2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

When birds meet fish: vision-force fusion for autonomous underwater docking in cross-domain avian-aquatic collaboration

Hongchang Liu[1,#], Ruiheng Wang[1,#], Yongkang Jiang[1,2,*], *Member, IEEE*, Shenli Zhang[2], Xiangdan Zhao[2], Xin Xu[2], Yulong Ding[1,2], Feng Lyu[2], Zhipeng Wang[1,2], Yanmin Zhou[1,2], Bin He[2], *Member, IEEE.*

***Abstract*—Unmanned aerial–aquatic vehicles (UAAVs) provide cross-domain adaptability and broad visions, while autonomous underwater vehicles (AUVs) support long-duration operations. This work integrates the two by developing a rapid underwater docking and releasing system. An autonomous clamping mechanism is designed to anchor UAAVs under varied landing attitudes, and a vision–tactile state perception algorithm based on decision-level dual-modal fusion is proposed to enable reliable underwater docking with no need of communications between the UAAV and AUV. Experimental results validate autonomous perception and reliable docking in fully underwater environments, achieving a docking time of 6 s and a landing gear recognition accuracy of 3 mm. The proposed framework offers an efficient solution for aerial–aquatic cooperation, advancing cross-domain robotic platforms for ocean monitoring, emergency response, and underwater exploration.**

I. Introduction

In nature, the distinct locomotion of birds and fish enables efficient wide-area aerial search and persistent aquatic navigation, respectively. This biological wisdom has inspired the development of Unmanned Aerial–Aquatic Vehicles (UAAVs) and Autonomous Underwater Vehicles (AUVs) [1], [2]. While UAAVs offer rapid deployment, they are constrained by limited payload and endurance; conversely, AUVs support long-duration deep exploration but lack aerial agility. Integrating these platforms through efficient docking systems combines their strengths, enabling seamless cross-domain collaboration. Most existing studies have focused on docking between aerial vehicles and fixed landing platforms [3], using methods such as vision-based autonomous landing with fiducial markers [4], magnetic recovery mechanisms [5], and adaptive mechanical structures [6]. While effective in above-water environments, these techniques rely on stable communication and air medium properties, as well as fixed platform support, making them unsuitable for complex underwater conditions.

Additionally, stable underwater communication for heterogeneous docking remains a bottleneck. Due to the high conductivity of seawater, Radio Frequency (RF) signals attenuate rapidly [7], making them impractical for long-range transmission. Although methods such as vision-based guidance [8–10], acoustic beacons [11], and electromagnetic structures [12] have been explored, they often suffer from significant limitations. Optical communication is range-limited by scattering [13], while acoustic methods [14] typically rely on bulky, energy-intensive devices unsuitable for lightweight aerial–aquatic platforms [15]. Moreover, most existing solutions assume homogeneous platforms or fixed bases, limiting adaptability in dynamic cross-domain scenarios.

To bypass communication constraints, autonomous underwater perception has emerged as a key solution. Traditional acoustic sensing [16] often suffers from low resolution and latency [17], while underwater vision [18] is constrained by turbidity and single-modal ambiguity. Consequently, multimodal fusion has become a research focus [19], such as integrating vision with acoustic [20] or sonar measurements [21]. Additionally, tactile sensing has been explored for its robustness to environmental variability [22]. However, integrating vision and tactile sensing for aerial–aquatic robots docking in complex aquatic environments remains unexplored.

Our work pioneers a zero-communication, self-sensing, and rapid autonomous underwater docking system that eliminates the need for additional payloads on the aerial vehicles. The proposed system employs a vision–tactile multimodal fusion algorithm to perceive and interpret aerial vehicle states under complex aquatic conditions, thereby overcoming the limitations of single-modal sensing and avoiding inefficient underwater communication. It further accommodates a wide range of approaching attitudes of UAAVs, ensuring robust and precise docking in real underwater scenarios. Beyond docking, the platform can be readily extended to applications such as energy replenishment, wired data exchange, payload transfer, and cooperative mission execution, providing a scalable and reliable foundation for integrated air–sea operations. The contributions of this paper are threefold:

- We propose a novel underwater docking-based cross-domain avian-aquatic collaboration workflow to integrate strengths of UAAVs and AUVs;
- Design of a dual-modal perception algorithm that fuses underwater vision and tactile sensing for robust docking of aerial-aquatic robots;
- Development of an underwater docking prototype and experimental validation of reliable collaboration of UAAVs and AUVs in challenging aquatic environments.

II. Design of the Underwater Docking System

A. System overview of the docking system

As illustrated in Fig. 1, the proposed collaboration workflow proceeds as follows: (1) After independent aerial and aquatic surveys, the UAAV locates the AUV, aligns, and dives; (2) Guided by visual markers, the UAAV lands on the

This research is supported by the National Key Research and Development Program of China under grant 2024YFB4709800, National Natural Science Foundation of China under grant 62303443, and Shanghai Municipal Science and Technology Major Project under grant 2021SHZDZX0100. *(Corresponding author: Yongkang Jiang)*

[1]College of Electronic and Information Engineering, Tongji University, Shanghai 201804, China.
[2]State Key Laboratory of Autonomous Intelligent Unmanned Systems, Tongji University, Shanghai 201210, China.

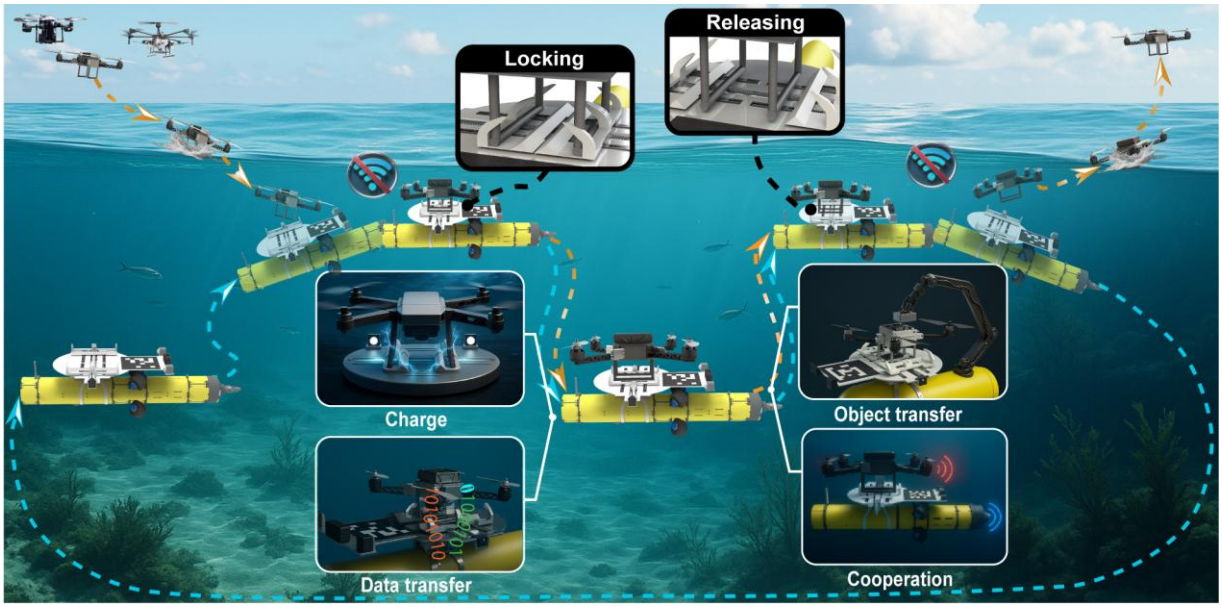

Figure 1. Workflow of the proposed zero-communication UAAV–AUV docking operations: the UAAV descends from the air into water, lands on the docking platform via vision marker-based guidance, recognized and clamped by the platform, and then performs tasks such as charging, object transfer, and data exchange. When departure is required, the UAAV applies downward thrust to trigger the release process, ascends from the water back into the air, and thus completes a closed-loop process.

docking platform; (3) The platform identifies the UAAV's landing intention and posture, triggering autonomous clamping; (4) The coupled system executes collaborative tasks, including battery charging, data transfer, and cooperative surveying; (5) Upon detecting the UAAV's departure intention, the platform releases it for independent operation.

To support this workflow, the docking system must achieve: (1) precise perception of UAAV states and docking/detaching intent; and (2) robust locking under diverse landing postures. As shown in Fig. 2(a), the system comprises a symmetrical clamping module and a vision–tactile sensing module. A Jetson Orin NX and an ESP32 manage high-level processing and low-level control, respectively, enabling event-driven integration.

The clamping module employs a bi-directional ball screw driven by a stepper motor. This mechanism ensures synchronized, mirrored displacement, preventing misalignment typical of dual-actuator designs while centering the UAAV and adapting to varying landing gear geometries.

The sensing module integrates a stereo camera and four piezoresistive pressure sensors, with a waterproof QR code on the platform facilitating initial guidance. Unlike sonar, which suffers from blind zones and low resolution at short ranges, calibrated stereo vision provides precise pose estimation and remains effective even in turbid environments. Complementing vision, tactile sensors capture real-time force distribution to verify contact states and operational intent.

B. Clamping mechanism design

A major challenge in reliable docking is ensuring that the clampers can tolerate position and posture deviations of the incoming vehicle. Initial tests showed that flat-front clampers, lacking guiding features, could not realign the landing gear before clamping. As a result, offsets during approach (as in Fig. 2(b1)) often caused misalignment, jamming, and failed sensing or releasing operations.

To overcome this, we introduced a shovel-shaped chamfer at the front of the clampers as in Fig. 2(a). The inclined surface provides a guide-before-clamp action, steering the landing gear toward the symmetry plane before engagement. This passive self-alignment arises from contact forces: the normal force decomposes into axial and radial components, with the radial component inducing a corrective torque that aligns the gear to the clamping axis. This geometry significantly increases tolerance to approach errors, improving docking robustness and success rate.

To theoretically analyze the range of the UAAV landing angles θ that the mechanism can successfully clamp with the shovel-based guiding design (Fig. 3), we conduct a

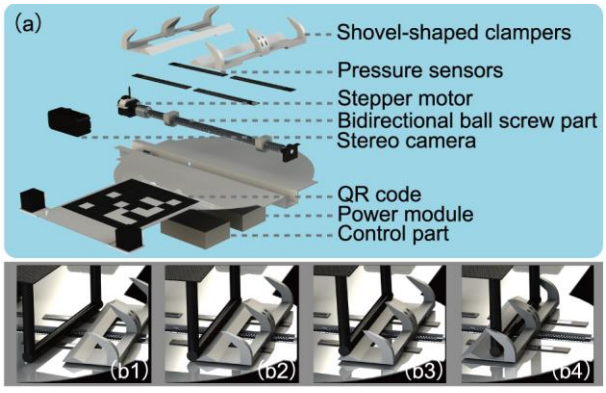

Figure 2. Components of the proposed underwater docking system and a brief workflow during locking of the UAAV landing gears. (a) The system comprises of a shovel-structured clamper, pressure and visual sensors, and actuating ball screw part, etc. (b1-b4) Clamping processes of the docking system to correct the landing angles of the UAAV and firmly lock it for further underwater cooperation.

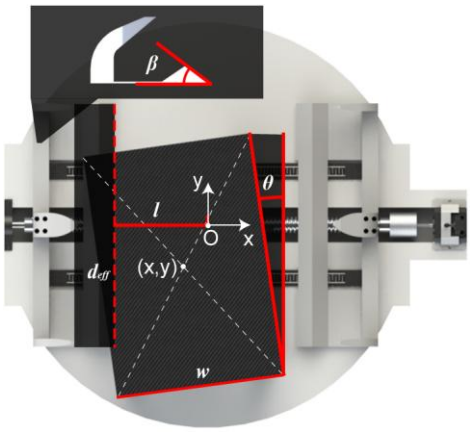

Figure 3. Parameterized description of the UAAV's landing on the docking system with an angle θ and an offset (x, y).

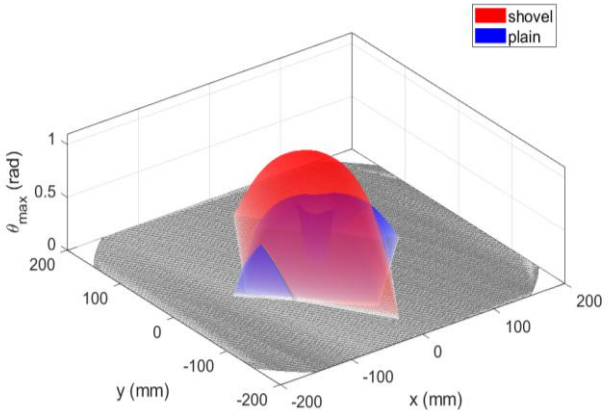

Figure 4. Maximum clampable angle maps of the docking system when designed with (red) and without (blue) the shovel-shaped structure on the front of the clampers.

mathematical model as derived as follows. According to the geometrical analysis, we have

$$\delta(x, y, \theta) = (l - |x|) \cdot tan\beta + w \cdot |sin\theta| - |y| \quad (1)$$

where $\delta(x, y, \theta)$ denotes the equivalent lateral occupation, combining lateral offset, chamfer-induced longitudinal error, and tilt-induced tip deflection into a single length scale that can be directly compared with d_{eff}. d_{eff} refers to the width of the clamper, (x, y) denote the center of the UAAV landing gears, β is the shovel-chamfer angle, w denotes the width of the UAAV landing gears, and l is the axial distance from the front tip of the clampers to the center of the clamping region (Fig. 3).

The maximum self-alignment angle provided by the shovel geometry is expressed as

$$\theta_{max} = \max\left\{0, arctan\left(\frac{d_{eff} - |y| + (l - |x|) \cdot tan\beta}{l}\right)\right\} \quad (2)$$

The combined feasibility condition is then

$$\Phi_{shovel} = \max\{\delta(x, y, \theta) - d_{eff}, |\theta| - \theta_{max}\} \leq 0 \quad (3)$$

The clamping is feasible only if Eq. (3) is satisfied.

To show the advantages of the proposed shovel-shaped design, we set β=0 in Eqs. (1) – (3), which indicate that the shovel-shaped structure disappears. In this case, the clamping criterion is obtained:

$$\delta_0 = w \cdot |sin\theta| - |y|$$
$$\theta_{max}(x, y) = \max\left\{0, arctan\left(\frac{d_{eff} - |y|}{l}\right)\right\} \quad (4)$$

Although Eq. (4) is algebraically contained in Eq. (3), their physical interpretations differ significantly. Simulation and experimental results validate the benefits of the shovel-shaped design on the clampers. In Fig. 4, the red and blue regions represent the maximum permissible landing angle θ_{max} with and without this design, respectively, under varying in-plane offsets (x, y). The shovel-shaped design consistently allows larger landing angle tolerance across most offset ranges, enabling the UAAV to land successfully with greater initial attitude error. The conventional design outperforms only when the offset x exceeds the axial distance l—an extreme case rarely encountered in practice and explained by Eq. (2), where $(l - |x|)$ becomes negative. The system achieves a landing radius of 150 mm and a maximum tolerable angle of 38°.

Our experimental results also confirm that the shovel-shaped design on the clampers not only improves misalignment tolerance while landing, but also smooths feasibility boundaries, effectively converting marginal docking attempts into robust successes.

C. Visual perception: underwater 3D geometric reconstruction via dual-domain mask fusion

The core objective of the vision system is to achieve accurate 3D geometric reconstruction of UAAV landing gears when landing on the underwater docking platform. The pipeline consists of four modules, as illustrated below.

Due to refraction-induced distortions and light attenuation in aquatic environments, we first calibrate the stereo camera system. The intrinsic matrices of the left and right cameras are denoted by $K_l, K_r \in R^{3\times3}$, while the relative pose between them is described by rotation $R \in SO(3)$ and translation $t \in R^3$. These parameters enable rectification and refraction compensation, ensuring that the corrected images satisfy the pinhole model and epipolar geometry.

To obtain a high-quality disparity map in scattering-prone underwater conditions, we adopt the Semi-Global Block Matching with 3-way aggregation (SGBM-3way) algorithm [23]. Its optimization objective is

$$E(D) = \sum_{q\in N(p)} P_1 \cdot [|D_p - D_q| = 1] + \sum_{q\in N(p)} P_2 \cdot [|D_p - D_q| > 1] + \sum_p C(p, D_p) \quad (5)$$

where $C(p, D_p)$ denotes the matching cost of pixel p at disparity D_p, and P_1, P_2 are smoothness penalties. The 3-way aggregation scheme reduces streaking artifacts and suppresses noise in low-texture or scattering regions. The recovered depth at pixel p is then computed as

$$Z(p) = \frac{f \cdot B}{d(p)} \quad (6)$$

with f the focal length, B the stereo baseline, and $d(p)$ the disparity value.

Accurately segmenting UAAV landing gears in cluttered underwater scenes is challenging, as general object recognition methods often suffer from visual ambiguity. To overcome this, we introduce Dual-Domain Mask Fusion (DDMF), which combines 2D image-based masks capturing appearance cues with 3D depth-based masks emphasizing geometric structure as in Fig. 5. We employ two lightweight YOLO-based networks [24]: one processes the rectified left image I_L to generate 2D instance segmentation masks $\mathcal{M}_{2\mathcal{D}} = \{m^i_{2D}\}$ with confidences $\mathcal{C}_{2\mathcal{D}} = \{c^i_{2D}\}$, while the other incorporates both I_L and the depth map D to predict 3D-aware masks $\mathcal{M}_{3\mathcal{D}} = \{m^i_{3D}\}$ with confidences $\mathcal{C}_{3\mathcal{D}} = \{c^i_{3D}\}$.

Masks are considered valid only if their confidence exceeds a threshold τ. For a candidate pair (M^i_{2D}, M^j_{3D}),the fusion function f_{fuse} operates as

$$f_{fuse}(M_i^{2D}, M_j^{3D}) = \begin{cases} M_i^{2D} \cap M_j^{3D}, \text{if } M_i^{2D} \cap M_j^{3D} \neq \emptyset \\ M_i^{2D} \cup M_j^{3D}, \text{if } M_i^{2D} \cap M_j^{3D} = \emptyset \end{cases} \quad (7)$$

Intersection strategy: If masks overlap spatially, we take the intersection, ensuring high precision by preserving only consistent detections across both domains.

Union strategy: If masks do not overlap, we take the union to retain complementary information from each domain, improving recall and ensuring structural completeness of the landing gear.

The fused mask M_{fused} is applied to the dense depth map D, isolating the landing gear region. Each pixel (u,v) within the mask is back-projected into 3D space

$$P = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = Z(u,v)K^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad (8)$$

where $Z(u,v)$ is the depth at (u,v) and K is the intrinsic matrix. This yields a raw point cloud containing only the UAAV landing gear.

Subsequently, we perform statistical denoising and outlier removal, followed by model fitting guided by the geometric priors of the UAAV landing gear—such as symmetry and parallelism among the landing gears. This process effectively prevents the whole system from disturbance by unrelated structure such as fishing net or marine debris. Furthermore, it enables precise estimation of geometric parameters and landing poses, which are critical for downstream control and docking tasks.

D. Tactile perception: interaction state recognition via dynamic sliding window and adaptive baseline

The pressure signal acquired in real time is represented as a discrete-time sequence: $p(t), \quad t \in N$.

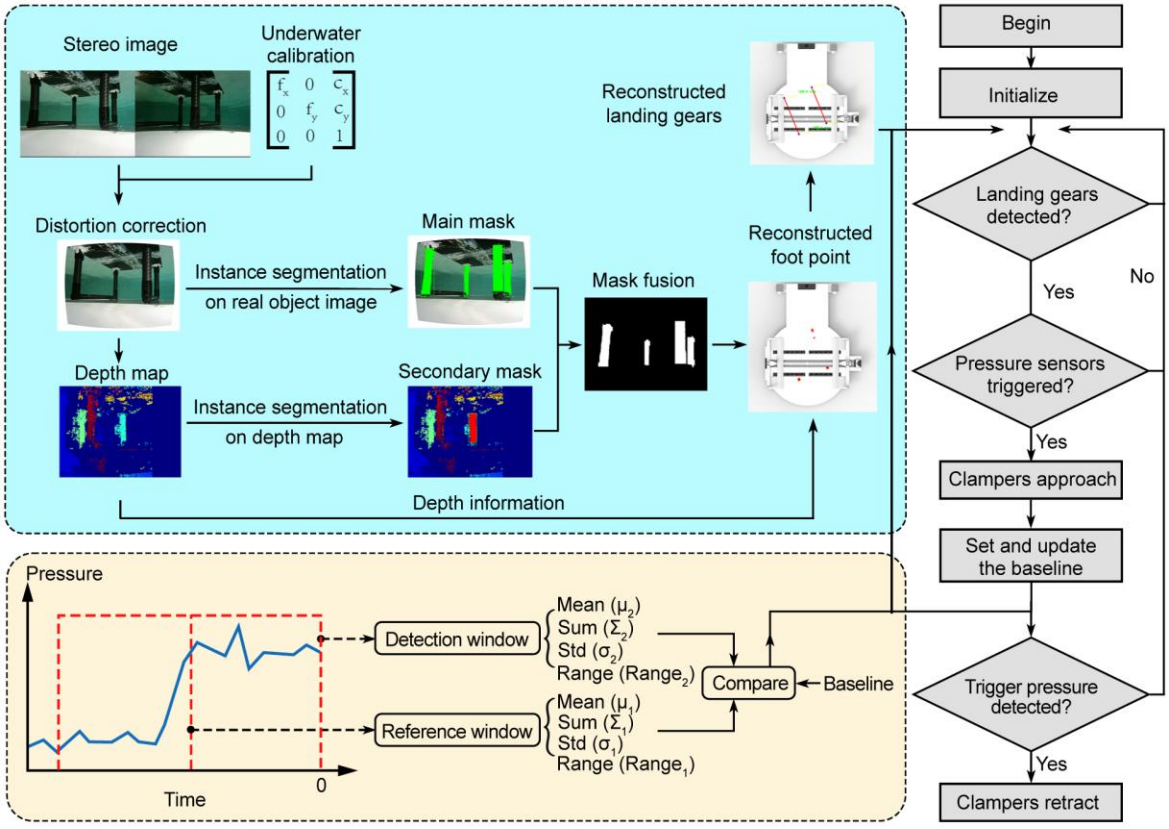

Figure 5. Overall workflow of the vision-tactile fusion module. The top-left subfigure depicts the visual processing pipeline, including stereo image rectification, depth mapping, instance segmentation, and 3D footprint reconstruction. The bottom-left subfigure illustrates pressure analysis using a sliding window to identify sustained high pressure after a sudden increase. The right panel presents the integrated decision process: clamping is triggered only when both visual and force parts confirm landing intentions of the UAAV, and release occurs when pressure remains consistently high for a certain period.

At any time step t, the system maintains two sliding windows of length W_r and W_d, respectively

$$\mathcal{R}(t) = \{p(t - W_r - W_d + 1), \dots, p(t - W_d)\} \quad (9)$$

which is used to estimate the dynamic baseline under the current hydrodynamic disturbances, and

$$\mathcal{D}(t) = \{p(t - W_d + 1), \dots, p(t)\} \quad (10)$$

which is used to capture instantaneous variations induced by interaction.

Within each window, statistical features such as the mean and standard deviation are computed. The interaction state is then inferred from the difference between the mean values of the two windows.

After the clamper achieves the initial locking of the UAAV, a stable segment of pressure data is collected. Its average value is calculated and stored as the baseline load

$$B = \frac{1}{T}\sum_{i=0}^{T-1} p(t_0 + i) \quad (11)$$

where $p(i)$ denotes the sampled pressure (aggregated from four sensors, and considered valid only when all four sensors are activated nearly simultaneously, which holds for regular landing gears such as UAAV skids), and t_0 is the time when the UAAV lands and becomes stable. This baseline reflects the average load of the UAAV in a static locked state, thereby enabling automatic adaptation to different UAAV platforms and payloads.

To avoid reliance on a fixed absolute threshold, a relative variation criterion is adopted. Specifically, the ratio

$$\eta(t) = \frac{|\Delta\,\mu(t)|}{B} \quad (12)$$

is monitored, where $|\Delta\,\mu(t)|$ denote the data difference of the detection and reference windows, respectively. An interaction is declared when $\eta(t)$ exceeds a predefined relative threshold θ:

$$State(t) = \begin{cases} Stable, & \eta(t) < \theta \\ Interaction, & \eta(t) \geq \theta \end{cases} \quad (13)$$

E. Dual-modal fusion strategy

To realize robust and fast-response locking and releasing of the UAAVs, we integrate the visual and tactile parts to construct a tightly coupled, event-driven perception–control unified system. The integrated decision process of the dual-modal fusion strategy is described in the right panel of Fig. 5. Once the visual part detects that the UAAV landing gear entering the designated touchdown zone, the system extracts the geometric parameters of the landing gear and maps them to the target closure displacement threshold of the clampers, thereby enabling full-scale adaptive clamping. Subsequently, the tactile part (pressure sensors) performs real-time monitoring. Upon detecting a contact signal, the clamper gradually closes to the optimal degree of constriction calculated from the visual recognition results, preventing both excessive compression and insufficient holding, thus ensuring mechanical stability and structural safety.

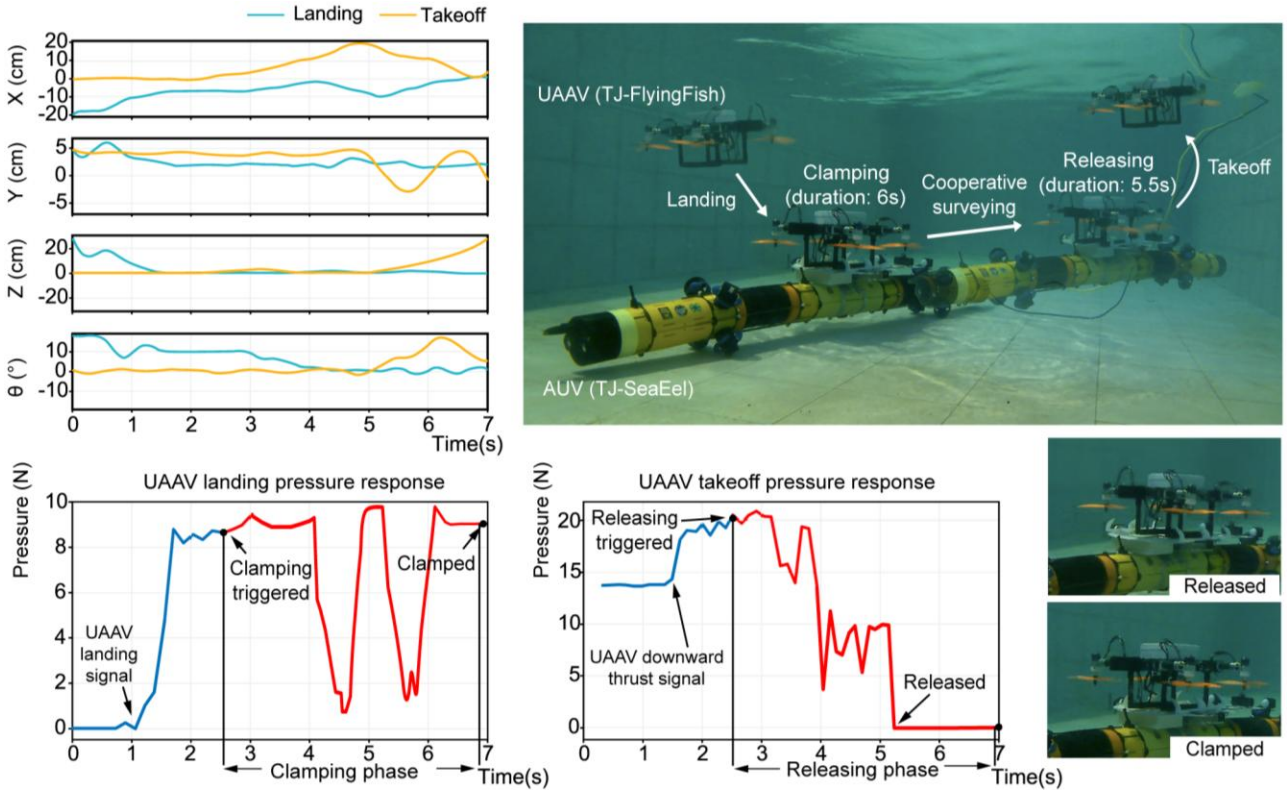

Figure 6. Demonstration and force response evaluation of the proposed underwater UAAV–AUV docking system. The top-right subfigure illustrates the complete docking process: the UAAV landed on the docking station, was secured by the clamping mechanism, conducted a collaborative survey with the AUV, was released, and finally took off. The top-left subfigure shows the positional changes of the UAAV's landing gear center in x, y, and z coordinates, as well as the deviation angle θ. The bottom subfigures present the pressure responses recorded by the docking system during the landing and takeoff phases of the UAAV, respectively.

After the clamping mechanism tightened, the camera and pressure monitor will continuously collect signal and analysis: if the pressure sensors detect static loading without the characteristic dynamic fluctuations of a functional UAAV or the camera couldn't detect the landing gear for a certain period, the mechanism automatically executes an emergency release to free the clamping area, which prevent the whole system from failure.

Distinct from traditional approaches that rely solely on contact detection, this work innovatively extends the tactile part into an interaction-intent recognition channel. When the UAAV intends to takeoff, it conveys the intention by exerting downward pressures on the docking platform. Once the pressure signal exhibits a significant and sustained upward trend, the system recognizes the UAAV's intention to disengage and activates the clamper's rapid-release mechanism, achieving a smooth and safe separation process.

III. Underwater Docking System Demonstrations

We conducted experiments as illustrated in Fig. 6. The UAAV utilized in the experiments is our proprietary aerial–aquatic vehicle "TJ-FlyingFish", and the AUV is another underwater robot "TJ-SeaEel". The top-right subfigure in Fig. 6 depicts the complete underwater docking process. As described, the UAAV successfully located the docking platform mounted on the AUV. The docking system then clamped the UAAV in 6 seconds to enable cooperative surveying and later released it in 5.5 seconds to allow for separate operations.

During clamping and release, the trajectories of the UAAV landing gear center show clear convergence (Fig. 6, top-left). The x-coordinate fluctuates markedly after initial contact, then stabilizes near −10 cm between 1.6 and 3.1 s, reflecting interaction with the shovel-shaped guide that redirects lateral deviation toward the symmetry axis. The y-position remains near 5 cm with minimal oscillation, indicating effective longitudinal constraint by the guide. The z-position decreases steadily from around 20 cm, stabilizing near the platform height by 1.5 s, demonstrating fast settling under guidance and gravity. The rotation angle θ oscillates initially (0–4 s), then converges to near 0°, as the guide structure produces a restoring torque that corrects attitude error, achieving a "guide before clamp" docking sequence.

During release, the landing gear center exhibits motion opposite to clamping. The x-displacement increases after 2 s, peaking around 5 s as the UAAV overcomes clamping force via upward thrust. A transient y-deviation occurs near 5 s, resulting from asymmetric thrust and hydrodynamic effects during takeoff. The z-position rises sharply from 5 to 7 s until complete detachment. The rotation angle θ briefly fluctuates before stabilizing, indicating a quick attitude correction after release. These results demonstrate that the system reliably detects takeoff intent and enables smooth disengagement under full autonomy.

The bottom subfigures in Fig. 6 show the pressure response of the docking system during the landing and takeoff phases, respectively. During landing, a sharp pressure increase occurs at 1.1 s upon contact between the UAAV and the platform, stabilizing around 9 N, which confirms reliable contact detection and adaptive locking by the clamping mechanism. Throughout the clamping process, the pressure exhibits fluctuations due to hydrodynamic effects. During takeoff, the UAAV applies a downward thrust, initiating the release phase as the force sensors detect a sustained increase in pressure starting at 1.6 s. A rapid pressure drop to the baseline level at 5.2 s marks the completion of the release and the successful takeoff of the UAAV.

These results mentioned above validate that the vision–tactile fusion framework ensures reliable contact detection and intent recognition without communication, supporting autonomous docking and disengagement.

IV. Conclusion

This paper presents an underwater rapid docking platform for UAAVs and AUVs, employing a symmetrically arranged clamper driven by a bidirectional lead screw and integrated with a shovel-shaped passive correction mechanism. Designed specifically for communication-denied underwater environments, the platform operates without relying on wireless links. Instead, it achieves accurate attitude estimation and clamping control through local vision–pressure dual-modal fusion. Experimental results show that the system ensures reliable docking under complex underwater conditions without communication.

Building upon this highly robust docking capability, future work will focus on expanding post-docking functionalities, including in-situ battery charging, high-speed wired data transmission, and mission parameter updates via physical interfaces. Further research will also explore hydrodynamic optimization and the development of foldable or reconfigurable clampers to reduce underwater drag. These enhancements will support the creation of an integrated underwater service station, facilitating long-endurance operations and multi-vehicle coordination for cross-medium UAAVs, thereby promoting applications in ocean observation, environmental monitoring, and air–water collaborative missions.

References

[1] X. Liu, M. Dou, D. Huang, B. Wang, J. Cui, Q. Ren, L. Dou, Z. Gao, J. Chen, and B. M. Chen, "TJ-FlyingFish: Design and implementation of an aerial-aquatic quadrotor with tiltable propulsion units," in 2023 IEEE International Conference on Robotics and Automation (ICRA), 2023.

[2] L. Li, S. Wang, Y. Zhang, S. Song, C. Wang, S. Tan, W. Zhao, G. Wang, W. Sun, F. Yang, J. Liu, B. Chen, H. Xu, P. H. Nguyen, M. Kovac, and L. Wen, "Aerial-aquatic robots capable of crossing the air-water boundary and hitchhiking on surfaces," Science Robotics, vol. 7, 2022.

[3] C. Grlj, N. Krznar, and M. Pranjić, "A Decade of UAV Docking Stations: A Brief Overview of Mobile and Fixed Landing Platforms," Drones, vol. 6, no. 1, p. 17, 2022.

[4] D. Falanga, A. Zanchettin, A. Simovic, J. Delmerico, and D. Scaramuzza, "Vision-based autonomous quadrotor landing on a moving platform," in IEEE International Symposium on Safety, Security, and Rescue Robotics, 2017.

[5] X. Yu et al., "Design and optimize an aerial precision docking system for UAVs based on magnetic vector fields," in 43rd Chinese Control Conference, Kunming, China, 2024, pp. 4693–4698.

[6] H. Nieuwoudt, J. Welgemoed, T. v. Niekerk, and R. Phillips, "Automated charging and docking station for security UAVs," in 14th International Conference on Mechanical and Intelligent Manufacturing Technologies (ICMIMT), Cape Town, South Africa, 2023, pp. 32–38, doi: 10.1109/ICMIMT59138.2023.10200192.

[7] X. Che, I. Wells, G. Dickers, P. Kear, and X. Gong, "Re-evaluation of RF electromagnetic communication in underwater sensor networks,"

IEEE Communications Magazine, vol. 48, no. 12, pp. 143–151, Dec. 2010.
[8] M. Myint, K. Yonemori, A. Yanou, M. Minami, and S. Ishiyama, "Visual-servo-based autonomous docking system for underwater vehicle using dual-eyes camera 3D-pose tracking," in IEEE/SICE International Symposium on System Integration (SII), Nagoya, Japan, 2015, pp. 989–994, doi: 10.1109/SII.2015.7405161.
[9] P.-M. Lee, B.-H. Jeon, and S.-M. Kim, "Visual servoing for underwater docking of an autonomous underwater vehicle with one camera," in Oceans 2003. Celebrating the Past ... Teaming Toward the Future (IEEE Cat. No.03CH37492), San Diego, CA, USA, 2003, vol. 2, pp. 677–682, doi: 10.1109/OCEANS.2003.178391.
[10] T. Ni, C. Sima, W. Zhang, J. Wang, J. Guo, and L. Zhang, "Vision-based underwater docking guidance and positioning: Enhancing detection with YOLO-D," Journal of Marine Science and Engineering, vol. 13, no. 1, p. 102, 2025, doi: 10.3390/jmse13010102.
[11] J. Ureña et al., "Acoustic local positioning with encoded emission beacons," Proceedings of the IEEE, vol. 106, no. 6, pp. 1042–1062, Jun. 2018.
[12] R. Lin, Y. Zhao, D. Li, M. Lin, and C. Yang, "Underwater electromagnetic guidance based on the magnetic dipole model applied in AUV terminal docking," Journal of Marine Science and Engineering, vol. 10, no. 7, p. 995, 2022.
[13] B. R. Angara, P. Shanmugam and H. Ramachandran, "Underwater Wireless Optical Communication System Channel Modelling With Oceanic Bubbles and Water Constituents Under Different Wind Conditions," in IEEE Photonics Journal, vol. 15, no. 2, pp. 1-11, April 2023, Art no. 7301611, doi: 10.1109/JPHOT.2023.3258500.
[14] Q. Wang, B. He, Y. Zhang, F. Yu, X. Huang, and R. Yang, "An autonomous cooperative system of multi-AUV for underwater targets detection and localization," Engineering Applications of Artificial Intelligence, vol. 121, 2023.
[15] M. Chitre, S. Shahabudeen, and M. Stojanovic, "Underwater acoustic communications and networking: Recent advances and future challenges," Marine Technology Society Journal, vol. 42, no. 1, pp. 103–116, 2008.
[16] L. Freitag, M. Grund, S. Singh, J. Partan, P. Koski, and K. Ball, "The WHOI micro-modem: An acoustic communications and navigation system for multiple platforms," in Proceedings of OCEANS, 2005, pp. 1086–1092.
[17] M. Stojanovic and J. Preisig, "Underwater acoustic communication channels: Propagation models and statistical characterization," IEEE Communications Magazine, vol. 47, no. 1, pp. 84–89, Jan. 2009.
[18] C. Li and S. Anwar, "Underwater scene prior inspired deep underwater image and video enhancement," Pattern Recognition, vol. 98, 2019.
[19] M. Yang, Z. Sha, and F. Zhang, "A multimodal approach based on large vision model for close-range underwater target localization," IEEE/ASME Transactions on Mechatronics, vol. 30, no. 4, pp. 2427–2437, Aug. 2025, doi: 10.1109/TMECH.2024.3449090.
[20] S. Liu, X. Fan, G. Wu, L. Yao, and S. Geng, "More modalities mean better: Vessel target recognition and localization through symbiotic transformer and multiview regression," IEEE Transactions on Geoscience and Remote Sensing, vol. 62, Art. no. 4203512, pp. 1–12, 2024, doi: 10.1109/TGRS.2024.3365711.
[21] M. E. Deowan, M. S. Y. Yousha, T. M. Hossain, S. Hassan, and R. Marxer, "Optimizing underwater robot navigation: A study of DRL algorithms and multi-modal sensor fusion," in 2025 IEEE International Conference on Robotics and Automation (ICRA), Atlanta, GA, USA, 2025, pp. 11270–11277, doi: 10.1109/ICRA55743.2025.11127836.
[22] J. Sun, Q. Zhang, Y. Lu, B. Huang, and Q. Li, "A review of touching-based underwater robotic perception and manipulation," Machines, vol. 13, no. 1, p. 41, Jan. 2025.
[23] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 2, pp. 328–341, Feb. 2008, doi: 10.1109/TPAMI.2007.1166
[24] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.