain sufficient forward airspeed to generate lift and cannot independently produce arbitrary lateral, longitudinal, and vertical translational motion. Lateral displacement is normally achieved through coordinated roll and heading changes, whereas vertical motion depends on pitch, flight-path angle, airspeed, and lift dynamics. Therefore, the three-dimensional docking errors cannot be mapped directly into three independent translational velocity commands.

A second distinctive feature is the long lever arm between the vehicle center of mass and the wingtip docking point. Small roll, pitch, or yaw errors may produce considerable displacement at the wingtip. Consequently, controlling the relative position of the aircraft centers of mass does not guarantee alignment of the mechanical interfaces. The relevant controlled variables should instead be defined between the actual docking points and should include relative position, relative attitude, and closing velocity. The flight tests of connected multibody aircraft and the wind-tunnel investigation of close wingtip interaction both indicate that roll attitude and wingtip geometry play critical roles in the stability and aerodynamic behavior of the system [3, 4].

Near-field aerodynamic interaction further increases the difficulty of terminal docking. Conventional flight models often treat another aircraft as an external disturbance source or neglect its influence entirely. Experimental evidence, however, indicates that the induced aerodynamic forces and moments during wingtip approach depend systematically on relative geometry rather than behaving as purely random disturbances [4]. As the separation decreases, the effective plant dynamics may therefore vary continuously, reducing the validity of controllers designed using isolated-aircraft models.

The perception and control processes are also strongly coupled. Aggressive roll or heading maneuvers may accelerate lateral-error correction but can simultaneously increase perspective distortion, motion

ISSN: 2414-1895 DOI: 10.6919/ICJE.202608_12(8).0007

blur, partial occlusion, or the probability of losing the target from the camera field of view. Faster longitudinal closure enlarges the target image and shortens docking time, but it also reduces the time available to reject erroneous measurements and execute a safe abort maneuver. Studies on air-to-air visual navigation and visual servoing for moving targets have demonstrated that target attitude, image location, observation geometry, and processing delay can significantly affect closed-loop performance [6-9].

Finally, wingtip docking is a safety-critical operation. In ordinary trajectory tracking, a temporary overshoot may be acceptable, whereas a comparable error during docking may cause wing collision, excessive impact loads, or incorrect engagement of the capture mechanism. The control system must therefore support not only convergence toward the target but also holding, retreat, re-alignment, and abort behaviors. Recent studies on aerial refueling and aerial recovery have increasingly incorporated closing-velocity limits, safety envelopes, minimum-airspeed constraints, actuator saturation, and collision-avoidance conditions into the control design [2, 10-13].

1.3 Related Research and Scope of This Article

Direct research on the complete autonomous wingtip docking of two fixed-wing UAVs remains limited. Most of the relevant methods are distributed across several neighboring areas, including probe-and-drogue autonomous aerial refueling, fixed-wing aerial recovery, visual autonomous landing, cooperative relative navigation, and visual servo control [14, 15].

Probe-and-drogue aerial refueling is one of the most closely related and systematically studied applications. Existing studies have developed relatively complete task-phase descriptions and have addressed drogue detection, monocular and stereo pose estimation, visual servoing, aerodynamic-disturbance rejection, actively controlled drogues, prescribed-performance control, and docking-success prediction [1, 2, 12, 16-35]. These results provide valuable foundations for task decomposition, terminal relative navigation, and constrained closure control. Nevertheless, the flexible hose-and-drogue dynamics and probe geometry differ from the rigid mechanical interface and long wingtip lever arm considered in wingtip docking.

Fixed-wing aerial recovery provides another important source of transferable methods. Moving arrest systems, aerial capture devices, and SideArm recovery have motivated research on RTK-GNSS navigation, visual guidance, prescribed-performance control, safe approach corridors, closing-velocity regulation, and safety-critical model predictive control [10-13, 28-32, 36-39]. These studies are particularly relevant to the problem of approaching a moving target while maintaining the fixed-wing flight envelope, although their capture geometry and allowable terminal errors differ from those of wingtip connection.

In relative navigation, cooperative fiducial markers, perspective-n-point solvers, circular features, and learning-based keypoint detection provide the basis for local pose measurement [20-24, 40-48]. Visual-inertial, GNSS-visual-inertial, and UWB-visual fusion methods further provide mechanisms for high-rate state propagation, coordinate alignment, delayed-measurement processing, and failure mitigation [6, 44-46, 49-66]. However, most conventional visual-inertial odometry and simultaneous localization and mapping methods estimate the motion of a single vehicle relative to a static environment. Their reported performance cannot be directly interpreted as the accuracy of relative navigation between two independently moving aircraft. Motion compensation for both vehicles, long lever arms, communication delay, correlated navigation information, and relative-state observability require dedicated treatment.

Closed-loop docking control has been studied using position-based visual servoing, image-based visual servoing, robust and sliding-mode control, prescribed-performance and predefined-time control, model predictive control, control barrier functions, iterative learning, and neural compensation [7, 11-13, 16-18, 25-32, 67, 68]. These approaches address different aspects of the problem, including disturbance rejection, convergence-rate specification, actuator constraints, visibility, and collision avoidance. Nevertheless, many control studies assume accurate, continuous, and delay-free relative-state feedback, whereas perception studies often evaluate only average pose errors without examining how uncertainty and outliers propagate into docking risk.

This article therefore focuses on the autonomous pre-contact process of fixed-wing UAV wingtip docking and organizes the relevant research around two tightly coupled themes: relative navigation and closed-loop control. The discussion covers vision-based relative measurement, multi-sensor relative-state estimation, fixed-wing docking guidance, visual servoing, robust and constrained control, uncertainty-aware safety decisions, near-field aerodynamic interaction, and contact-transition modeling. Research on aerial refueling, aerial recovery, multirotor docking, and general visual-inertial navigation is included only where it provides transferable concepts, models, or implementation experience. The objective is to clarify the current methodological foundations, distinguish direct evidence from cross-domain analogies, identify unresolved system-level challenges, and provide a structured basis for future autonomous wingtip docking research.

2. Task Characteristics and Technical Requirements of Fixed-Wing UAV Wingtip Docking

Fixed-wing UAV wingtip docking is not a simple extension of formation flight toward a smaller separation distance. In conventional formation flight, each aircraft tracks a prescribed nonzero relative position while collision avoidance remains a fundamental constraint. Wingtip docking instead requires the actual mechanical interfaces of two moving aircraft to progressively enter a limited capture region and, during this process, simultaneously satisfy constraints on relative position, relative attitude, closing velocity, approach direction, vehicle flight envelope, and sensing reliability. The task therefore combines the characteristics of cooperative navigation, precision guidance, safety-critical control, and contact preparation.

2.1 Task Process and Research Scope

Autonomous aerial refueling and aerial recovery studies generally divide the mission into multiple stages according to sensing range, control objective, and safety requirement. Probe-and-drogue aerial refueling is commonly described through rendezvous, joining, contact or refueling, and separation phases, while practical sensor architectures further distinguish global navigation, relative target acquisition, visual tracking, terminal alignment, and abort operations [1, 2]. Fixed-wing aerial recovery systems similarly employ global navigation for coarse convergence and switch to local relative sensing as the UAV approaches the moving recovery device [11, 36-38].

Based on these related task structures, the pre-contact process of fixed-wing wingtip docking can be divided into five functional stages: rendezvous, target acquisition, approach, precise alignment, and terminal closure. This division is used here as an analytical framework rather than as an established industrial standard.

During the **rendezvous stage**, the two UAVs enter a common mission region and establish communication and coarse relative navigation. The dominant information sources are typically GNSS, inertial navigation, flight-controller states, and inter-UAV communication. The main objective is to reduce the large-scale separation in position, altitude, velocity, and heading while maintaining sufficient safety distance. At this stage, centimeter-level relative accuracy is unnecessary, but navigation continuity and communication reliability are essential.

During the **target-acquisition stage**, the host UAV establishes a stable observation of the target UAV or its cooperative marker. Whole-aircraft appearance, line-of-sight direction, apparent size, and coarse range may be used to confirm target identity and initialize visual tracking [6, 41]. The control objective is no longer limited to trajectory convergence; the host UAV must also keep the target within the effective field of view and create favorable observation geometry for subsequent pose estimation.

During the **approach stage**, the host UAV moves toward a predefined pre-docking region. Global navigation information is gradually supplemented or replaced by local relative measurements from

ISSN: 2414-1895 DOI: 10.6919/ICJE.202608_12(8).0007

vision, UWB, RTK-GNSS, or other ranging devices [36-38, 51, 54, 55]. The main task is to reduce relative position and velocity errors without entering the collision-sensitive region prematurely. Smooth switching or weighting between navigation sources is important because abrupt changes in the reference frame or measurement bias may lead to control transients.

During the **precise-alignment stage**, the system must align the actual wingtip docking points rather than merely maintain a relative position between vehicle centers of mass. Six-degree-of-freedom relative pose, relative velocity, measurement quality, and state-estimation confidence become increasingly important. The controller should primarily reduce lateral, vertical, roll, pitch, and yaw misalignment while restricting the longitudinal closing rate.

During the **terminal-closure stage**, the two docking interfaces enter the capture region of the mechanical mechanism. The relative position and attitude errors must remain within the allowable capture envelope, and the closing velocity must be sufficiently low to avoid excessive impact loads or incorrect engagement. Terminal closure should be permitted only after the alignment, state-confidence, visibility, airspeed, and actuator conditions have been verified. If any critical condition is violated, the system should hold its relative position, retreat to the pre-docking region, or execute an abort maneuver.

The complete wingtip docking mission also includes contact, mechanical capture, locking, and post-connection flight. However, the present discussion focuses primarily on the autonomous pre-contact process from target acquisition to terminal closure. The docking mechanism is considered through its capture tolerance and admissible contact conditions, whereas detailed structural design, impact mechanics, and connected-flight control are discussed only when they affect navigation and control requirements.

2.2 Relative Navigation and Closed-Loop Control Requirements

The geometric relationship between the aircraft center of mass and the wingtip docking interface is fundamental to the formulation of the relative-navigation and control problem. Let the host and target UAVs be denoted by the subscripts h and t, respectively. For UAV i, where $i \in \{h, t\}$, let $\mathbf{p}_{B_i}^{I}$ denote the position of the body-frame origin in the inertial frame $\{I\}$, $\mathbf{R}_{B_i}^{I}$ denote the rotation matrix from the body frame $\{B_i\}$ to the inertial frame, and $\mathbf{r}_{D_i}^{B_i}$ denote the fixed lever-arm vector from the body-frame origin to the wingtip docking point D_i, expressed in $\{B_i\}$. The inertial position of the corresponding docking point is

$$\mathbf{p}_{\mathrm{D_i}}^{\mathrm{I}} = \mathbf{p}_{\mathrm{B_i}}^{\mathrm{I}} + \mathbf{R}_{\mathrm{B_i}}^{\mathrm{I}} \mathbf{r}_{\mathrm{D_i}}^{\mathrm{B_i}}, \quad \mathrm{i} \in \{\mathrm{h}, \mathrm{t}\}. \tag{1}$$

The relative position error between the target and host docking points can then be expressed in the host body frame as

$$\boldsymbol{e}_p^{B_h} = \boldsymbol{R}_I^{B_h}\left(\boldsymbol{p}_{D_t}^{I} - \boldsymbol{p}_{D_h}^{I}\right). \tag{2}$$

where $\mathbf{R}_I^{B_h} = \left(\mathbf{R}_{B_h}^{I}\right)^T$. This definition makes the three components of $\mathbf{e}_p^{B_h}$ directly interpretable as longitudinal, lateral, and vertical docking-point errors relative to the host UAV. Because the wingtip is located far from the aircraft center of mass, the error is influenced by both translational displacement and attitude variation. A small roll, pitch, or yaw error may therefore generate a considerable displacement at the docking interface even when the aircraft centers of mass are accurately positioned.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202608_12(8).0007

The velocity of docking point D_i is determined by both the translational velocity of the aircraft and the velocity induced by body rotation. Let $\mathbf{v}_{B_i}^{I}$ be the translational velocity of the body-frame origin and $\boldsymbol{\omega}_{B_i}^{B_i}$ be the body angular velocity. The docking-point velocity in the inertial frame is

$$\boldsymbol{v}_{D_i}^{I} = \boldsymbol{v}_{B_i}^{I} + \boldsymbol{R}_{B_i}^{I}\left(\boldsymbol{\omega}_{B_i}^{B_i} \times \boldsymbol{r}_{D_i}^{B_i}\right). \tag{3}$$

Accordingly, the relative docking-point velocity expressed in the host body frame can be written as

$$\boldsymbol{v}_{rel}^{B_h} = \boldsymbol{R}_{I}^{B_h}\left(\boldsymbol{v}_{D_t}^{I} - \boldsymbol{v}_{D_h}^{I}\right). \tag{4}$$

The longitudinal component of $\mathbf{v}_{\mathrm{rel}}^{B_h}$ represents the closing velocity, whereas its lateral and vertical components describe the residual relative motion perpendicular to the desired docking direction. The rotational contribution in the docking-point velocity cannot be neglected when the aircraft executes rapid roll or yaw corrections, because the long wingtip lever arm can amplify the velocity induced by angular motion.

The relative attitude between the two docking interfaces is also required during precise alignment. Let $\mathbf{R}_{D_h}^{I}$ and $\mathbf{R}_{D_t}^{I}$ denote the orientations of the host and target docking-point frames with respect to the inertial frame. The relative attitude of the target docking frame with respect to the host docking frame is

$$\boldsymbol{R}_{D_t}^{D_h} = \left(\boldsymbol{R}_{D_h}^{I}\right)^{T}\boldsymbol{R}_{D_t}^{I}. \tag{5}$$

If $\mathbf{R}_d$ denotes the desired relative docking attitude, a compact attitude-error vector can be defined on the rotation group as

$$\boldsymbol{e}_R = Log\left(\boldsymbol{R}_d^{T}\boldsymbol{R}_{D_t}^{D_h}\right)^{\vee}. \tag{6}$$

where $\mathrm{Log}(\cdot)$ is the logarithmic map from $SO(3)$ to its Lie algebra and $(\cdot)^{\vee}$ converts a skew-symmetric matrix into a three-dimensional vector. The vector $\mathbf{e}_R$ describes the rotational misalignment that must be reduced before terminal closure.

Based on these geometric relationships, the minimum control-oriented relative state can be represented as

$$\boldsymbol{x}_{rel} = \left[\left(\boldsymbol{e}_p^{B_h}\right)^{T} \quad \left(\boldsymbol{v}_{rel}^{B_h}\right)^{T} \quad \boldsymbol{e}_R^{T}\right]^{T}. \tag{7}$$

For practical closed-loop operation, the navigation output should not be limited to the estimated mean state $\hat{\mathbf{x}}_{\mathrm{rel}}$. It should also include an estimation covariance $\mathbf{P}_{\mathrm{rel}}$, accurate timestamps, visual-quality indicators, and sensor-health flags. The information supplied to the controller can therefore be represented by the belief state

$$\boldsymbol{b}_{rel} = (\hat{\boldsymbol{x}}_{rel}, \boldsymbol{P}_{rel}). \tag{8}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202608_12(8).0007

A smooth estimated trajectory is not necessarily reliable when the associated uncertainty is large. For example, temporary visual loss may be bridged by inertial propagation, but the covariance should increase continuously during the loss interval. The controller should use this increasing uncertainty to reduce the closing velocity, maintain the current separation, or return to a safe pre-docking state.

The sensing requirements vary substantially across the docking process. Long-range rendezvous emphasizes coverage, navigation continuity, and global consistency, whereas precise alignment requires high local accuracy, low latency, and robustness to rapid perspective variation. Vision provides rich geometric information but is sensitive to illumination, image scale, occlusion, motion blur, and processing delay. The IMU provides high-rate motion information but suffers from sensor bias and integration drift. GNSS, RTK, UWB, and inter-UAV communication extend the operating range but may be affected by blockage, interference, data-link latency, and inconsistent coordinate frames [1, 6, 49-55]. Therefore, no single sensor is generally sufficient for all docking stages.

The relative-navigation system should first ensure cross-scale continuity. When the dominant navigation source changes from global navigation to local visual sensing, coordinate alignment, bias compensation, and source-transition logic should prevent discontinuities from being transmitted to the controller. The system should also maintain temporal consistency. Camera exposure time, image-processing completion time, IMU sampling time, flight-controller timestamps, and remote-state reception time are generally different. A delayed measurement should be associated with the state at its actual observation time rather than being directly applied to the current state.

Measurement uncertainty should be adjusted according to the operating condition. The covariance of a visual pose measurement may be modeled as a function of target distance d, viewing angle α, target image area A_{img}, reprojection error e_{rep}, image-blur quality q_{blur}, and occlusion quality q_{occ}:

$$\boldsymbol{R}_v = f\big(d, \alpha, A_{img}, e_{rep}, q_{blur}, q_{occ}\big). \tag{9}$$

This formulation indicates that the measurement-noise covariance $\mathbf{R}_v$ should not be treated as a constant. Visual measurements obtained at long range, under large viewing angles, or during partial occlusion should be assigned lower confidence than measurements acquired under favorable imaging conditions.

The closed-loop controller must also respect the kinematic and dynamic limitations of fixed-wing aircraft. Unlike a multirotor, a fixed-wing UAV cannot independently command arbitrary Cartesian velocities. The relative docking-point errors must be converted into feasible airspeed, flight-path, heading, roll, and pitch commands. The command vector may be expressed as

$$\boldsymbol{u}_c = [V_{a,d} \quad \phi_d \quad \theta_d \quad \psi_d]^T. \tag{10}$$

where $V_{a,d}$, ϕ_d, θ_d, and ψ_d denote the desired airspeed, roll angle, pitch angle, and heading angle, respectively. These commands must satisfy the vehicle flight envelope and actuator limitations:

$$V_a \geq V_{min}, \tag{11}$$

$$|\phi| \leq \phi_{max}, \qquad |\theta| \leq \theta_{max}, \tag{12}$$

$$\boldsymbol{u}_{min} \leq \boldsymbol{u}_c \leq \boldsymbol{u}_{max}. \tag{13}$$

A hierarchical architecture is therefore preferable. The outer-loop docking controller generates desired airspeed, path, or attitude commands from the relative state, whereas the existing autopilot stabilizes the aircraft and tracks these commands. This architecture reduces the risk associated with replacing flight-proven inner-loop controllers, although the outer-loop design must account for the response delay, bandwidth, and saturation characteristics of the inner loop [17, 37].

Terminal closure should not be initiated solely because the estimated position error is small. A typical terminal-entry condition may be expressed as

$$\|\boldsymbol{e}_p^{B_h}\| \leq e_{p,max}, \qquad \|\boldsymbol{e}_R\| \leq e_{R,max}, \tag{14}$$

$$|v_{cl}| \leq v_{cl,max}, \qquad \lambda_{max}(\boldsymbol{P}_{rel}) \leq P_{max}. \tag{15}$$

Here, $e_{p,\max}$ and $e_{R,\max}$ are the admissible position and attitude errors, $v_{\mathrm{cl},\max}$ is the maximum allowable closing velocity, and $\lambda_{\max}(\mathbf{P}_{\mathrm{rel}})$ is the maximum eigenvalue of the relative-state covariance matrix. Additional requirements should include valid visual tracking, sufficient airspeed margin, and adequate control authority. If any condition is violated, the system should hold, retreat, re-align, or abort rather than continue the terminal approach [10-13, 33].

2.3 Key Technical Conflicts

The difficulty of fixed-wing wingtip docking can be summarized through several coupled technical conflicts.

The first conflict is between **cross-scale operation and limited sensor capability**. Global navigation systems are suitable for long-range rendezvous but generally cannot independently provide the local accuracy required for terminal docking. Vision can offer precise local geometry but has a limited field of view and becomes unreliable when the target is too distant, blurred, occluded, or poorly illuminated. The central problem is therefore not to identify a single universal sensor, but to combine global, relative, and local measurements within a consistent state-estimation framework.

The second conflict is between **underactuated vehicle motion and multi-degree-of-freedom alignment**. Wingtip docking requires simultaneous reduction of three-dimensional position errors and attitude errors, whereas fixed-wing translational motion is indirectly generated through attitude and aerodynamic forces. Roll and heading changes are needed to correct lateral position, but the same attitude changes directly displace the wingtip docking point through the lever arm. The variables used to control the trajectory are therefore also part of the docking geometry.

The third conflict is between **rapid error convergence and visual observability**. Aggressive maneuvers may reduce relative-position errors quickly but can move the target toward the image boundary, increase perspective distortion, reduce the apparent marker area, or produce motion blur. A spatially efficient trajectory may therefore be unfavorable for visual estimation. Future docking guidance should consider both geometric reachability and expected sensing quality.

The fourth conflict is between **docking efficiency and safety margin**. A large closing speed reduces mission time and exposure to environmental disturbances, but it also amplifies the effect of sensing delay and leaves less time for corrective or abort actions. Conversely, excessively conservative closure increases mission duration and may lead to repeated exposure to wind disturbances or visual degradation. The appropriate closing speed should therefore depend on alignment error, relative dynamics, estimation uncertainty, available control authority, and mechanical capture tolerance.

The fifth conflict is between **modular subsystem development and system-level consistency**. Vision algorithms are commonly evaluated using image-based detection or pose-estimation metrics, navigation filters using trajectory error, controllers using terminal tracking error, and docking mechanisms using static capture tolerance. These separate metrics do not automatically guarantee system-level success. A small average pose error may still contain rare but dangerous jumps, while a controller with good nominal convergence may be unsafe under delayed or biased state estimates. Navigation, control, and mechanical capture should therefore be designed using shared definitions of docking-point error, closing velocity, uncertainty, and success criteria.

These conflicts demonstrate that fixed-wing wingtip docking should be treated as an integrated closed-loop problem. Relative navigation must provide not only an estimated state but also a representation of its reliability, while the controller must regulate the approach without degrading visibility, violating the flight envelope, or exceeding the mechanical capture conditions. This task-oriented relationship provides the basis for the relative-navigation methods discussed in the next section.

3. Relative Navigation for Fixed-Wing Aerial Docking

Relative navigation provides the state information required for target acquisition, close-range approach, precise alignment, and terminal closure. For fixed-wing wingtip docking, the required navigation output is not merely the pose of a visual marker with respect to a camera, but the relative position, velocity, and attitude between the two physical docking interfaces. The estimation process must also compensate for host-vehicle motion, camera and marker lever arms, asynchronous sensor measurements, visual dropouts, and the continuously changing observation geometry between two moving aircraft.

3.1 Vision-Based Relative Measurement

Vision is particularly suitable for terminal docking because it can provide bearing, scale, structural features, and six-degree-of-freedom pose information without requiring physical contact. Existing approaches can be broadly categorized into cooperative-marker methods, structure-based non-cooperative methods, stereo or multi-view methods, and learning-geometry hybrid methods.

3.1.1 Cooperative-Marker-Based Measurement

Cooperative methods install a target with known geometry near the docking interface. Typical targets include AprilTag, ArUco, circular patterns, concentric rings, and specially designed multi-plane marker assemblies [20-24, 42, 43, 46-48]. Because the three-dimensional coordinates of the marker features are known in advance, the relative pose can be recovered from their image projections.

Let the coordinates of feature point i in the marker frame $\{M\}$ be $\mathbf{P}_i^M$, and let its image coordinates be $\mathbf{u}_i = [u_i, v_i]^T$. Under the pinhole camera model,

$$s_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \boldsymbol{K}[\boldsymbol{R}_M^C \quad \boldsymbol{t}_M^C] \begin{bmatrix} \boldsymbol{P}_i^M \\ 1 \end{bmatrix}. \tag{16}$$

where $\mathbf{K}$ is the camera intrinsic matrix, $\mathbf{R}_M^C$ and $\mathbf{t}_M^C$ represent the marker orientation and position with respect to the camera frame $\{C\}$, and s_i is the projective scale.

The pose can be estimated by minimizing the reprojection error

$$\min_{\boldsymbol{R}_M^C, \boldsymbol{t}_M^C} \sum_{i=1}^{N} \|\boldsymbol{u}_i - \pi(\boldsymbol{K}[\boldsymbol{R}_M^C \mid \boldsymbol{t}_M^C] \bar{\boldsymbol{P}}_i^M)\|^2. \tag{17}$$

where $\pi(\cdot)$ denotes perspective normalization and $\bar{\mathbf{P}}_i^M$ is the homogeneous representation of the three-dimensional feature point. EPnP provides an efficient solution to the perspective-n-point problem [44], whereas IPPE is specifically designed for planar targets and explicitly handles the two-solution ambiguity that frequently occurs under weak-perspective conditions [45]. In practical

ISSN: 2414-1895 DOI: 10.6919/ICJE.202608_12(8).0007

docking systems, the physically valid solution should be selected using positive-depth constraints, reprojection error, frame-to-frame continuity, inertial prediction, and feasible relative-motion limits.

Planar markers are easy to manufacture and detect, but their performance may deteriorate when the target appears small, the viewing direction is nearly parallel to the marker plane, or part of the marker is occluded. Multi-plane or three-dimensional cooperative targets can improve viewpoint robustness, although they increase aerodynamic drag, structural complexity, and calibration requirements. For wingtip docking, the marker should be positioned so that it remains visible during roll and yaw corrections and does not interfere with the capture mechanism.

The measured marker pose must be converted into the relative pose between the physical docking interfaces. Let $\mathbf{t}_M^C$ denote the measured marker position, $\mathbf{r}_{D_t}^M$ the vector from the target marker to the target docking point, and $\mathbf{r}_{D_h}^C$ the vector from the host camera to the host docking point.

The target docking-point position with respect to the host docking point can be expressed as

$$\boldsymbol{p}_{D_t D_h}^C = \boldsymbol{t}_M^C + \boldsymbol{R}_M^C \boldsymbol{r}_{D_t}^M - \boldsymbol{r}_{D_h}^C. \tag{18}$$

This transformation is essential because direct control of the camera-to-marker pose may leave a residual offset between the mechanical interfaces. The calibration accuracy of the camera-to-host-docking-point and marker-to-target-docking-point lever arms therefore directly affects terminal docking accuracy.

3.1.2 Structure-Based Non-Cooperative Measurement

Non-cooperative methods estimate relative information from the natural structure of the target aircraft without requiring dedicated markers. Whole-aircraft contours, wing edges, fuselage geometry, control surfaces, and key structural points can be detected and matched to a known three-dimensional model [6, 41]. Such methods are attractive for medium-range acquisition because the entire aircraft can remain visible before a small wingtip marker becomes detectable.

Learning-based object detectors can identify the target UAV and provide a region of interest, after which geometric methods estimate line of sight, apparent scale, range, or pose. Causa et al. combined air-to-air visual detection with navigation filtering for cooperative UAV navigation in GNSS-challenging environments [6]. Xu et al. used target-part detection and geometric reconstruction to obtain vision-only relative distance and pose information for multi-UAV systems [41].

The principal advantage of structure-based methods is their larger acquisition range and reduced dependence on dedicated target hardware. Their main limitations are model dependence, sensitivity to aircraft attitude, background confusion, partial occlusion, and comparatively low depth accuracy at long range. Consequently, they are more suitable for target acquisition and coarse approach than for final wingtip alignment. A practical docking system may therefore employ aircraft-level detection for long- and medium-range acquisition, followed by dedicated marker tracking for short-range precision measurement.

3.1.3 Stereo and Multi-View Measurement

Stereo vision estimates depth from image disparity and can reduce the scale ambiguity inherent in monocular systems. Binocular aerial-refueling studies have demonstrated the feasibility of recovering the relative position of a drogue from synchronized stereo images [21, 22]. For a rectified stereo system, the depth Z of a feature can be written as

$$Z = \frac{fb}{d_{px}}. \tag{19}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202608_12(8).0007

where f is the focal length, b is the stereo baseline, and d_{px} is the pixel disparity. Stereo vision can provide direct geometric depth, but its accuracy deteriorates rapidly as the target distance increases and the disparity approaches pixel-level uncertainty. A sufficiently large baseline improves depth accuracy but increases installation difficulty and sensitivity to structural deformation. Accurate camera synchronization and online or frequent extrinsic calibration are also required.

Multi-camera systems can extend the field of view and reduce the probability of target loss. However, asynchronous acquisition, different exposure conditions, and changing extrinsic parameters must be handled consistently. Multi-camera visual-inertial systems such as MIMC-VINS provide useful methods for asynchronous and redundant sensors, although they were not developed specifically for two-aircraft docking [64].

3.1.4 Learning-Geometry Hybrid Measurement

Purely geometric methods are interpretable and require limited training data, but they are sensitive to feature-extraction quality. End-to-end learning methods can improve robustness to illumination, blur, and background variation, but may produce physically inconsistent outputs outside the training distribution. A hybrid architecture combines the advantages of both approaches: a neural network performs target detection, segmentation, or keypoint prediction, while a geometric solver recovers the final relative pose.

AARPose uses lightweight learning-based keypoint extraction and geometric optimization for drogue-pose measurement [20]. AeroNet combines efficient object detection with a cooperative target for relative localization [43]. DeepArUco++ improves fiducial-marker detection under challenging lighting conditions while retaining the known geometric structure of the marker [47]. These methods indicate that learning is particularly effective for improving feature extraction, whereas the final pose should remain constrained by camera geometry and known target structure.

For fixed-wing wingtip docking, the visual front end should provide more than a pose estimate. It should also output the number and spatial distribution of valid features, target image area, reprojection error, image-blur score, occlusion ratio, and possible pose ambiguity. These quantities can be used by the fusion estimator to adapt the measurement covariance and reject unreliable observations.

3.2 Multi-Sensor Fusion and Relative-State Estimation

A single visual measurement cannot normally provide the temporal continuity and failure tolerance required for fixed-wing docking. Visual measurements have bounded long-term drift but may be delayed, intermittent, or affected by outliers. Inertial sensors provide high-rate motion information but accumulate drift. GNSS/RTK, UWB, inter-UAV communication, and flight-controller states can extend the sensing range and improve observability, but they introduce coordinate-alignment and timing problems. Multi-sensor fusion is therefore required to produce a continuous and control-oriented relative state.

3.2.1 Relative-State Formulation for Two Moving Aircraft

Most conventional visual-inertial odometry systems estimate the motion of a single sensor platform relative to a static environment [60-62]. Wingtip docking instead involves two independently moving vehicles. The relative state should therefore represent the motion of the target UAV with respect to the host UAV.

A control-oriented relative state can be defined as

$$\boldsymbol{x}_{rel} = \left[\left(\boldsymbol{p}_{rel}^{B_h}\right)^T \quad \left(\boldsymbol{v}_{rel}^{B_h}\right)^T \quad \boldsymbol{q}_{rel}^T \quad \boldsymbol{b}_a^T \quad \boldsymbol{b}_g^T \right]^T. \tag{20}$$

where $\mathbf{p}_{\mathrm{rel}}^{B_h}$ and $\mathbf{v}_{\mathrm{rel}}^{B_h}$ are the relative position and velocity expressed in the host body frame, $\mathbf{q}_{\mathrm{rel}}$ is the relative-attitude quaternion, and $\mathbf{b}_a$ and $\mathbf{b}_g$ denote the relevant accelerometer and gyroscope biases.

When both aircraft share inertial and flight-controller states, the relative motion can be propagated using information from both platforms. If only the host UAV carries a camera and an IMU, the motion of the host camera must be removed from the observed target motion. Otherwise, a rotation of the host aircraft may be incorrectly interpreted as target motion. The relative-position dynamics expressed in the rotating host frame can be written in simplified form as

$$\dot{\boldsymbol{p}}_{rel}^{B_h} = \boldsymbol{v}_{rel}^{B_h} - \boldsymbol{\omega}_h^{B_h} \times \boldsymbol{p}_{rel}^{B_h}, \tag{21}$$

where $\boldsymbol{\omega}_h^{B_h}$ is the angular velocity of the host body frame. The second term represents apparent relative motion induced by host-frame rotation. Neglecting this term may produce significant errors during the roll and yaw maneuvers required for lateral alignment.

Dual-quaternion formulations provide a unified representation of relative translation and rotation. Lee and Sung developed a cooperative relative-navigation method based on dual quaternions and monocular visual-inertial integration [49]. Such formulations are attractive for docking because they avoid treating translation and attitude as completely independent quantities.

3.2.2 Loosely Coupled Fusion

In a loosely coupled architecture, the visual front end independently estimates a relative position or pose. The result is then used as a measurement by an extended Kalman filter, unscented Kalman filter, or another state estimator. Kayhani et al. fused AprilTag pose measurements with inertial data using an on-manifold extended Kalman filter [50]. Causa et al. combined visual line-of-sight and range information with onboard navigation data in a customized filtering framework [6].

For an extended Kalman filter, the innovation associated with a visual measurement can be written as

$$\boldsymbol{r}_k = \boldsymbol{z}_k - h(\hat{\boldsymbol{x}}_k^-), \tag{22}$$

with innovation covariance

$$\boldsymbol{S}_k = \boldsymbol{H}_k \boldsymbol{P}_k^- \boldsymbol{H}_k^T + \boldsymbol{R}_k, \tag{23}$$

where $\mathbf{z}_k$ is the visual measurement, $h(\cdot)$ is the nonlinear measurement model, $\mathbf{H}_k$ is its Jacobian, $\mathbf{P}_k^-$ is the prior covariance, and $\mathbf{R}_k$ is the visual-measurement covariance.

Loosely coupled fusion is modular and facilitates integration with an existing vision subsystem and autopilot. Its main limitation is that image-level feature geometry is compressed into a single pose estimate. If the pose estimate is degraded by an ambiguous planar solution, a mistracked corner, or poor feature geometry, the filter may not have sufficient information to identify the source of the error.

3.2.3 Tightly Coupled Fusion

In a tightly coupled architecture, feature coordinates, bearing vectors, or reprojection residuals are directly incorporated into the estimator. The filter or optimization process jointly estimates navigation states, inertial biases, and sometimes sensor extrinsic and temporal parameters.

VINS-Mono employs inertial preintegration and sliding-window nonlinear optimization [61]. ORB-SLAM3 uses maximum-a-posteriori estimation for visual and visual-inertial states [60]. OpenVINS provides an error-state filtering framework with online calibration capabilities [62]. Visual-inertial teach-and-repeat, MIMC-VINS, DC-VINS, and IC-GVINS further address embedded

ISSN: 2414-1895 DOI: 10.6919/ICJE.202608_12(8).0007

implementation, asynchronous multi-sensor fusion, changing extrinsics, and integration with GNSS [63-66].

These systems provide important methodological foundations but are generally formulated for a single vehicle moving through a static environment. Applying them to wingtip docking requires replacing static landmarks with features attached to a moving target and introducing target motion into the state model. Directly treating a marker on the target UAV as a static environmental landmark would violate the underlying model assumptions.

Tightly coupled fusion preserves image-level information and can improve robustness when only part of the target is visible. It also enables direct estimation of camera-IMU extrinsics and time offsets. However, it increases computational complexity, implementation difficulty, and sensitivity to model inconsistency.

3.2.4 Fusion with GNSS, UWB, and Inter-UAV Information

GNSS or RTK-GNSS can provide long-range relative position when both aircraft exchange their global navigation states. However, subtracting two absolute navigation solutions does not automatically yield an accurate relative state if the measurements have different timestamps, reference frames, or correlated errors.

UWB provides direct relative range and is less dependent on lighting conditions. UWB-vision docking and UWB-assisted visual-inertial navigation have demonstrated that range measurements can improve scale and reduce drift in GNSS-denied environments [54,55]. Distributed UWB/vision/INS fusion has also been investigated for micro-UAV cooperative localization [251]. A scalar UWB measurement between transceivers can be modeled as

$$z_{uwb} = \|\boldsymbol{p}_{U_t}^{I} - \boldsymbol{p}_{U_h}^{I}\| + n_{uwb}, \tag{24}$$

where $\mathbf{p}_{U_h}^{I}$ and $\mathbf{p}_{U_t}^{I}$ are the transceiver positions and n_{uwb} is measurement noise. When the UWB antennas are not located at the aircraft centers of mass, antenna lever arms and body attitudes must be included in the measurement model.

Inter-UAV communication can provide the target vehicle’s GNSS, attitude, velocity, air-data, and health information. This information can significantly improve relative-state propagation, but delayed packets and packet loss must be considered. The target state should be propagated from its transmission time to the current estimation time before fusion. Cooperative-navigation studies have shown that communication topology, correlated information, and failed nodes influence estimation consistency [56-59]. Although wingtip docking normally involves only two vehicles, double counting may still occur if the same navigation information is exchanged and repeatedly reused by both estimators.

3.2.5 Time Synchronization and Delayed Measurements

Accurate spatial estimation does not guarantee accurate relative navigation if the measurements are temporally inconsistent. Camera exposure, image processing, IMU sampling, autopilot output, UWB ranging, and remote-data reception may all occur at different times.

Let a visual measurement acquired at time t_{k-d} become available at the current time t_k. It should be modeled as

$$\boldsymbol{z}_k = h(\boldsymbol{x}_{k-d}) + \boldsymbol{n}_k. \tag{25}$$

rather than as a measurement of the current state $\mathbf{x}_k$. Applying the delayed measurement directly to the current state creates an error approximately proportional to the relative velocity, angular rate, and delay duration.

Possible solutions include hardware triggering, timestamp synchronization, historical-state buffering, out-of-sequence measurement updates, state interpolation, time-offset augmentation, and sliding-window optimization. The selection depends on the available computational resources and the maximum expected delay.

3.2.6 Adaptive Covariance and Outlier Rejection

Visual-measurement quality changes continuously with target distance, viewing angle, image scale, blur, and occlusion. A constant measurement covariance cannot accurately represent these conditions. The visual covariance can be adjusted according to observable quality indicators:

$$\boldsymbol{R}_v = f\left(d, \alpha, A_{img}, e_{rep}, q_{blur}, q_{occ}, N_{feat}\right). \tag{26}$$

where d is the target distance, α is the viewing angle, A_{img} is the target image area, e_{rep} is the reprojection error, q_{blur} and q_{occ} describe blur and occlusion, and N_{feat} is the number of valid features.

Innovation-based gating can be performed using the Mahalanobis distance

$$\gamma_k = \boldsymbol{r}_k^T \boldsymbol{S}_k^{-1} \boldsymbol{r}_k. \tag{27}$$

The measurement can be rejected when

$$\gamma_k > \chi^2_{v,\alpha}, \tag{28}$$

where $\chi^2_{v,\alpha}$ is the chi-square threshold for measurement dimension v and confidence level α. A practical system may classify visual measurements into normal-update, down-weighted-update, rejected-update, and tracking-loss modes. During short visual interruptions, the relative state can be propagated using inertial and motion-model information, but the covariance should increase. Once uncertainty exceeds a safety threshold, the docking controller should reduce the closing speed or terminate the approach.

3.3 Methodological Development and Remaining Challenges

Relative-navigation methods for aerial docking are evolving from isolated visual-pose estimation toward integrated, uncertainty-aware relative-state estimation. Several development trends can be identified. First, the measurement target is shifting from the camera-to-marker pose toward the relative state between the actual mechanical interfaces, which requires explicit calibration and compensation of camera, body, marker, antenna, and docking-point lever arms. Second, visual perception is moving from purely geometric or purely learning-based methods toward hybrid architectures in which learning improves target and keypoint detection while geometric optimization maintains physical consistency. Third, the estimation model is shifting from single-platform navigation toward dual-moving-platform relative dynamics, requiring explicit consideration of host-frame rotation, target acceleration, asynchronous shared states, and correlated navigation information. Fourth, performance evaluation is expanding beyond average position and attitude errors to include latency, output frequency, covariance consistency, outlier rate, visual-loss duration, state-recovery time, and the influence of estimation errors on terminal docking safety.

A suitable relative-navigation architecture for fixed-wing wingtip docking may use GNSS/INS and inter-UAV communication for long-range rendezvous, aircraft-level visual detection together with UWB or RTK information for medium-range approach, and cooperative wingtip markers with visual-inertial fusion for precise alignment. The estimator should output relative docking-point position, velocity, attitude, covariance, and health status to the controller.

Several unresolved issues remain. The relative observability of two moving fixed-wing vehicles under nearly parallel flight has not been sufficiently characterized. Low relative excitation may make some translational, attitude, bias, or time-offset states weakly observable. Near-field aerodynamic disturbances may invalidate simple constant-velocity or constant-acceleration target models. Visual measurements can also become highly non-Gaussian under planar-pose ambiguity, partial occlusion, or target loss. Addressing these problems requires coordinated development of motion models, measurement-quality assessment, active observation trajectories, and uncertainty-aware control.

4. Guidance and Closed-Loop Control for Fixed-Wing Aerial Docking

The objective of a docking controller is to drive the two physical docking interfaces into the admissible capture region while maintaining target visibility, the fixed-wing flight envelope, collision avoidance, and sufficient control authority. Compared with conventional trajectory tracking, the controlled variables are not limited to the aircraft centers of mass. The controller must regulate the relative position, velocity, and attitude of the actual wingtip docking points and coordinate the transition from coarse approach to constrained terminal closure. The relevant literature can be organized into position-based visual servoing, image-based visual servoing, robust and prescribed-performance control, model predictive and safety-critical control, and learning-based compensation. These approaches differ in their feedback variables, model dependence, constraint-handling capability, and compatibility with existing autopilots.

4.1 Control Objectives and Command Interfaces

Let the relative position of the target docking point expressed in the host docking frame be $\mathbf{p}_{D_t}^{D_h}$, and let the desired relative position be $\mathbf{p}_{D_t,d}^{D_h}$. The docking-point position error is defined as

$$\boldsymbol{e}_p = \boldsymbol{p}_{D_t}^{D_h} - \boldsymbol{p}_{D_t,d}^{D_h}. \tag{29}$$

Let the relative docking attitude be $\mathbf{R}_{D_t}^{D_h}$, and let the desired relative attitude be $\mathbf{R}_d$. A compact attitude-error vector can be written as

$$\boldsymbol{e}_R = Log\left(\boldsymbol{R}_d^T \boldsymbol{R}_{D_t}^{D_h}\right)^{\vee}. \tag{30}$$

The relative-velocity error is

$$\boldsymbol{e}_v = \boldsymbol{v}_{rel}^{D_h} - \boldsymbol{v}_{rel,d}^{D_h}. \tag{31}$$

During terminal docking, the desired lateral and vertical components of $\mathbf{v}_{\mathrm{rel},d}^{D_h}$ are normally close to zero, whereas the desired longitudinal component is a bounded closing velocity. This closing command should decrease when the alignment error, state-estimation uncertainty, or predicted collision risk increases. Therefore, the terminal objective is not instantaneous zero error, but entry into and continued residence within the mechanical capture envelope.

An admissible capture set can be defined as

$$\mathcal{D}_{cap} = \left\{\boldsymbol{x}_{rel} \mid \|\boldsymbol{e}_p\| \le e_{p,max},\ \|\boldsymbol{e}_R\| \le e_{R,max},\ |v_{cl}| \le v_{cl,max}\right\}. \tag{32}$$

Successful terminal closure requires the relative state to remain in $\mathcal{D}_{\text{cap}}$ for a prescribed interval rather than merely crossing it at one instant. The allowable limits should be determined jointly from the mechanical capture range, impact tolerance, navigation uncertainty, and controller response. This formulation also prevents a method from being judged only by its minimum instantaneous position error.

The controller interface strongly affects implementation. Some autonomous aerial-refueling studies directly command control-surface deflections and thrust, allowing the controller to exploit a complete nonlinear aircraft model 26, 27]. Although this approach can provide high bandwidth, it depends heavily on model accuracy and requires modification of the flight-control system. For experimental fixed-wing platforms, a more practical architecture is to retain the existing inner-loop autopilot and let the docking controller generate high-level commands such as desired airspeed, roll angle, pitch angle, heading, flight-path angle, or virtual path points [17, 37]. A typical high-level command vector is

$$\boldsymbol{u}_c = [V_{a,d} \quad \phi_d \quad \theta_d \quad \psi_d]^T. \tag{33}$$

The realized aircraft response can be represented in a general form as

$$\dot{\boldsymbol{x}}_a = f_a(\boldsymbol{x}_a, \boldsymbol{u}_c, \boldsymbol{d}_a), \tag{34}$$

where $\mathbf{x}_a$ denotes the aircraft state and $\mathbf{d}_a$ represents aerodynamic disturbances and model uncertainty. The outer loop must account for inner-loop bandwidth, response delay, command saturation, and cross-axis coupling. A controller that assumes instantaneous tracking of $\mathbf{u}_c$ may become oscillatory or unsafe during terminal closure.

A phase-dependent structure is particularly suitable for fixed-wing docking. During rendezvous and coarse approach, the controller emphasizes path convergence and velocity coordination. During precise alignment, lateral, vertical, roll, pitch, and yaw errors become dominant. During terminal closure, longitudinal motion is enabled only after alignment and confidence conditions have been verified. A conceptual confidence-dependent closing command can be expressed as

$$v_{cl,d} = -v_{max}\sigma_p\big(\|\boldsymbol{e}_{p,\perp}\|\big)\sigma_R(\|\boldsymbol{e}_R\|)\sigma_P\big(\lambda_{max}(\boldsymbol{P}_{rel})\big), \tag{35}$$

where $\mathbf{e}_{p,\perp}$ contains the lateral and vertical position errors, and σ_p, σ_R, and σ_P are bounded scheduling functions. This expression is a system-design concept rather than an established standard law: its purpose is to show that terminal closure should automatically diminish when alignment quality or state confidence deteriorates.

4.2 Main Guidance and Control Paradigms

4.2.1 Position-Based Visual Servoing

Position-based visual servoing (PBVS) reconstructs the three-dimensional relative position and attitude and then regulates errors in physical space. The method is compatible with the relative-state output of an EKF, UKF, or optimization-based estimator and facilitates the definition of pre-docking points, approach corridors, collision boundaries, and closing-speed constraints. Fixed-wing moving-recovery systems commonly use line-of-sight guidance, virtual target points, or relative trajectories to convert spatial errors into heading, roll, altitude, and airspeed commands [37]. The heterogeneous recovery system of Zheng et al. used GNSS and NMPC for coarse convergence and an ArUco-derived pose with proportional-derivative corrections for terminal alignment; the dynamic part was demonstrated in AirSim, while the mechanism was evaluated by a ground static docking experiment

ISSN: 2414-1895 DOI: 10.6919/ICJE.202608_12(8).0007

[36]. Its terminal visual loop is more appropriately classified as PBVS than as image-based visual servoing because the feedback variable is reconstructed pose rather than raw image error.

PBVS provides direct physical interpretation and makes it relatively straightforward to impose three-dimensional safety constraints. Its accuracy, however, depends on camera calibration, marker geometry, lever-arm calibration, and pose-estimation consistency. A small but persistent pose bias can be converted directly into a steady docking-point offset. Planar-pose ambiguity or a delayed pose update may also generate abrupt command changes. Consequently, PBVS should normally use fused states, filtered derivatives, and explicit validity checks instead of differentiating frame-by-frame pose estimates.

For fixed-wing platforms, the mapping from position error to control command must respect underactuation. Lateral error is usually corrected through roll and heading, vertical error through pitch or flight-path angle, and longitudinal error through airspeed and path scheduling. Because roll and yaw corrections simultaneously displace the wingtip through the lever arm, a controller designed only around center-of-mass kinematics may create additional docking-point error. A suitable PBVS formulation should therefore compute the error at the docking interface and include inner-loop dynamics in the command-generation model.

4.2.2 Image-Based and Hybrid Visual Servoing

Image-based visual servoing (IBVS) uses image features directly as feedback. Let $\mathbf{s}$ denote the current feature vector and $\mathbf{s}_d$ the desired feature vector. The feature error and its kinematics are

$$\boldsymbol{e}_s = \boldsymbol{s} - \boldsymbol{s}_d, \qquad \dot{\boldsymbol{e}}_s = \boldsymbol{L}_s \boldsymbol{v}_c + \frac{\partial s}{\partial t}, \tag{36}$$

where $\mathbf{L}_s$ is the image interaction matrix, $\mathbf{v}_c$ is the camera spatial velocity, and $\partial \mathbf{s}/\partial t$ represents target-feature motion that is not caused by the host camera. For a moving target UAV, this additional term cannot be neglected.

IBVS can reduce sensitivity to some camera-to-body and target-model errors because the loop directly drives the observed image toward a desired appearance. Quan et al. formulated an image-based docking controller for autonomous aerial refueling [25], and Liu et al. further considered a velocity-control interface, autopilot dynamics, camera installation error, wake turbulence, and gust disturbances [17]. These studies indicate that an image-space controller can be connected to a fixed-wing autopilot without directly replacing the inner loop. Their reported validation, however, is numerical rather than a complete fixed-wing docking flight experiment.

Pure IBVS has several limitations for wingtip docking. Image-feature convergence does not by itself guarantee a safe three-dimensional separation, and feature depth is required by the interaction matrix. The target may also leave the field of view under large initial error or aggressive maneuvering. In addition, two different physical configurations may produce similar image features under unfavorable geometry. For these reasons, a hybrid architecture is generally more suitable: fused three-dimensional states are used for approach, collision avoidance, and flight-envelope management, while image errors are used for terminal bias correction and visibility regulation. The transition between PBVS and IBVS should be scheduled continuously to avoid a command discontinuity.

4.2.3 Robust, Adaptive, and Prescribed-Performance Control

Aerodynamic disturbances, model mismatch, actuator dynamics, and target maneuvers motivate the use of robust and adaptive control. A general uncertain docking model can be written as

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{g}(\boldsymbol{x})\boldsymbol{u} + \boldsymbol{d}(\boldsymbol{x}, t), \tag{37}$$

where $\mathbf{d}(\mathbf{x}, t)$ collects wake effects, gusts, aerodynamic coupling, unmodeled dynamics, and input disturbances. Backstepping, sliding-mode observers, adaptive dynamic-surface control, and neural approximators have been used to compensate for such uncertainty in aerial-refueling and aerial-recovery problems [26, 27, 29-32]. Su et al. combined backstepping with a high-order sliding-mode disturbance observer for exact docking control [26], whereas Wu et al. used adaptive dynamic-surface control with neural approximation of uncertain nonlinearities [27]. These methods provide formal stability arguments, but their performance depends on disturbance assumptions, gain selection, and the quality of the underlying model.

Prescribed-performance control constrains the tracking error within a predefined time-varying envelope. For a scalar error $e(t)$, the requirement can be expressed as

$$-\underline{\delta}\rho(t) < e(t) < \overline{\delta}\rho(t), \qquad \rho(t) = (\rho_0 - \rho_\infty)e^{-l} + \rho_\infty, \tag{38}$$

where $\rho(t)$ specifies the allowable transient and steady-state error bounds. Direct probe control, neuroadaptive prescribed-performance control, and flexible prescribed-performance recovery methods have been developed to limit overshoot and enforce convergence requirements under wind disturbances and carrier maneuvers [10, 28, 30, 31]. Predefined-time control further attempts to make the convergence-time upper bound independent of the initial condition [12]. The latter study was supported by six-degree-of-freedom numerical simulations and should not be interpreted as hardware-in-the-loop or flight validation.

These methods are attractive for docking because the capture mechanism naturally defines admissible error envelopes. Nevertheless, mathematical error bounds should not be equated automatically with physical safety. The controller must also consider state-estimation uncertainty, actuator saturation, minimum airspeed, image visibility, and the geometry of both wings. Aggressive gains chosen to satisfy a tight theoretical envelope may cause high-frequency commands, structural load, or visual degradation.

4.2.4 Model Predictive and Safety-Critical Control

Model predictive control (MPC) is well suited to multi-variable fixed-wing docking because it can optimize future motion while explicitly handling state and input constraints. A finite-horizon formulation can be written as

$$\min_{\boldsymbol{u}_{0:N-1}} \sum_{k=0}^{N-1} \left(\left\| \boldsymbol{x}_k - \boldsymbol{x}_{k,d} \right\|_{\boldsymbol{Q}}^2 + \left\| \boldsymbol{u}_k \right\|_{\boldsymbol{R}}^2 \right) + \left\| \boldsymbol{x}_N - \boldsymbol{x}_{N,d} \right\|_{\boldsymbol{Q}_f}^2. \tag{39}$$

The optimization is subject to $\mathbf{x}_{k+1} = \mathbf{F}(\mathbf{x}_k, \mathbf{u}_k)$, $\mathbf{x}_k \in \mathcal{X}$, and $\mathbf{u}_k \in \mathcal{U}$. The constraint set $\mathcal{X}$ can include minimum airspeed, attitude limits, field-of-view boundaries, collision-avoidance regions, alignment corridors, and terminal capture conditions. The set $\mathcal{U}$ can represent command and actuator limits. SideArm recovery research has used constrained nonlinear control to regulate the approach trajectory and closing velocity under maritime disturbances [11]. The results were numerical simulations under different initial and environmental conditions, rather than a physical wingtip-docking experiment. Safety-critical MPC has also been explored for net recovery of fixed-wing UAVs, with minimum-airspeed and unsafe-region constraints represented through control barrier functions [13].

For a continuously differentiable safety function $h(\mathbf{x})$, the safe set can be defined as $\mathcal{C} = \{\mathbf{x} \mid h(\mathbf{x}) \geq 0\}$. A standard control-barrier condition is

$$\dot{h}(\boldsymbol{x}, \boldsymbol{u}) + \alpha\big(h(\boldsymbol{x})\big) \geq 0, \tag{40}$$

where $\alpha(\cdot)$ is an extended class-$\mathcal{K}$ function. For wingtip docking, multiple barrier functions may be required to describe minimum airspeed, wing collision boundaries, target visibility, and actuator margins. A hierarchical safety supervisor can override the nominal controller and issue hold, retreat, or abort commands when these conditions cannot be maintained.

The main challenge of MPC is model and computation dependence. A prediction model based on isolated-aircraft aerodynamics may become inaccurate in the near-field interaction region. Online optimization must also complete within a deterministic time budget. Practical implementation may therefore require a reduced-order prediction model, warm starting, explicit constraint prioritization, and a verified fallback controller. Incorporating state covariance or chance constraints is an important direction because deterministic constraints based only on the estimated mean can become unsafe when visual uncertainty increases.

4.2.5 Learning-Based Compensation and Iterative Methods

Learning methods are mainly used to predict target motion, approximate unknown disturbances, tune control parameters, or improve performance over repeated docking attempts. Terminal iterative learning control uses the terminal error from previous trials to modify the reference or feedforward command for subsequent trials [68]. This approach is useful for repeatable aerodynamic disturbances but is less effective when the target maneuver or wind field changes significantly between attempts.

Neural networks and reinforcement learning have been combined with model-based controllers for aerial recovery and refueling [27, 29, 30, 32]. A generic composite control command can be written as

$$\boldsymbol{u} = \boldsymbol{u}_{model} + \boldsymbol{u}_{learn} + \boldsymbol{u}_{safe}, \tag{41}$$

where $\mathbf{u}_{\text{model}}$ is generated by the nominal physical controller, $\mathbf{u}_{\text{learn}}$ compensates for unknown or repeatable effects, and $\mathbf{u}_{\text{safe}}$ enforces hard constraints or supervisory actions. This structure is preferable to an unconstrained end-to-end policy because it separates nominal stabilization, learned compensation, and safety protection.

Learning-based methods face several difficulties in wingtip docking. Flight data containing near-collision and failed-contact cases are scarce, simulation-to-flight differences are substantial in near-field aerodynamics, and distribution shifts may occur because of lighting, aircraft configuration, wind, or mechanism wear. Therefore, a learned component should provide bounded compensation or prediction together with confidence information. Its output should be monitored by a model-based safety layer, and performance should be evaluated under out-of-distribution and sensor-failure conditions rather than only in nominal simulations.

4.3 Comparative Analysis and Development Trends

The available control paradigms address different parts of the docking problem. PBVS provides direct physical interpretation and supports spatial constraints, but it is sensitive to pose bias and calibration error. IBVS directly regulates image features and may reduce some model dependence, but it requires depth information and does not automatically guarantee three-dimensional safety. Robust and adaptive methods improve disturbance tolerance but may rely on conservative uncertainty bounds. Prescribed-performance methods provide explicit transient envelopes, yet aggressive envelopes can conflict with actuator, visibility, and structural limitations. MPC and barrier-function methods are particularly suitable for flight-envelope and collision constraints, although their reliability depends on prediction-model fidelity and real-time computation. Learning and iterative methods can compensate for repeatable or unknown effects but require explicit safeguards against data scarcity and distribution shift.

Several development trends are apparent. First, the controlled object is moving from the aircraft center of mass toward the actual probe, recovery tip, or docking interface [26, 28, 32, 40]. This shift

ISSN: 2414-1895 DOI: 10.6919/ICJE.202608_12(8).0007

is essential for wingtip docking because attitude motion is amplified by the lever arm. Second, high-level velocity, path, and attitude interfaces are becoming more attractive than direct control-surface commands because they allow the docking function to be added above a flight-proven autopilot [17, 37]. Third, the control objective is expanding from asymptotic convergence toward prescribed transient performance, flight-envelope compliance, collision avoidance, and reliable abort behavior [10-13, 28-32]. Fourth, PBVS and IBVS are increasingly viewed as complementary layers rather than mutually exclusive alternatives. Fifth, learning is moving toward model augmentation and prediction instead of unrestricted replacement of the physical controller.

A major unresolved issue is the weak connection between navigation uncertainty and control action. Many controllers assume accurate, continuous, and delay-free relative states, while actual vision-based estimates contain range-dependent covariance, outliers, and temporary loss. A credible terminal controller should use the estimated covariance, innovation statistics, target visibility, and time since the last valid observation to schedule gains and closing speed. When confidence falls below a threshold, the correct action may be to hold or retreat rather than to increase control effort.

Another limitation is the treatment of near-field aerodynamics. Most docking controllers model aerodynamic influence as a bounded disturbance, whereas wingtip wind-tunnel results indicate that the induced moments vary systematically with relative position [4]. A useful future architecture may combine a position-dependent low-order aerodynamic model with robust feedback. Such a model can provide feedforward compensation or prediction within MPC, while the robust layer accommodates residual uncertainty.

The most suitable system architecture for fixed-wing wingtip docking is therefore likely to be hierarchical. A mission manager selects rendezvous, approach, alignment, terminal-closure, hold, retreat, or abort modes. A safety supervisor monitors minimum airspeed, field of view, collision boundaries, actuator limits, and navigation confidence. A guidance layer generates a visibility-preserving reference trajectory, and an outer-loop controller produces feasible airspeed, path, or attitude commands for the existing autopilot. Terminal closure is authorized only when position, attitude, velocity, and confidence conditions are satisfied. This architecture does not prescribe a single controller; rather, it defines how PBVS, IBVS, MPC, robust compensation, and learning modules can be combined without losing system-level safety.

5. Integrated Discussion and Future Directions

The preceding sections indicate that autonomous wingtip docking cannot be developed as a simple sequence of perception, state estimation, and control modules. Visual measurements are influenced by the maneuvers commanded by the controller, while the resulting estimation uncertainty should in turn affect closure speed, mode transitions, and abort decisions. As the separation decreases, near-field aerodynamic interaction, mechanical capture, and the transition to connected flight become increasingly important. Future progress therefore depends on coordinated system design rather than isolated improvements in nominal pose accuracy or tracking convergence.

5.1 Perception-Estimation-Control Coupling and Uncertainty-Aware Autonomy

Perception and control are bidirectionally coupled during terminal docking. Large roll or heading commands may rapidly reduce lateral error, but can also move the target marker toward the image boundary, increase perspective distortion, reduce the usable marker area, and introduce motion blur. A high closing speed can improve image resolution as the target becomes larger, yet it also reduces the time available for rejecting abnormal measurements and executing a safe retreat. Studies on air-to-air visual navigation, high-speed visual interception, and field-of-view-constrained visual servoing show that observation geometry, target motion, and processing delay can materially influence closed-loop performance [6-9]. These results do not directly validate fixed-wing wingtip docking, but they provide a strong methodological basis for visibility-aware guidance.

A reliable docking controller should therefore use more than the estimated relative state. The navigation system should also provide covariance, innovation consistency, target image size, reprojection error, blur and occlusion indicators, tracking status, and sensor-health information. These quantities allow the controller to distinguish between a geometrically aligned state supported by reliable measurements and an apparently aligned state produced during visual degradation. When uncertainty increases, the system should reduce closure speed, maintain separation, or return to the pre-docking region instead of continuing with unchanged commands.

An appropriate system architecture should contain three complementary layers. The nominal docking controller generates the desired approach motion. A constraint-handling layer modifies the command to preserve minimum airspeed, field of view, collision clearance, actuator margins, and the mechanical capture envelope. An independent mission supervisor monitors sensor health, estimator consistency, communication status, and the feasibility of continued approach. This supervisor should be capable of commanding hold, retreat, re-alignment, and abort actions without depending on the nominal controller remaining fully functional.

Active perception is another important direction. Rather than treating visual quality as an uncontrollable disturbance, the guidance system can select trajectories that maintain favorable observation geometry. Candidate criteria include the predicted target image area, distance from the image boundary, distribution of visible features, pose-estimation conditioning, and expected uncertainty reduction. Such observation-aware planning is particularly relevant during nearly parallel flight, where limited relative excitation may make depth, attitude, sensor bias, and timing errors weakly observable.

The most important implication is that estimation confidence should participate directly in task progression. Mode switching should not depend only on estimated distance or alignment error. Terminal closure should be enabled only when geometric alignment, covariance consistency, target visibility, airspeed margin, communication status, and available control authority are simultaneously acceptable. This confidence-aware logic provides a practical link between relative navigation and safety-critical control.

5.2 Near-Field Aerodynamics, Contact Transition, and System Co-Design

Most existing docking controllers either neglect the aerodynamic influence of the other aircraft or represent it as a bounded external disturbance. Direct wind-tunnel experiments, however, show that the rolling moment acting on a fixed-wing UAV during close wingtip approach varies systematically with longitudinal, lateral, and vertical separation [4]. Rapid aerodynamic prediction methods based on aerodynamic theory, experimental data, and parameter correction offer a possible route toward real-time estimation of these interaction effects [69]. The available evidence remains configuration-dependent and does not yet provide a general six-degree-of-freedom model for two maneuvering fixed-wing UAVs.

A realistic development path is likely to combine several levels of aerodynamic treatment. Robust feedback can provide a baseline tolerance to unmodeled interaction. Reduced-order models derived from wind-tunnel or computational-fluid-dynamics data can support feedforward compensation and gain scheduling. Online parameter adaptation or disturbance estimation can then correct residual mismatch. This layered approach is less dependent on a perfect physical model than purely model-based control and provides clearer safety boundaries than an unrestricted black-box predictor.

Mechanical capture introduces a second major transition. Before contact, the aircraft are independent rigid bodies coupled through sensing, communication, aerodynamic interaction, and control. During capture, intermittent contact, mechanism compliance, impact, misalignment correction, and possible rebound may occur. After locking, the system becomes a connected multibody aircraft with shared structural loads and altered stability characteristics [3]. These three operating conditions require different models, objectives, and safety criteria, and should be coordinated through explicit mode-transition logic.

Related research on air-ground robot combination and modular multirotor docking has demonstrated the value of explicitly modeling contact regions, capture forces, mechanical constraints, visual-inertial navigation, and connected control [70-72]. These studies provide useful architectural concepts but cannot be transferred directly to fixed-wing platforms. Multirotors can hover and generate relatively independent translational motion, whereas fixed-wing vehicles must maintain forward airspeed and use attitude changes to create lateral and vertical motion. Contact-transition strategies must therefore be reformulated around fixed-wing flight-envelope and aerodynamic constraints.

Navigation, control, and docking-mechanism requirements should be established jointly. A larger mechanical capture range can relax sensing and tracking accuracy requirements, but may increase mass, drag, structural flexibility, and locking clearance. A compact mechanism reduces these penalties but demands more accurate calibration, lower latency, smaller tracking error, and tighter synchronization. Similarly, mechanical compliance can reduce impact loads but may introduce oscillation and uncertainty in the connected state. The design objective should therefore be system-level capture probability and safe load transfer rather than minimum error in any single subsystem.

Post-capture behavior should also be considered before terminal control is finalized. Connected-flight experiments indicate that a wingtip-connected configuration may possess stability characteristics substantially different from those of an isolated aircraft [3]. The capture mechanism, terminal attitude, residual relative velocity, and controller handover strategy all influence the initial condition of connected flight. A successful contact event is not sufficient if it produces excessive loads or an unstable post-connection state.

5.3 Validation Framework, Evaluation Criteria, and Research Priorities

The maturity of a docking method cannot be judged solely by whether a study reports an experiment. Numerical simulation, high-fidelity simulation, software-in-the-loop testing, model-in-the-loop testing, processor-in-the-loop testing, hardware-in-the-loop testing, ground dynamic testing, non-contact flight approach, physical aerial capture, and connected-flight testing provide fundamentally different levels of evidence. These levels should be reported explicitly rather than grouped under the broad term experimental validation.

A progressive validation route is recommended. Initial work should verify geometry, observability, estimator consistency, and nominal closed-loop behavior. Six-degree-of-freedom simulations should then incorporate wind, sensor noise, processing delay, actuator dynamics, communication loss, and uncertain aerodynamic interaction. Monte Carlo analysis and fault injection should be used to examine rare pose jumps, visual loss, inertial bias, GNSS degradation, timing errors, and actuator saturation. Software-, processor-, and hardware-in-the-loop tests should expose implementation and interface defects before real flight. Ground motion platforms can evaluate cameras, markers, fusion software, and capture mechanisms. Flight testing should progress from single-aircraft target observation to two-aircraft non-contact approach, low-risk physical capture, and finally connected-flight stabilization [73, 74].

Evaluation should cover the complete information and control chain. Perception reporting should include detection probability, feature accuracy, pose error, update rate, processing delay, outlier rate, and recovery time after target loss. Estimation reporting should include relative-state error, covariance consistency, drift during visual interruption, innovation statistics, and reinitialization behavior. Control reporting should include docking-point error, attitude error, closing speed, maximum overshoot, command smoothness, control effort, and approach duration. System-level reporting should include capture success rate, safe-abort rate, collision-boundary violations, visual-loss frequency, computational load, and the highest completed validation level.

A docking attempt should be declared successful only when the relative state remains within a clearly specified capture envelope for a defined duration and the mechanism reaches the required engagement condition. Reported success rates should state the number of trials, initial-condition distribution, wind and lighting conditions, sensing configuration, mechanical tolerance, and statistical confidence. Without these definitions, results from different platforms cannot be meaningfully compared.

Several priorities emerge for future research. Dual-platform relative observability should be characterized under nearly parallel fixed-wing motion. Visual uncertainty models should reflect range, viewing angle, feature geometry, blur, and occlusion. Closing speed and mode transitions should depend on both estimated error and confidence. Guidance should maintain visual observability while satisfying flight-envelope, aerodynamic, and collision constraints. Near-field aerodynamic models require validation across different wing geometries and operating conditions. Free flight, mechanical capture, and connected flight should be incorporated into a unified mode-management framework. Finally, shared datasets, repeatable benchmarks, and clearly defined validation levels are needed to support quantitative comparison and reproducible development.

6. Conclusion

Autonomous wingtip docking of fixed-wing UAVs is a transition from close cooperative flight to mechanical connection. Its difficulty arises from the simultaneous interaction of dual-moving-platform relative navigation, fixed-wing underactuation, the long lever arm between the center of mass and the docking interface, near-field aerodynamic interference, terminal collision risk, and the change in system dynamics after contact.

Direct wingtip-related research has established several important foundations. Connected multibody flight tests have shown that a wingtip-connected configuration can be maintained under dedicated control, while wind-tunnel experiments have identified position-dependent aerodynamic interaction during close approach. Compound-VTOL experiments have also demonstrated the integration of a wingtip mechanism, monocular localization, and flight control under a hovering-target condition . Nevertheless, these achievements do not yet constitute a complete autonomous demonstration between two continuously forward-flying fixed-wing UAVs.

Relative-navigation research provides the sensing and estimation basis for the task. Cooperative markers and learning-geometry visual methods can support precise local measurement, whereas visual-inertial, GNSS/RTK-assisted, and UWB-assisted fusion can improve continuity and operating range. However, conventional single-platform navigation cannot be transferred directly without accounting for both aircraft motions, lever arms, asynchronous measurements, correlated shared information, and relative-state observability. The navigation output should describe the physical docking interfaces and should include velocity, uncertainty, and health information in addition to pose.

Closed-loop control has progressed from center-of-mass trajectory tracking toward docking-point regulation, visual servoing, prescribed-performance control, model predictive control, safety constraints, and learning-based compensation. A practical system should combine fused three-dimensional states for approach and safety management, image information for terminal refinement and visibility preservation, an established autopilot for inner-loop stabilization, and an independent supervisor for hold, retreat, and abort decisions.

The central challenge is therefore not to improve perception accuracy or controller convergence in isolation. It is to construct an uncertainty-aware autonomous system that preserves visual observability, respects fixed-wing and collision constraints, compensates for near-field aerodynamic interaction, and manages the transition through capture into connected flight. Reliable wingtip docking will depend on integrated modeling, navigation-control-mechanism co-design, quantitative safety assessment, and progressively structured experimental validation.

References

[1] Parry, J., & Hubbard, S. (2023). Review of sensor technology to support automated air-to-air refueling of a probe configured uncrewed aircraft. Sensors, 23(2), 995.

[2] Ren, J., & Quan, Q. (2024). Progress in modeling and control of probe-and-drogue autonomous aerial refueling. Chinese Journal of Aeronautics, 37(5), 6–26.

[3] Meng, Y., An, C., Xie, C., et al. (2022). Conceptual design and flight test of two wingtip-docked multi-body aircraft. Chinese Journal of Aeronautics, 35(12), 144–155.

[4] Zhou, W., Ma, P., Wei, B., et al. (2023). Experimental study on aerodynamic characteristics of fixed-wing UAV air docking. Aerospace Science and Technology, 137, 108257.

[5] Huo, J., Liu, H., Huang, R., et al. (2026). Autonomous aerial docking for compound VTOL UAVs with thrust vectoring rotors. Mechanical Systems and Signal Processing, 244, 113782.

[6] Causa, F., Opromolla, R., & Fasano, G. (2022). Closed loop integration of air-to-air visual measurements for cooperative UAV navigation in GNSS challenging environments. Aerospace Science and Technology, 130, 107947.

[7] Yang, K., Bai, C., She, Z., et al. (2025). High-speed interception multicopter control by image-based visual servoing. IEEE Transactions on Control Systems Technology, 33(1), 119–135.

[8] Guo, H., Song, T., Ye, J., et al. (2026). Image-based visual servoing tracking control of a quadrotor with target estimation and field of view constraints. Chinese Journal of Aeronautics, 104354.

[9] Zhao, W., Liu, H., & Wang, X. (2021). Robust visual servoing control for quadrotors landing on a moving target. Journal of the Franklin Institute, 358(4), 2301–2319.

[10] Wang, Y., Wang, H., Zhu, J., et al. (2025). Flexible prescribed performance-based approaching control for UAV aerial recovery under multiwind disturbances. Chinese Journal of Aeronautics, 38(4), 103282.

[11] Su, Z., Xing, Z., Li, X., et al. (2024). Constrained docking control for shipborne UAV SideArm recovery. Chinese Journal of Aeronautics, 37(5), 39–59.

[12] Wu, Z., Wang, L., Zhang, Z., et al. (2026). Practical predefined-time control for stable and safe docking of fixed-wing receiver in aerial refueling. Aerospace Science and Technology, 112451.

[13] Esfahani, H. N., & Mohammadpour Velni, J. (2024). Safe net-recovery of fixed-wing unmanned aerial vehicles using safety-critical MPC. IFAC-PapersOnLine, 58(28), 324–329.

[14] Xin, L., Tang, Z., Gai, W., et al. (2022). Vision-based autonomous landing for the UAV: A review. Aerospace, 9(11), 634.

[15] Semerikov, S. O., Nechypurenko, P. P., Vakaliuk, T. A., et al. (2025). Vision-based autonomous UAV landing: A comprehensive review of technologies, techniques, and applications. Journal of Intelligent & Robotic Systems, 111, 115.

[16] Funke, A., Lück, V., & Glabeke, G. (2026). Robust control of an actively controlled drogue for autonomous aerial docking. CEAS Aeronautical Journal, 17, 677–693.

[17] Liu, R., Huang, Y., Ren, J., et al. (2026). Probe-and-drogue autonomous aerial refueling with a velocity control mode: An image-based visual servo method. Aerospace Science and Technology, 176, 112007.

[18] Ren, J., Du, J., Xu, B., et al. (2026). Multi-degree-of-freedom reliable docking control for probe-and-drogue autonomous aerial refueling. Aerospace Science and Technology, 168, 111086.

[19] Lin, P., He, Y., & Chen, Y. (2024). Unmanned autonomous air-to-air refueling intelligent docking technology. Chinese Journal of Aeronautics, 37(5), 1–5.

[20] Wen, S., Gao, Y., Hu, B., et al. (2025). AARPose: Real-time and accurate drogue pose measurement based on monocular vision for autonomous aerial refueling. Chinese Journal of Aeronautics, 38(6), 103307.

[21] Gong, K., Liu, B., Xu, X., et al. (2023). Research of an unmanned aerial vehicle autonomous aerial refueling docking method based on binocular vision. Drones, 7(7), 433.

[22] Gao, Y., Yang, S., & Liu, X. (2023). Drogue position measurement of autonomous aerial refueling based on embedded system. Sensors and Actuators A: Physical, 353, 114251.

[23] Ma, Y., Zhao, R., Liu, E., et al. (2019). A novel autonomous aerial refueling drogue detection and pose estimation method based on monocular vision. Measurement, 136, 132–142.

[24] Sun, Y., Deng, Y., Duan, H., et al. (2019). Bionic visual close-range navigation control system for the docking stage of probe-and-drogue autonomous aerial refueling. Aerospace Science and Technology, 91, 136–149.

[25] Quan, Q., Liu, R., Liu, H., et al. (2025). Image-based visual servo docking control for autonomous aerial refueling. Journal of Guidance, Control, and Dynamics, 48(7), 1689–1698.

[26] Su, Z., Wang, H., Li, N., et al. (2018). Exact docking flight controller for autonomous aerial refueling with back-stepping based high order sliding mode. Mechanical Systems and Signal Processing, 101, 338–360.

[27] Wu, J., Luo, H., & Ai, J. (2020). Docking controller for autonomous aerial refueling with adaptive dynamic surface control. IEEE Access, 8, 99846–99857.

[28] Su, Z., Liu, Y., & Wang, H. (2022). Probe dynamics direct control for aerial recovery with preassigned docking performance. IEEE Transactions on Aerospace and Electronic Systems, 58(4), 3509–3523.

[29] Liu, Y., Wang, H., Liu, B., et al. (2023). Learning-based compound docking control for UAV aerial recovery: Methodology and implementation. IEEE/ASME Transactions on Mechatronics, 28(3), 1706–1717.

[30] Wang, Y., Wang, H., Liu, Y., et al. (2023). Performance-guaranteed neuroadaptive docking control for UAV aerial recovery under carrier maneuvering flight. IEEE Transactions on Aerospace and Electronic Systems, 59(6), 8355–8368.

[31] Su, Z., Li, C., Wu, J., et al. (2022). Neuro-adaptive prescribed performance control for aerial-recovery drogue with actuator constraints. Journal of Guidance, Control, and Dynamics, 45(8), 1451–1465.

[32] Liu, Y., Wang, H., & Fan, J. (2019). Novel docking controller for autonomous aerial refueling with probe direct control and learning-based preview method. Aerospace Science and Technology, 94, 105403.

[33] Liu, Y., Zhao, Z., Ma, H., et al. (2021). A stochastic approximation method for probability prediction of docking success for aerial refueling. Applied Soft Computing, 103, 107139.

[34] Hang, B., Guo, P., Yan, S., et al. (2025). Robust docking control and safety evaluation of autonomous aerial refueling for unmanned aerial vehicles. Journal of the Franklin Institute, 362(10), 107736.

[35] Hang, B., Guo, P. (2026). Precise docking and safety prediction for unmanned aerial vehicle autonomous aerial refueling: An adaptive control and deep learning approach. Journal of the Franklin Institute, 363(9), 108693.

[36] Zheng, T., Richardson, T. S., & Meier, K. (2026). Design of an autonomous airborne recovery system: A fixed-wing UAV-quadrotor platform using improved NMPC and vision-based control. Drones, 10(3), 212.

[37] Gryte, K., Sollie, M. L., & Johansen, T. A. (2021). Control system architecture for automatic recovery of fixed-wing unmanned aerial vehicles in a moving arrest system. Journal of Intelligent & Robotic Systems, 103, 73.

[38] Wang, Y., Wang, H., Liu, B., et al. (2021). A visual navigation framework for the aerial recovery of UAVs. IEEE Transactions on Instrumentation and Measurement, 70, 5019713.

[39] Liu, Y., Wang, Y., Li, H., et al. (2024). Runway-free recovery methods for fixed-wing UAVs: A comprehensive review. Drones, 8(9), 463.

[40] Worth, D., Choate, J., Lynch, J., et al. (2024). Relative vectoring using dual object detection for autonomous aerial refueling. Neural Computing and Applications, 36, 10143–10163.

[41] Xu, X., Zhuge, S., Li, C., et al. (2023). A vision-only relative distance calculation method for multi-UAV systems. Aerospace Science and Technology, 142, 108665.

[42] Luo, H., Zhang, K., Su, Y., et al. (2022). Monocular vision pose determination-based large rigid-body docking method. Measurement, 204, 112049.

[43] Shen, K., Zhuang, Y., Chen, Y., et al. (2023). AeroNet: An efficient relative localization and object detection network for cooperative aerial-ground unmanned vehicles. Pattern Recognition Letters, 171, 28–37.

[44] Lepetit, V., Moreno-Noguer, F., & Fua, P. (2009). EPnP: An accurate O(n) solution to the PnP problem. International Journal of Computer Vision, 81(2), 155–166.

[45] Collins, T., & Bartoli, A. (2014). Infinitesimal plane-based pose estimation. International Journal of Computer Vision, 109(3), 252–286.

[46] Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F. J., et al. (2014). Automatic generation and detection of highly reliable fiducial markers under occlusion. Pattern Recognition, 47(6), 2280–2292.

[47] Berral-Soler, R., Muñoz-Salinas, R., Medina-Carnicer, R., et al. (2024). DeepArUco++: Improved detection of square fiducial markers in challenging lighting conditions. Image and Vision Computing, 152, 105313.

[48] Olson, E. (2011). AprilTag: A robust and flexible visual fiducial system. In 2011 IEEE International Conference on Robotics and Automation (pp. 3400–3407). IEEE.

ISSN: 2414-1895
DOI: 10.6919/ICJE.202608_12(8).0007

[49] Lee, B., & Sung, S. (2024). A study on dual quaternion based cooperative relative navigation of multiple UAVs with monocular vision-inertial integration. Chinese Journal of Aeronautics, 37(11), 335–354.

[50] Kayhani, N., Zhao, W., McCabe, B., et al. (2022). Tag-based visual-inertial localization of unmanned aerial vehicles in indoor construction environments using an on-manifold extended Kalman filter. Automation in Construction, 135, 104112.

[51] Xiong, C., Lu, W., Xiong, H., et al. (2024). Onboard cooperative relative positioning system for Micro-UAV swarm based on UWB/Vision/INS fusion through distributed graph optimization. Measurement, 234, 114897.

[52] Chang, Y., Cheng, Y., Manzoor, U., et al. (2023). A review of UAV autonomous navigation in GPS-denied environments. Robotics and Autonomous Systems, 170, 104533.

[53] Phan, T. D., & Kim, G. W. (2025). Toward specialized learning-based approaches for visual odometry: A comprehensive survey. Journal of Intelligent & Robotic Systems, 111, 44.

[54] Cheng, C., Li, X., Xie, L., et al. (2022). Autonomous dynamic docking of UAV based on UWB-vision in GPS-denied environment. Journal of the Franklin Institute, 359(7), 2788–2809.

[55] Lin, H. Y., & Zhan, J. R. (2023). GNSS-denied UAV indoor navigation with UWB incorporated visual inertial odometry. Measurement, 206, 112256.

[56] Li, J., Yang, G., Cai, Q., et al. (2022). Cooperative navigation for UAVs in GNSS-denied area based on optimized belief propagation. Measurement, 192, 110797.

[57] Xiong, J., Xiong, Z., Cheong, J. W., et al. (2020). Cooperative positioning for low-cost close formation flight based on relative estimation and belief propagation. Aerospace Science and Technology, 106, 106068.

[58] Wang, S., Zhan, X., Zhai, Y., et al. (2021). Performance estimation for Kalman filter based multi-agent cooperative navigation by employing graph theory. Aerospace Science and Technology, 112, 106628.

[59] Shen, J., Wang, S., & Zhan, X. (2022). Multi-UAV cluster-based cooperative navigation with fault detection and exclusion capability. Aerospace Science and Technology, 124, 107570.

[60] Campos, C., Elvira, R., Gómez Rodríguez, J. J., et al. (2021). ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multimap SLAM. IEEE Transactions on Robotics, 37(6), 1874–1890.

[61] Qin, T., Li, P., & Shen, S. (2018). VINS-Mono: A robust and versatile monocular visual-inertial state estimator. IEEE Transactions on Robotics, 34(4), 1004–1020.

[62] Geneva, P., Eckenhoff, K., Lee, W., et al. (2020). OpenVINS: A research platform for visual-inertial estimation. In 2020 IEEE International Conference on Robotics and Automation (pp. 4666–4672). IEEE.

[63] Nitsche, M., Pessacg, F., & Civera, J. (2020). Visual-inertial teach and repeat. Robotics and Autonomous Systems, 131, 103577.

[64] Eckenhoff, K., Geneva, P., & Huang, G. (2021). MIMC-VINS: A versatile and resilient multi-IMU multi-camera visual-inertial navigation system. IEEE Transactions on Robotics, 37(5), 1360–1380.

[65] Rebello, J., Li, C., & Waslander, S. L. (2021). DC-VINS: Dynamic camera visual inertial navigation system with online calibration. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (pp. 2559–2568). IEEE.

[66] Tang, H., Zhang, T., Niu, X., et al. (2023). IC-GVINS: A robust, real-time, INS-centric GNSS-visual-inertial navigation system. IEEE Robotics and Automation Letters, 8(1), 216–223.

[67] Tandale, M. D., Bowers, R., & Valasek, J. (2006). Trajectory tracking controller for vision-based probe and drogue autonomous aerial refueling. Journal of Guidance, Control, and Dynamics, 29(4), 846–857.

[68] Dai, X., Quan, Q., Ren, J., et al. (2018). Terminal iterative learning control for autonomous aerial refueling under aerodynamic disturbances. Journal of Guidance, Control, and Dynamics, 41(7), 1576–1583.

[69] Ma, P., Zhou, W., Han, S., et al. (2026). Rapid prediction of aerodynamic parameters changes for close-range wings based on feature parameter correction. Aerospace Science and Technology, 168, 110707.

[70] Li, F., Sheng, H., Xiang, Y., et al. (2026). Modular UAV autonomous aerial docking and control based on multi-sensor fusion. Chinese Journal of Aeronautics, 104073.

[71] Jin, X., Liu, H., Dong, G., et al. (2025). Contact dynamics modeling and control during the combination process of air-ground robots. Chinese Journal of Mechanical Engineering, 38, 39.

[72] Šćurić, A., Krznar, N., Penđer, A., et al. (2025). Autonomous multirotor UAV docking and charging: A comprehensive review of systems, mechanisms, and emerging technologies. Symmetry, 17(11), 1988.

[73] Shen, T. J., & Chen, C. L. (2025). Model-in-the-loop design and flight test validation of flight control laws for a small fixed-wing UAV. Drones, 9(9), 624.

[74] Tahir, M. A., Mir, I., & Islam, T. U. (2023). Control algorithms, Kalman estimation and near actual simulation for UAVs: State of art perspective. Drones, 7(6), 339.