Towards Autonomous Reef Surveying: An Online Underwater Localization and Photorealistic Mapping System based on Multi-sensor Fusion

1st Hanwen Zhang 2nd Yipeng Zhu 3rd Xiaopeng Guo
Integrative Systems and Design (ISD)
The Hong Kong University of Science and Technology (HKUST)
HongKong, China
{hzhangfr, yzhudg, xguoay}@connect.ust.hk

4th Huajian Huang
School of Interdisciplinary Science
Beijing Institute of Technology
Beijing, China
huajian@bit.edu.cn †

5th Sai-Kit Yeung
ISD
HKUST
HongKong, China
saikit@ust.hk

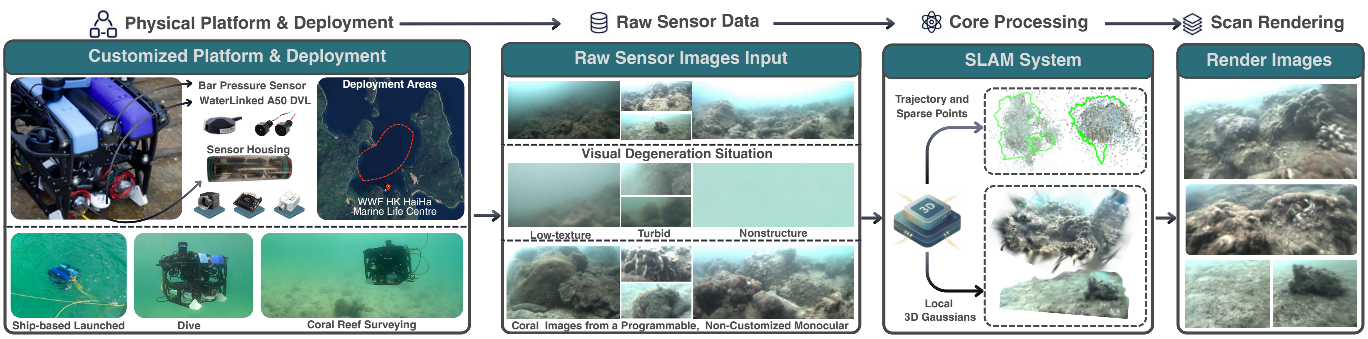

Fig. 1. The proposed real-time multi-sensor SLAM system for robust underwater localization and mapping. Our framework addresses visual degradation by integrating inertial, acoustic, and pressure modalities with monocular vision. The system achieves consistent state estimation and leverages 3D Gaussians to generate photorealistic 3D renderings, facilitating precise coral reef monitoring and representation.

***Abstract*—Resilient localization and photorealistic mapping constitute the fundamental components of intelligent reef robotics. However, unstructured geometry and degraded visibility severely undermine both localization robustness and mapping fidelity. To address these, we propose a real-time multi-sensor SLAM system that integrates inertial, acoustic, pressure and visual modalities, ensuring consistent state estimation and high-fidelity photorealistic mapping in degraded underwater environments. Furthermore, to mitigate the scarcity of benchmarks derived from real-world robotic platforms, this work contributes a coral reef surveying dataset featuring multi-modal data collected via BlueROV deployments. Extensive experiments conducted on the public dataset and the proposed benchmark demonstrate that our framework delivers superior localization precision and photorealistic representation while maintaining real-time online performance.**

***Index Terms*—Multi-Modal Sensing and Fusion, Underwater Scene Representations**

I. Introduction

Autonomous underwater robots facilitate high-resolution, non-invasive coral reef surveying in complex environments [1]–[3]. Robust state estimation is central to these operations, providing the essential foundation for safe navigation and spatial data integration. Furthermore, photorealistic 3D reconstruction is increasingly vital for ecological assessment, as it offers immersive digital twins that overcome the limitations of traditional geometric mapping [4].

However, purely visual underwater SLAM [5]–[7] often fails in marine environments due to turbidity and light attenuation, necessitating multi-sensor fusion to maintain consistent robustness [8]–[10]. Despite this, existing underwater multi-sensor SLAM systems often struggle with limited estimation accuracy and rely heavily on rigid stereo configurations that lack flexibility across diverse platforms. Furthermore, these systems typically fail to provide real-time photorealistic mapping, leaving a gap between robust state estimation and high-fidelity environmental representation [11]–[13].

Developing such systems requires datasets that mirror real-world conditions. However, current benchmarks remain a significant bottleneck. Most existing datasets are restricted to laboratory tanks or lack synchronized multi-modal sequences obtained from robotic platforms in coral reef environments [14]–[16]. This scarcity hinders the rigorous evaluation of SLAM systems in these ecologically significant environments.

To bridge these gaps, a real-time multi-sensor SLAM framework is proposed that integrates monocular vision, inertial, acoustic Doppler Velocity Log (DVL), and pressure modalities for robust state estimation and high-fidelity mapping, as shown in Figure 1. Unlike existing methods, our system incorporates an incremental 3D Gaussians [17] mapping engine, enabling real-time, photorealistic reconstruction through explicit water-media modeling and efficient densification of primitives. Furthermore, we contribute a multi-modal dataset collected via

TABLE I
QUANTITATIVE COMPARISON OF ATE (M) ON THE TANK DATASET.

Method	Sensors	Structure	HalfTank	WholeTank	Average
ORB3	S+I	2.17	1.35	1.53	1.70
VINS	S+I	0.22	29.80	13.10	–
UVA	S+D	0.37	0.58	0.29	0.43
AQUA	S+I+D	0.25	0.31	0.37	0.30
Ours	**M+I+D+P**	**0.17**	**0.26**	**0.25**	**0.22**

Bold indicates the best performance. M: Mono, S: Stereo, I: IMU, D: DVL, P: Pressure.

BlueROV deployments in coral reef ecosystems, addressing the scarcity of field-derived benchmarks. Extensive experiments on both public datasets and our proposed benchmark demonstrate that delivers superior localization precision and mapping fidelity with real-time online performance.

II. METHOD

In contrast to terrestrial SLAM, underwater visual degradation involves frequent short-term and sustained long-term failures, which can destabilize factor-graph optimization. We address this by constructing a non-degradable subsystem fusing inertial, acoustic, and pressure data based on ESKF [18]. To fuse visual input, the factor graph leverages this subsystem's stable estimation to maintain state continuity, enabling rapid re-initialization upon visual recovery. The final state output is fused by coupling the two subsystems based on their respective noise weights to ensure optimal precision and reliability.

Under non-degraded visual conditions, the mapping system integrates the estimated state and leverages high-confidence sparse feature points optimized by the factor graph to initialize the 3D Gaussians. We incorporate a physics-based water-medium model to compensate for light attenuation and scattering effects online. To address the geometric sparsity inherent in monocular configurations and weak texture, an efficient densification scheme is proposed to adaptively populate 3D Gaussian primitives. Our approach enables high-fidelity and structurally consistent scene representation even in complex underwater conditions.

To bridge the gap of benchmark for coral mapping evaluation, we present a coral reef surveying dataset collected in the unstructured coral reef environments of Hong Kong waters using BlueROV. The sensor comprises a fixed forward-looking industrial monocular camera, a 6-axis IMU, a Bar50 pressure sensor, and a Waterlinked A50 DVL. A key feature of this dataset is its inclusion of diverse environmental degradations, ranging from low-texture surfaces and structureless voids to high-turbidity conditions.

III. EXPERIMENT

We evaluate the proposed system on the public Tank [16] dataset and our self-collected Coral Reef Surveying dataset. Tracking performance is benchmarked against state-of-the-art classical SLAM (ORB-SLAM3 [19] and VINS-Fusion [20]) and underwater multi-sensor SLAM (UVA [21], and AQUA-SLAM [13]) based on the Absolute Trajectory Error [22] (ATE RMSE) on the Tank dataset. For mapping evaluation, we conducted qualitative and quantitative evaluations based on PSNR metrics on our own dataset compared with Go-SLAM [23], MonoGS [24], Photo-SLAM [25] and VINGS [26].

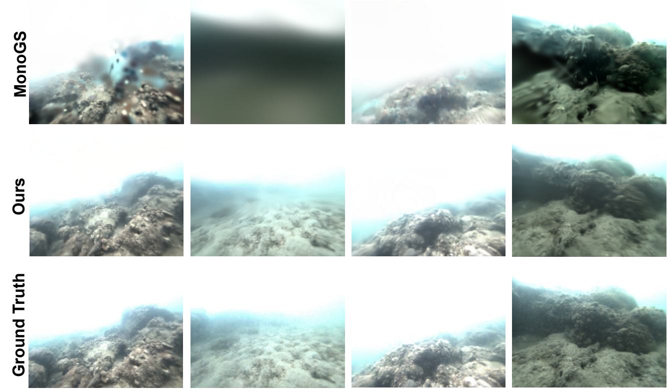

Fig. 2. Qualitative results on the Coral Reef Surveying dataset. Our method achieves best performance with high-fidelity texture and sharp details.

The proposed method achieves the superior tracking performance on the Tank dataset, as detailed in Table I. Classical stereo-inertial SLAM systems fail to maintain stable trajectories on sequences characterized by significant environmental degradation, leading to substantial drift or tracking failure. Furthermore, the proposed framework consistently outperforms existing underwater solutions, such as UVA and AQUA (0.22m vs. 0.30m), despite the reliance of those methods on stereo configurations. These experiments validate the robust and accurate localization capabilities of the integrated multi-sensor approach in challenging underwater scenarios.

In real-world data by robot deployment, Go-SLAM and Photo-SLAM with a pure monocular setup fail tracking and lead to fragmented reconstructions due to the substantial sensor noise and severe visual degradation stemming from unstructured coral geometries and limited underwater visibility. Although VINGS leverages inertial cues, its unspecialized deep optical flow front-end also fails to mapping. As shown in Figure 2, while MonoGS produces complete maps, its limited localization accuracy results in blurred textures (24.14 PSNR). In contrast, our method demonstrates superior robustness, delivering photorealistic reconstructions (29.05 PSNR) with fine-grained and spatially consistent details.

IV. CONCLUSION

We introduce a real-time multi-modal SLAM system and a new coral reef dataset to address the challenges of underwater localization and mapping. By integrating IMU, acoustic, pressure, and visual sensors, our framework ensures robust state estimation and high-fidelity reconstruction despite degraded visibility. Extensive evaluations demonstrate that our system outperforms existing methods in both precision and mapping quality, making it a reliable tool for intelligent reef robotics.

References

[1] Z. Li, Y. Zhang, S. He, X. Zhu, T. Wei, and C. Hu, "Centimeter-scale submarine robot for monitoring coral reef ecosystem," in *2024 IEEE International Conference on Mechatronics and Automation (ICMA)*. IEEE, 2024, pp. 345–350.

[2] A. Gupta, A. Abdullah, X. Li, V. Ramesh, I. Rekleitis, and M. J. Islam, "Demonstrating cavepi: Autonomous exploration of underwater caves by semantic guidance," in *Robotics: Science and Systems (RSS)*, 2025.

[3] S. McCammon, L. Cai, D. Yang, J. Walsh, J. D. Cast, T. A. Mooney, and Y. Girdhar, "Autonomous seeking and mapping coral reef biodiversity hotspots with a multimodal auv," *Science Robotics*, vol. 11, no. 114, p. eadx9939, 2026.

[4] D. Yang, J. Hong, J. J. Leonard, and Y. Girdhar, "Reefmapgs: Enabling large-scale underwater reconstruction by closing the loop between multimodal slam and gaussian splatting," *arXiv preprint arXiv:2604.11992*, 2026.

[5] J. Zhang, F. Han, D. Han, J. Yang, W. Zhao, and H. Li, "Integration of sonar and visual–inertial systems for slam in underwater environments," *IEEE Sensors Journal*, vol. 24, no. 10, pp. 16 792–16 804, 2024.

[6] R. Liu, S. Fan, W. Wang, and Y. Yang, "Underwater visual slam with depth uncertainty and medium modeling," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025, pp. 970–980.

[7] K. Wang, S. Zou, C. Jiang, Y. Dai, S. Chen, S. Shen, and G. Wang, "Watersplat-slam: Photorealistic monocular slam in underwater environment," *IEEE Robotics and Automation Letters*, 2026.

[8] H. M. Ali, H. J. P. Williams, L. Arulmozhiselvan, F. T. Ayasrah, A. Smerat, S. Venkatesh, H. R. Suresh, and S. Sengan, "Advancing underwater robotic localization accuracy through multi-sensor data fusion and evolutionary optimization techniques," 2025.

[9] M. Guo, B. Wang, L. Wei, M. Zhang, C. Zhang, and H. Lu, "Robust ins/gnss/dvl integrated navigation for mass based on gradient-adaptive factor graph optimization," *Electronics*, vol. 15, no. 3, p. 634, 2026.

[10] X. Chen, Y. Feng, B. Nie, W. Lu, and X. Wang, "Sonid: underwater sonar slam integrated with neural inertial-dvl odometry," in *Fifth International Conference on Artificial Intelligence, Automation, and Algorithms (AI2A 2025)*, vol. 14073. SPIE, 2026, pp. 28–36.

[11] J. Song, O. Bagoren, R. Andigani, A. Sethuraman, and K. A. Skinner, "Turtlmap: Real-time localization and dense mapping of low-texture underwater environments with a low-cost unmanned underwater vehicle," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024, pp. 1191–1198.

[12] S. Rahman, A. Quattrini Li, and I. Rekleitis, "Svin2: A multi-sensor fusion-based underwater slam system," *The International Journal of Robotics Research*, vol. 41, no. 11-12, pp. 1022–1042, 2022.

[13] S. Xu, K. Zhang, and S. Wang, "Aqua-slam: Tightly-coupled underwater acoustic-visual-inertial slam with sensor calibration," *IEEE Transactions on Robotics*, 2025.

[14] M. Bernardi, B. Hosking, C. Petrioli, B. J. Bett, D. Jones, V. A. Huvenne, R. Marlow, M. Furlong, S. McPhail, and A. Munafò, "Aurora, a multi-sensor dataset for robotic ocean exploration," *The International Journal of Robotics Research*, vol. 41, no. 5, pp. 461–469, 2022.

[15] C. Boittiaux, C. Dune, M. Ferrera, A. Arnaubec, R. Marxer, M. Matabos, L. Van Audenhaege, and V. Hugel, "Eiffel tower: A deep-sea underwater dataset for long-term visual localization," *The International Journal of Robotics Research*, vol. 42, no. 9, pp. 689–699, 2023.

[16] S. Xu, J. Scharff Willners, J. Roe, S. Katagiri, T. Luczynski, Y. Petillot, and S. Wang, "Tank dataset: An underwater multi-sensor dataset for slam evaluation," *The International Journal of Robotics Research*, p. 02783649251364904, 2025.

[17] B. Kerbl, G. Kopanas, T. Leimkühler, G. Drettakis, *et al.*, "3d gaussian splatting for real-time radiance field rendering." *ACM Trans. Graph.*, vol. 42, no. 4, pp. 139–1, 2023.

[18] J. Sola, "Quaternion kinematics for the error-state kalman filter," *arXiv preprint arXiv:1711.02508*, 2017.

[19] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, "Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam," *IEEE transactions on robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.

[20] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE transactions on robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.

[21] M. M. M. Manhães, S. A. Scherer, M. Voss, L. R. Douat, and T. Rauschenbach, "Uuv simulator: A gazebo-based package for underwater intervention and multi-robot simulation," in *Oceans 2016 Mts/Ieee Monterey*. Ieee, 2016, pp. 1–8.

[22] M. Grupp, "evo: Python package for the evaluation of odometry and slam," 2017.

[23] Y. Zhang, F. Tosi, S. Mattoccia, and M. Poggi, "Go-slam: Global optimization for consistent 3d instant reconstruction," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 3727–3737.

[24] H. Matsuki, R. Murai, P. H. Kelly, and A. J. Davison, "Gaussian splatting slam," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2024, pp. 18 039–18 048.

[25] H. Huang, L. Li, H. Cheng, and S.-K. Yeung, "Photo-slam: Real-time simultaneous localization and photorealistic mapping for monocular stereo and rgb-d cameras," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 21 584–21 593.

[26] K. Wu, Z. Zhang, M. Tie, Z. Ai, Z. Gan, and W. Ding, "Vings-mono: Visual-inertial gaussian splatting monocular slam in large scenes," *IEEE Transactions on Robotics*, 2025.