

Spectral Whitening and Confidence Fusion for Robust Sonar Scan Matching

Abstract— Robust sonar scan registration is a cornerstone of underwater mapping and navigation. While Fourier-SOFT in 2D (FS2D) and related Fourier-based approaches provide an efficient foundation, they remain sensitive to yaw ambiguities and translation reliability in cluttered conditions. In this work, we strengthen the FS2D pipeline with three targeted modifications: (i) multi-ring spectral whitening for more stable yaw estimation, (ii) phase-only cross-power spectrum with peak-to-sidelobe ratio (PSR) scoring for sharper translation recovery, and (iii) fusion of PSR with zero-mean normalized cross-correlation (ZNCC) for more reliable hypothesis selection. We validate across synthetic ground-truth transforms, Gazebo simulations of river and shipwreck environments, and a real-world pool experiment with obstructions. Across these settings, our approach consistently improves yaw stability and translation accuracy while maintaining runtime efficiency. These findings demonstrate that confidence-aware spectral correlation is a practical and efficient path toward more reliable sonar-based SLAM in cluttered underwater environments.

I. INTRODUCTION

Reliable navigation and mapping are fundamental capabilities for autonomous underwater vehicles (AUVs). Accurate localization underpins tasks such as infrastructure inspection, environmental monitoring, and search-and-rescue, where GPS is unavailable and visual cues are often degraded by turbidity, and low light. In these conditions, sonar remains the most reliable sensor [1]–[3].

To build consistent maps, consecutive sonar scans must be aligned. Unlike instantaneous 2D images, a Ping360 scan is acquired over 40 seconds, during which the robot may move. As a result, consecutive scans are only partially overlapping: they contain repeated structures from the previous scan as well as new regions, while some areas may drop out entirely. *Sonar scan registration* provides the relative pose between views and is a core component of sonar-based SLAM [1], [4]–[11]. As a result, without robust registration, pose graph optimization and mapping drift rapidly accumulate, and degrade navigation performance.

Feature-based approaches have been explored, such as extracting keypoints from side-scan sonar images [12], using graph matching on sonar intensity patterns [13], or adapting visual descriptors such as SIFT/SURF to sonar mosaicing [14]. However, sonar imagery is often speckled, low-contrast, and view-dependent. This makes reliable keypoint extraction difficult. Early probabilistic scan-matching frameworks [15] also struggle in these conditions.

Spectral methods offer an attractive alternative. By operating in the Fourier domain, they exploit shift and rotation invariances and are computationally efficient. Classical phase correlation methods [16], [17] and log-polar Fourier

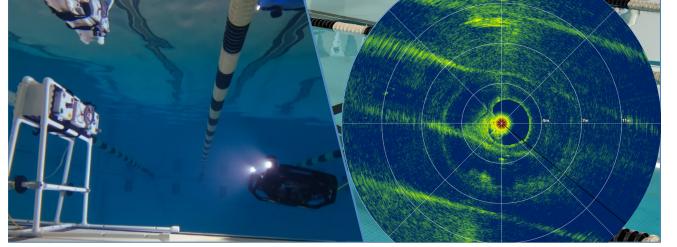


Fig. 1. Experimental platform overview. Left: BlueROV2 Heavy approaching a modular task box designed to emulate underwater intervention tasks. Right: Corresponding Ping360 sonar scan showing long-range detection of the task box in a pool environment.

transforms [18], [19] have been widely studied in vision, radar, and medical imaging. Recent work further shows that preprocessing and normalization can improve phase correlation robustness on featureless imagery [20], conditions that resemble sonar. In the sonar domain, Fourier-based methods [1], [4] and FS2D [21] demonstrate the promise of spectral approaches.

Spectral whitening is one such normalization strategy. It rescales the Fourier magnitude spectrum so that all frequency bands contribute more equally, suppressing bias from dominant low-frequency components and emphasizing angular variations. This improves correlation stability in cluttered or low-contrast imagery, and makes it particularly suitable for sonar scan registration where raw intensities can vary widely across scans.

Despite this progress, two challenges remain prominent in practice: (i) yaw estimation with FS2D often yields competing peaks due to reliance on a single frequency ring; and (ii) translation correlation peaks vary in absolute height across scans, which complicates hypothesis selection under speckle, padding, and intensity drift.

To address these challenges, we revisit the FS2D pipeline and introduce a strengthened approach based on *spectral whitening and confidence fusion*. Our contributions are three-fold:

- We stabilize yaw estimation by sampling and whitening multiple concentric frequency rings.
- We adopt phase-only correlation with PSR scoring for sharper and more consistent translation recovery.
- We introduce PSR-ZNCC fusion for hypothesis selection, combining spectral peak sharpness with image-level consistency.

We evaluate the approach on synthetic ground-truth transforms (rotation sweeps and random translations), Gazebo

simulated underwater environments, and real-world pool experiments with clutter and occlusions. Together, these experiments show improved robustness and reduced drift without additional runtime cost.

Fig. 1 shows our experimental platform: a BlueROV2 Heavy with a Ping360 mechanical scanning sonar and a Blueprint Subsea SeaTrac USBL for long-range mapping and navigation in a pool environment. The object of interest is a modular *task box*, designed to emulate real-world underwater manipulation tasks including valve turning, line cutting, and button pressing. Our objective is to detect and localize the task box from long range using sonar for mapping and path planning, then transition to vision-based perception for close-range manipulation. A further advantage of the Ping360 is its low cost, enabling large-scale deployment in multi-robot or long-duration missions.

The rest of the paper is structured as follows. A discussion of related works is provided in Sec.II. Methodology is described in Sec.III. Experiments and results are presented in Sec.IV. Sec.V concludes the paper.

II. RELATED WORKS

A. Feature-Based Methods

Previous works have adapted visual feature detectors and descriptors to sonar imagery. For instance, side-scan sonar images have been registered using keypoint matching frameworks [12], while graph-based correspondences have been leveraged to improve loop closure and scan alignment [13]. Sonar-adapted descriptors (e.g., SIFT, SURF, and variants) have also been proposed for SLAM [14], but their reliability is limited by speckle and low contrast. Mechanical scanning imaging sonar (MSIS) data such as that from the Blue Robotics Ping360 is particularly challenging. Early probabilistic scan-matching frameworks like MSISpIC [15] applied Bayesian models for pose refinement but remained sensitive to initialization and clutter. More recent distributed frameworks such as DRACo-SLAM2 [8] and graph-based MSIS SLAM [9] highlight the critical role of robust pairwise scan registration in large-scale sonar mapping, motivating our focus on strengthening FS2D to improve yaw stability and translation accuracy.

B. Spectral Correlation Methods

Spectral methods utilize Fourier-domain representations to achieve invariance to translation and rotation. Classical phase correlation [16], [17] has long been applied in radar, and vision, while extensions into the log-polar domain [18], [19] and multiple polar Fourier transforms [18] enable estimation of similarity transforms. Preprocessing and normalization have also been shown to improve the robustness of phase correlation on featureless imagery [20], conditions that resemble MSIS returns. In the sonar domain, spectral approaches have been leveraged for mosaicing [1], [4] and as front-ends for SLAM pipelines [5]. Recent developments emphasize robust peak scoring and multi-sensor integration [10], [22], underscoring the need for confidence-aware registration

methods. These advances prompt our adoption of phase-only correlation and explicit confidence fusion.

C. Mechanical Scanning Sonar Registration

Mechanical scanning sonars are widely used for navigation and mapping in low-visibility environments due to their affordability and robustness. Hansen and Birk introduced FS2D [21], which applies the spherical Fourier transform (SOFT) to the magnitude spectra of MSIS pings to estimate yaw from spectral correlation and translation via phase correlation. Earlier spectral registration methods for sonar range data [23] and broader reviews of AUV navigation [24] highlight these challenges, which we tackle through improved yaw estimation and translation recovery within FS2D.

III. METHODOLOGY

Our approach extends the FS2D pipeline [21] with three targeted modifications designed to improve robustness in sonar scan registration: (i) multi-ring spectral whitening for yaw estimation, (ii) phase-only cross-power spectrum for translation, and (iii) PSR–ZNCC fusion rule for hypothesis selection.

A. Multi-Ring Spectral Whitening

Fourier magnitudes are normalized across multiple concentric rings so that angular information dominates over global intensity bias.

In FS2D, yaw estimation is obtained by correlating the Fourier magnitude spectrum in the spherical Fourier domain. In practice, this is often implemented by sampling one or a few high-radius rings, where angular variation is most pronounced. However, reliance on a narrow radial band makes the estimate sensitive to the chosen radius and can introduce bias from low-frequency components.

We introduce *multi-ring spectral whitening*. A band of L_r concentric rings is sampled around $r_0 = N/2 - 2$. Each ring is normalized by its mean intensity, approximating spectral whitening, and smoothly weighted by a Hamming window:

$$\widetilde{M}_i(\theta, \phi) = \frac{\sum_{r=r_s}^{r_e} w_r (|F_i|(x_r, y_r) / \bar{M}_i(r))}{\sum_{r=r_s}^{r_e} w_r}, \quad (1)$$

where $\bar{M}_i(r)$ is the mean magnitude along ring r .

B. Yaw Estimation via SOFT Correlation

Following spectral whitening, the sonar data are transformed into the spherical Fourier domain. Representing the magnitude spectra on the sphere enables the use of SOFT correlation, which is designed to detect relative rotations.

We compute the correlation between the whitened spectra \widetilde{M}_1 and \widetilde{M}_2 using SOFT, which yields a 2D response surface $C(\theta, \phi)$. To focus on horizontal alignment, this surface is projected onto the azimuthal axis to produce a 1D yaw response curve. Local maxima on this curve define candidate yaw angles, denoted $\{\hat{\psi}_k\}$, where k indexes the detected peaks.

TABLE I
ROTATION-ONLY EXPERIMENT. YAW RMSE (DEG) AND RUNTIME (MS).

Method	$N = 64$		$N = 256$	
	RMSE [°]	Time [ms]	RMSE [°]	Time [ms]
FS2D (orig)	36.9	59.1 ± 10.3	22.9	4672.8 ± 234.8
FS2D+PhaseOnly (ours)	43.0	60.1 ± 10.8	41.17	(4669.8 ± 236.0)
FS2D+ZNCC (ours)	33.5	67.6 ± 11.4	20.1	4760.9 ± 231.5

TABLE II
TRANSLATION-ONLY EXPERIMENT. TRANSLATION ERROR (PIXELS) AND AVERAGE RUNTIME (MS).

Method	$N = 64$		$N = 256$	
	Error [px]	Time [ms]	Error [px]	Time [ms]
FS2D (orig)	1.36 ± 6.3	56.8 ± 10.7	0.5 ± 8.9	4597.6 ± 216.3
FS2D+PhaseOnly (ours)	0.06 ± 1.3	56.62 ± 10.9	0.0 ± 0.0	(4593.1 ± 206.2)
FS2D+ZNCC (ours)	0.33 ± 3.45	66.7 ± 14.6	0.1 ± 4.0	4701.1 ± 216.0

C. Phase-Only Translation Estimation

Translation is recovered by suppressing amplitude information and aligning scans using only phase cues.

For each yaw hypothesis $\hat{\psi}_k$, the second image is derotated and aligned with zero-padding. We then compute the *phase-only cross-power spectrum*:

$$\hat{R}(\mathbf{k}) = \frac{F_1(\mathbf{k}) \overline{F_2(\mathbf{k})}}{|F_1(\mathbf{k}) \overline{F_2(\mathbf{k})}| + \varepsilon}, \quad (2)$$

with ε preventing division by zero. The inverse FFT yields a correlation surface $r(x, y)$, where the maximum value

$$r_{\max} = \max_{x, y} r(x, y)$$

at location (x^*, y^*) indicates the translation offset.

Robustness is quantified via the peak-to-sidelobe ratio (PSR):

$$\text{PSR} = \frac{r_{\max} - \mu_{\text{SL}}}{\sigma_{\text{SL}}}, \quad (3)$$

where $\mu_{\text{SL}}, \sigma_{\text{SL}}$ are statistics of sidelobe regions computed outside a small window (typically $w \times w$ with $w \approx 11$, tunable) centered on the peak.

The resulting translation estimate for hypothesis k is denoted $\mathbf{t}_k = (x^*, y^*)$.

D. Median ZNCC Consistency

Visual consistency between aligned scans is verified using zero-mean normalized cross correlation (ZNCC) on local crops.

While PSR measures sharpness of the correlation peak, it does not guarantee visual consistency. We therefore compute zero-mean normalized cross correlation (ZNCC) over B crops of the derotated pair:

$$\widetilde{\text{ZNCC}} = \text{median}\{\text{ZNCC}(\mathcal{W}_b)\}_{b=1}^B. \quad (4)$$

Taking the median across random or tiled crops reduces sensitivity to local noise, occlusions, and zero-padding artifacts.

E. PSR-ZNCC Fusion

Confidence scores from PSR and ZNCC are combined to select the most reliable alignment.

Each hypothesis h_k yields $(\hat{\psi}_k, \mathbf{t}_k, \text{PSR}_k, \widetilde{\text{ZNCC}}_k)$. Hypotheses are first filtered using thresholds $(\tau_{\text{psr}}, \tau_{\text{zncc}})$; if none passes, the trial is marked low confidence. For the remaining candidates, metrics are normalized and combined. If one metric is dominant, it is trusted directly; otherwise, a weighted fusion is applied:

$$S_k = \alpha \widehat{\text{PSR}}_k + (1 - \alpha) \widehat{\text{ZNCC}}_k, \quad (5)$$

with $\alpha = 0.6$ in all experiments. This fusion leverages the complementary strengths of PSR (peak sharpness) and ZNCC (image-level consistency).

F. Complexity

All FFT operations remain $O(N^2 \log N)$. Multi-ring whitening adds only $O(L_r)$ radial samples, negligible for $L_r \leq 11$. Phase-only normalization is a pointwise operation, and PSR-ZNCC fusion involves a small number of local correlations. Thus, runtime is dominated by the SOFT correlation stage, as in FS2D.

IV. EXPERIMENTS AND RESULTS

We evaluate our method across three complementary settings: (A) synthetic transforms with exact ground truth, (B) simulated consecutive sonar scans in Gazebo environments, and (C) real-world pool experiments with obstructions. Each setting addresses a specific goal: correctness under controlled transforms, robustness in cluttered but repeatable simulation, and generalization to real acoustic artifacts.

A. Synthetic Transforms with Ground Truth

Synthetic data enables precise ground-truth evaluation with controlled transforms. Separate rotation and translation tests isolate FS2D failure modes and assess our modifications, while combined experiments capture their interaction to benchmark overall registration robustness.

TABLE III

TRANSLATION-ROTATION COMBINED EXPERIMENT. YAW RMSE (DEG), TRANSLATION ERROR (PIXELS) AND AVERAGE RUNTIME (MS).

Method	$N = 64$			$N = 256$		
	RMSE [°]	Error [px]	Time [ms]	RMSE [°]	Error [px]	Time [ms]
FS2D (orig)	1.88	0.004 ± 0.07	58.0 ± 11.0	0.45	0.003 ± 0.05	4696.5 ± 298.1
FS2D+PhaseOnly (ours)	1.88	0.0 ± 0.0	58.0 ± 10.8	0.45	0.02 ± 0.2	4693.0 ± 288.6
FS2D+ZNCC (ours)	1.85	0.0 ± 0.0	67.1 ± 12.9	0.43 ± 0.02	0.02 ± 0.14	4792.7 ± 320.2

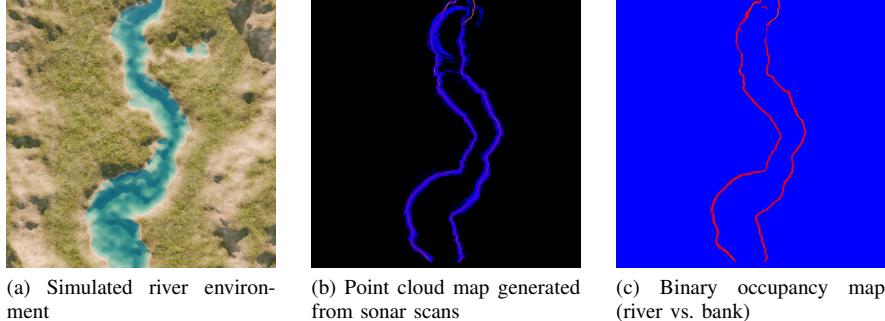


Fig. 2. Simulated river experiment. (a) Ground-truth Gazebo environment with elongated banks and curved channel. (b) Map reconstruction from sonar returns with dense edge structure along the banks. (c) Binary occupancy map derived from the reconstruction, used for navigation and planning.

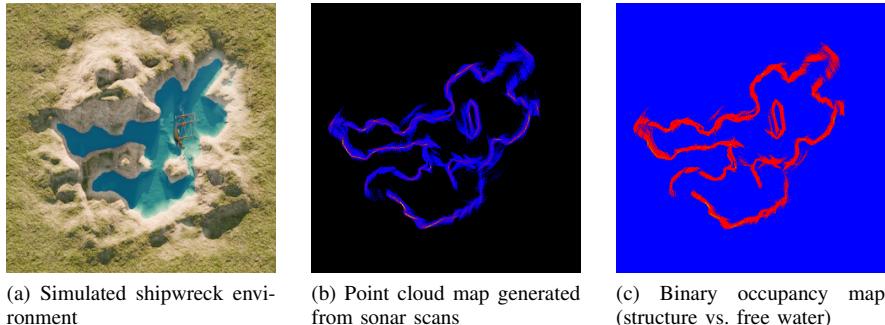


Fig. 3. Simulated shipwreck experiment. (a) Ground-truth Gazebo environment containing compact, high-contrast geometry with occlusions. (b) Sonar-based reconstruction of the wreck, capturing prominent edges and symmetric structures. (c) Binary occupancy map derived from the reconstruction.

In this experiment, 111 sonar scans are collected from Ping360 sonar and resampled to square grids $N \in \{64, 256\}$. Synthetic target images are generated by:

- *Rotation-only*: Target scans are generated by applying rotations $\psi \sim \mathcal{U}(5^\circ, 120^\circ)$ in 5° increments, using bilinear interpolation with zero padding to avoid wrap-around. This sweep tests yaw stability over a wide range.
- *Translation-only*: Target scans are generated by translations $(t_x, t_y) \sim \mathcal{U}(-5, 5)$ m, with 10 targets per scan to reflect realistic short-baseline displacements. Independent seeds for t_x and t_y prevent correlations.
- *Rotation+Translation*: Perturbations are restricted to smaller motions: $\psi \sim \mathcal{U}(5^\circ, 15^\circ)$ combined with translations < 1 m. This evaluates robustness under modest simultaneous yaw and translation.

The rotation-only experiment (Table I) shows that our ZNCC-enhanced variant achieves the lowest yaw RMSE at both grid sizes (33.5° at $N = 64$, 20.1° at $N = 256$),

outperforming both the original FS2D and the phase-only variant. Phase-only normalization alone degrades yaw accuracy, especially at high resolution. This suggests that magnitude cues suppressed by normalization aid yaw estimation. When combined with ZNCC, this drawback is mitigated, and yields both higher stability and accuracy. Performance also improves from $N = 64$ to $N = 256$, which indicates that finer spatial sampling reduces aliasing and strengthens yaw discriminability.

The translation-only experiment (Table II) shows the complementary strengths of our modifications. Phase-only normalization yields nearly perfect translation recovery (0.0 ± 0.0 px error at $N = 256$), and produces the expected sharp correlation impulse. ZNCC further reduces translation error at $N = 64$, making it more robust in lower-resolution conditions where sidelobe structure is stronger.

The combined rotation–translation experiment Table III) confirms these findings under modest perturbations. FS2D+ZNCC attains the best yaw RMSE (1.85° at $N = 64$,

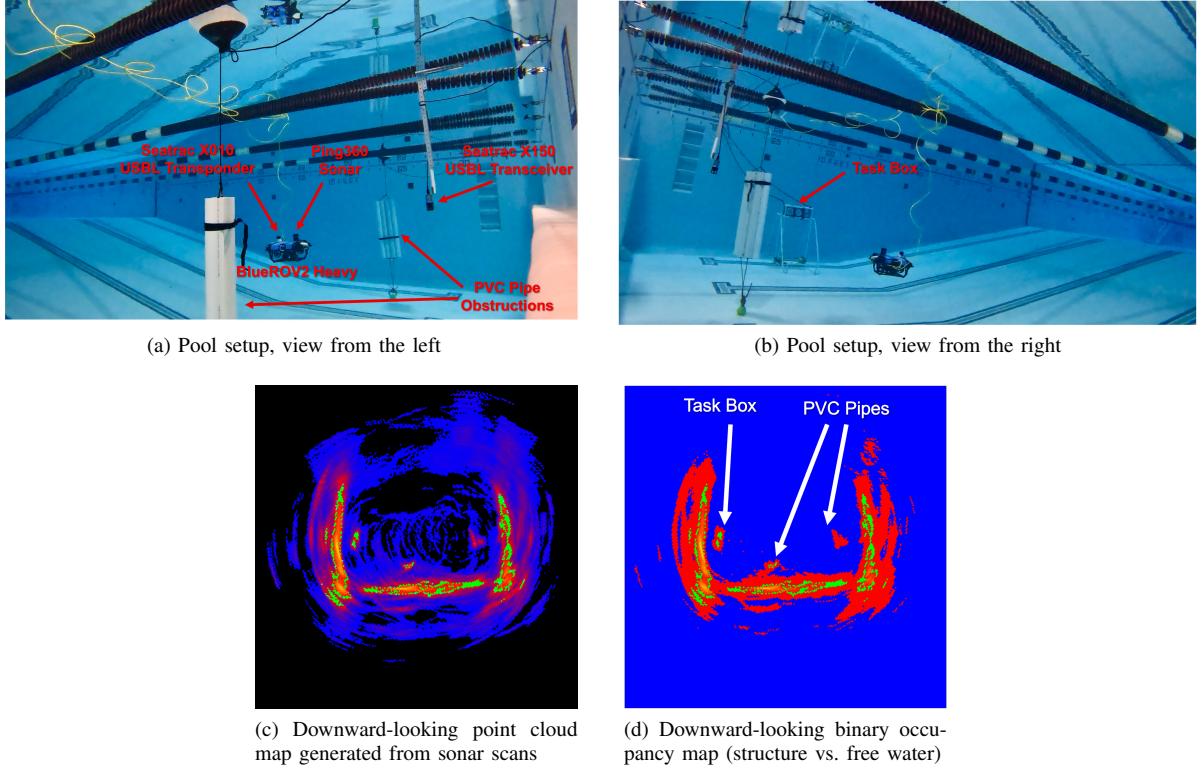


Fig. 4. Experimental pool mapping. (a-b) Ground-truth photo of pool site. (b) Point cloud reconstruction from accumulated Ping360 sonar scans, capturing the pool walls and corner structures. (c) Binary occupancy map derived from the reconstruction, highlighting obstacles and free-water regions for navigation.

0.43° at $N = 256$), while both phase-only and ZNCC yield perfect or near-perfect translation recovery. FS2D (orig) remains competitive in translation at $N = 256$, but lags slightly in yaw stability. Runtime differences are negligible across all methods, with our variants matching the efficiency of FS2D.

Taken together, these results highlight an important insight: *phase-only correlation is critical for accurate translation, while ZNCC fusion is essential for robust yaw estimation*. FS2D alone struggles at large rotations and with low-resolution inputs, but our targeted modifications directly address these weaknesses.

B. Gazebo Simulation (River and Shipwreck)

Simulation provides an essential intermediate step between purely synthetic transforms and real-world experiments. By deploying sonar in simulated environments, we can introduce clutter, noise, and structural complexity while still retaining ground-truth geometry for validation. This allows controlled evaluation of both yaw estimation and translation robustness in scenarios that closely resemble operational conditions.

We evaluate in two representative Gazebo environments (Figs. 2–3) modeled in Blender. The *river* world features elongated banks with sparse clutter, where extended low-texture structures hinder alignment and accentuate drift. The *shipwreck* world contains compact, high-contrast geometry with repeated beams and occlusions, generating aliased sonar returns that challenge yaw disambiguation. Together, these

environments expose distinct weaknesses of FS2D and provide a comprehensive benchmark for our improvements.

In both settings, a simulated Ping360 in Gazebo Harmonic synthesizes sonar returns, while a simulated LiDAR provides range. Each ping is rasterized into 1200 intensity bins by painting a radial Gaussian at the measured range, with exponential attenuation and multiplicative speckle noise to approximate acoustic backscatter. The sensor sweeps one bearing at a time clockwise, replicating Ping360 operation and yielding time-synchronized sonar-like frames for registration experiments.

C. Real-World Pool Experiment with Obstructions

Real sonar data exhibits multipath, specular reflections, and acoustic shadows that are difficult to simulate. To assess practical robustness, we conducted a controlled pool experiment with obstructions emulating underwater manipulation. A Blue Robotics Ping360 was mounted on a BlueROV2 and operated in a pool site containing PVC pipe frames and a modular task box, introducing occlusions, repeated linear features, and compact corners that stress registration. The setup follows our OCEANS study but with added obstructions.

Localization relied only on an onboard IMU and a Blueprint Subsea SeaTrac USBL system, creating a constrained navigation scenario. The IMU accumulates drift during dead reckoning, while USBL updates are infrequent (seconds) and can jump under reflections or multipath.

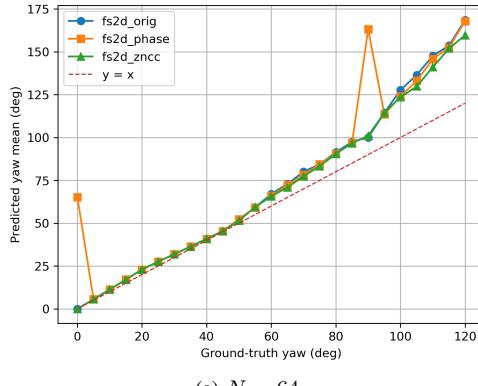
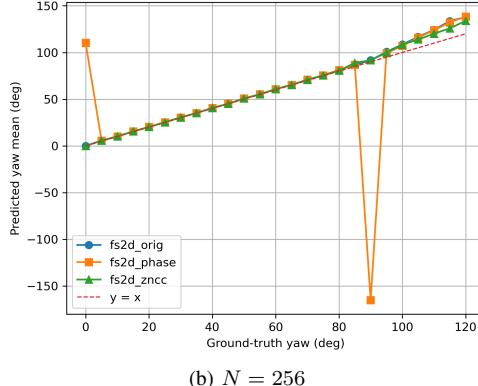
(a) $N = 64$ (b) $N = 256$

Fig. 5. Predicted yaw mean versus ground-truth in rotation-only experiment.

This makes scan-to-scan registration difficult as yaw drift accumulates between updates and translation priors become noisy, effectively restricting mapping to the deep end of the pool where USBL geometry is more reliable.

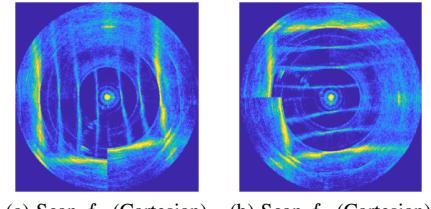
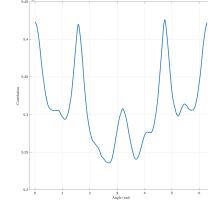
Despite these challenges, the experiment highlights the robustness of our method under realistic sensor limitations. The combination of sonar ambiguities, inertial drift, and sparse position fixes shows the necessity of robust correlation: without phase-only constraints and confidence fusion, registration quickly degrades. These results underscore the value of complementary sensing and demonstrate that even with minimal sensors, our method sustains mapping performance in cluttered underwater environments.

D. Failure Cases and Limitations

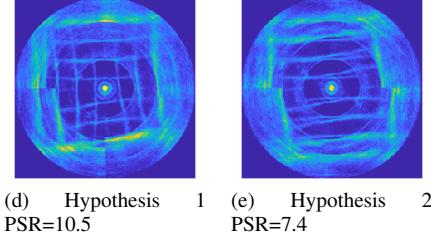
While our modifications improve robustness overall, certain structural conditions remain challenging.

Environments with strong symmetries (e.g., parallel pool lanes) produce competing yaw peaks in the correlation surface. This is most evident for FS2D+PhaseOnly, which shows anomalies near 0° and 90° in Fig. 5 and Fig. 6. Since yaw candidates are not disambiguated directly, the final selection defaults to the hypothesis with the highest PSR score, which is computed on the translation surface. When multiple yaws yield comparable correlation energy, this translation-based criterion often locks onto the wrong mode.

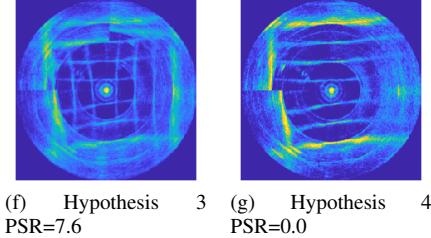
Performance also degrades faster at lower resolution ($N = 64$). As seen in Fig. 5a, yaw error grows beyond 60° ,

(a) Scan f_1 (Cartesian) (b) Scan f_2 (Cartesian)

(c) SOFT correlation vs. yaw



(d) Hypothesis 1 PSR=10.5 (e) Hypothesis 2 PSR=7.4



(f) Hypothesis 3 PSR=7.6 (g) Hypothesis 4 PSR=0.0

Fig. 6. Ambiguous registration from FS2D+PhaseOnly. (a-b) Pair of sonar scans offset by a 90° rigid rotation. (c) SOFT correlation curve with multiple peaks. (d-g) Top yaw hypotheses visualized as derotated overlays.

reflecting aliasing and loss of frequency detail at coarse sampling. This resolution dependence suggests that low-resolution scans may require additional priors (e.g., inertial constraints) to remain reliable under larger rotations.

V. CONCLUSIONS

This work revisited the FS2D sonar registration pipeline with three targeted modifications: multi-ring spectral whitening, phase-only translation estimation, and PSR-ZNCC confidence fusion. These additions improved yaw stability and translation accuracy across synthetic, simulated, and pool experiments, while preserving Fourier-domain efficiency.

Two broader insights emerge. Phase-only normalization is well suited for translation recovery, while image-level consistency checks are critical for resolving yaw ambiguities. Moreover, lightweight preprocessing such as multi-ring whitening can enhance robustness without runtime cost. Overall, these findings suggest that confidence-aware spectral methods offer a practical foundation for sonar-based SLAM.

REFERENCES

- [1] L. Paull, S. Saeedi, M. Seto, and H. Li, "AUV navigation and localization: A review," *IEEE Journal of Oceanic Engineering*, vol. 39, no. 1, pp. 131–149, 2014. doi:10.1109/JOE.2013.2278891.
- [2] M. Fallon, M. Kaess, H. Johannsson, and J. J. Leonard, "Efficient AUV navigation fusing acoustic ranging and side-scan sonar," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 2011, pp. 2398–2405. doi:10.1109/ICRA.2011.5979665.
- [3] Z. Zhang, J. Xie, Y. Ling, and J. Folkesson, "A fully-automatic side-scan sonar SLAM framework," arXiv:2304.01854 [cs.RO], 2023.
- [4] N. Hurtós, D. Ribas, X. Cuf, Y. Petillot, and J. Salvi, "Fourier-based registrations for two-dimensional forward-looking sonar image mosaicing," in **Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, 2012, pp. 5298–5305. doi:10.1109/IROS.2012.6385813.*
- [5] J. Li, M. Kaess, R. M. Eustice, and M. Johnson-Roberson, "Pose-graph SLAM using forward-looking sonar," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2330–2337, 2018. doi:10.1109/LRA.2018.2810958.
- [6] L. Chen, A. Yang, H. Hu, and W. Naeem, "RBPF-MSIS: Toward Rao-Blackwellized particle filter SLAM for autonomous underwater vehicle with slow mechanical scanning imaging sonar," *IEEE Systems Journal*, vol. 14, no. 3, pp. 3301–3312, 2020. doi:10.1109/JST.2019.294694.
- [7] P. Vial, J. Solà, N. Palomeras, and M. Carreras, "Underwater Pose SLAM using GMM scan matching for a mechanical profiling sonar," *Journal of Field Robotics*, vol. 41, no. 3, pp. 511–537, 2024. doi:10.1002/rob.22272.
- [8] Y. Huang, J. McConnell, X. Lin, and B. Englot, "DRACo-SLAM2: Distributed robust acoustic communication-efficient SLAM for imaging sonar equipped underwater robot teams," arXiv:2507.23629, 2025.
- [9] I. T. Ruiz, S. de Raucourt, Y. Petillot, and D. M. Lane, "Concurrent mapping and localization using sidescan sonar," *IEEE Journal of Oceanic Engineering*, vol. 29, no. 2, pp. 442–456, 2004. doi:10.1109/JOE.2004.829812.
- [10] S. Rahman, A. Quattrini Li, and I. Rekleitis, "SVIn2: A Multi-Sensor Fusion-Based Underwater SLAM System," *The International Journal of Robotics Research*, 2022. doi:10.1177/02783649221110259.
- [11] X. Wang, X. Fan, P. Shi, J. Ni, and Z. Zhou, "An Overview of Key SLAM Technologies for Underwater Scenes," *Remote Sensing*, vol. 15, no. 10, 2023. doi:10.3390/rs15102496.
- [12] X. Han, J. Sun, S. Zhang, J. Dong, and H. Yu, "Sonar-based simultaneous localization and mapping using the semi-direct method," *Journal of Marine Science and Engineering*, vol. 12, no. 12, p. 2234, 2024. doi:10.3390/jmse12122234.
- [13] L. Zhuang, X. Chen, W. Lu, and Y. Yan, "Graph matching for underwater simultaneous localization and mapping using multibeam sonar imaging," *Journal of Marine Science and Engineering*, vol. 12, no. 10, p. 1859, 2024. doi:10.3390/jmse12101859.
- [14] M. D. Aykin and M. Negahdaripour, "On feature matching and image registration for two-dimensional forward-scan sonar imaging," *Journal of Field Robotics*, vol. 30, no. 3, pp. 352–372, 2013. doi:10.1002/rob.21461.
- [15] E. Hernández, P. Ridao, D. Ribas, and A. Mallios, "Probabilistic Sonar Scan Matching for an AUV (MSISpIC)," in *Proc. *IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, 2009, pp. —. doi:10.1109/IROS.2009.5354656.*
- [16] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Transactions on Image Processing*, vol. 5, no. 8, pp. 1266–1271, 1996. doi:10.1109/83.506761.
- [17] Q.-S. Chen, M. Deffrise, and F. Deconinck, "Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 12, pp. 1156–1168, 1994. doi:10.1109/34.387502.
- [18] Y. Dong, W. Jiao, T. Long, G. He, and C. Gong, "An extension of phase correlation-based image registration to estimate similarity transform using multiple polar Fourier transform," *Remote Sensing*, vol. 10, no. 11, p. 1719, 2018. doi:10.3390/rs10111719.
- [19] S. Chelbi, and A. Mekhmoukh, "Features-based image registration using cross-correlation and log-polar transform," *Alexandria Engineering Journal*, vol. 57, no. 4, pp. 2297–2307, 2018. doi:10.1016/j.aej.2017.08.018.
- [20] J. Lentz, H. Sevil, and D. Fries, "Image preprocessing to enhance phase correlation of featureless imagery," *Scientific Reports*, vol. 15, no. 1, 2025, Art. no. 1234. doi:10.1038/s41598-025-XXXXX.
- [21] T. Hansen and A. Birk, "Using Registration with Fourier-SOFT in 2D (FS2D) for robust scan matching of sonar range data," in *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2023, pp. 12067–12073. doi:10.1109/ICRA48891.2023.10160519.
- [22] M. Heshmat, L. Saoud, M. Abujabal, A. Sultan, M. Elmezain, L. Seneviratne, and I. Hussain, "Underwater SLAM Meets Deep Learning: Challenges, Multi-Sensor Integration, and Future Directions," *Sensors*, 2025.
- [23] H. Bülow, M. Pfingsthorn, and A. Birk, "Using robust spectral registration for scan matching of sonar range data," in *Proc. IFAC Symp. Intelligent Autonomous Vehicles (IAV)*, 2010, pp. 1–6.
- [24] J. J. Leonard and A. Bahr, "Autonomous underwater vehicle navigation," in *Springer Handbook of Ocean Engineering*, M. R. Dhanak and N. I. Xiros, Eds. Cham: Springer, 2016, pp. 341–358. doi:10.1007/978-3-319-16649-0_14.