Analyzing the Impact of Integrated Signal and Image Processing Techniques on Pattern Recognition and Decision Accuracy in Complex Systems

[1*]Shinnu Jangra, [2]C. J. Jacintha Cathrin, [3]Dr. Poornima Raikar, [4]Harish Chandra Sharma, [5]Dr. M. Mary Adline Priya, [6]Deepak Murlidhar Sonje, [7]Nyan Lin Satt

[1*]Assistant Professor, Department of Artificial Intelligence and machine learning, Manipal University Jaipur, Jaipur, Rajasthan 303007, Email Id: shinnujangra204@gmail.com, Orcid Id: https://orcid.org/0000-0002-6636-4291
[2]Assistant Professor, Department of Electronics and communication Engineering, Specialization in Communication system & Transmission line, Arunai Engineering College Autonomous , Tamil Nadu, Tiruvannamalai Dst, 605757, Email Id: jacinthacathrin@arunai.org
[3]Assistant Professor, Department of Computer Science and Engineering (AI&ML), Specialization in Image processing, Artificial Intelligence, KLS VDIT Haliyal, Pin code: 581329, Haliyal, Karnataka, India, Orcid Id: 0000-0003-4132-4848, Email Id: pnr@klsvdit.edu.in
[4]Associate Professor, SGRR University, Dehradun, Email Id: hcs19@yahoo.com, Orcid Id: 0000-0002-1263-9325
[5]Associate Professor, Specialization in Image Processing, Department of Medical Electronics, Saveetha Engineering College, Pin code: 600056, Chennai, India, Orcid Id: 0000-0001-6528-1518, Email Id: michaelpriya21@gmail.com
[6]Associate professor, Department of Electrical engineering, R H Sapat College of Engineering Management Studies and Research Nashik, Savitribai Phule Pune University, Pune, Orcid Id: 0000 0001 8511 6288, Email Id: deepaksonje123@gmail.com
[7]Master In Information Technology, Inti International University, Malaysia, Email Id: i25033498@student.newinti.edu.my, Orcid Id:0009-0005-3592-5091

Abstract
Integrated signal and image processing is being used more to enhance pattern recognition and decision quality in complex sensing conditions, but the quality of benefit of combining underwater sonar and camera data to ensure continuous localization is context-dependent. This work quantifies the effect of feature-level fusion of multibeam sonar and monocular camera imagery on diver localization accuracy. A publicly available underwater dataset was synchronized to produce 1,193 paired samples across four experimental sets. Each frame was converted to grayscale, resized to 64×64 pixels, and represented using global intensity statistics and regional mean features. Three Random Forest regression models were trained under identical conditions: sonar-only, camera-only, and fused features. Localization error was evaluated using Euclidean distance and standard regression metrics. Sonar-only processing produced the highest error, camera-only significantly reduced error, and multimodal fusion achieved the lowest overall RMSE and MAE. The greatest improvements from fusion were observed in scenarios with higher spatial variability, while in visually favorable conditions, camera-only performance remained competitive. These findings indicate that integrating acoustic and visual features enhances overall localization robustness, though the magnitude of improvement depends on environmental characteristics. Adaptive fusion strategies may further improve stability in real-world underwater systems.
Keywords: machine learning, computer vision, data fusion, underwater robotics, pattern recognition

1. Introduction

Recent systems of pattern recognition and smart decision-making lie in technology: signal, image processing. The computational processes and sensing technologies have undergone an improvement in the past decade, enabling more advanced processing structures (which can be utilized in dynamic and complex environments) [1]. Such developments have transcended the ideas of traditional filtering and enhancement models to far more integrated algorithm designs that include mathematical modelling and the use of data-based learning [2]. The recent theoretical and practical papers emphasize the augmented importance of combined signal and image processing in real-world systems that demand non-homogeneous data to sensory data to be decoded simultaneously [3]. Pattern recognition also plays a key part in such systems, particularly in cases where there is uncertainty in the environment, noise, distortions due to modality and so forth [4]. The feature extraction strategies utilized in order to transform the raw signals to the discriminative ones are of great importance in recognition pipelines and decision-making pipelines [5]. The pattern recognition systems have evolved from the classical rule-based classification to the nonlinear relationship-based neural network classification systems [6]. Despite these advancements, the basic principles of feature representation and permanence in classifiers remain the key pillars

of the process of providing stable execution to the various conditions of operation [7]. The issues in underwater environments are of special interest, as optical degradation and acoustic noise introduce significant complexity to the sensing and interpretation tasks.

Signal processing techniques employed to address visual deficiencies to create underwater perception systems rely on sonar. It has been demonstrated that the sonar image segmentation and quality assessment technologies improve the acoustic scene interpretation under challenging conditions [8]. How acoustic information can be used to characterize the environment and identify structures in the environment is also evidenced by the fact that marine sonar images can be used to identify the presence of structures using texture-based recognition methods [9]. Conversely, there are still developments of object detection and localization processes that are performed using the vision-based processing methods in situations where the degree of visual clarity is high. The types of machine learning structures based on image pattern recognition have shown good results in the localization of objects, classification and structural interpretation [10]. However, camera-only systems too are susceptible to lighting, turbidity and occlusion variations, and, therefore, cannot be used in the changing environment of underwater. The latest tendencies in the sphere of computational intelligence are devoted to the synthesis of algorithms and their performance by emphasizing the necessity of the combination of different sources of information to enhance the consistency of the decision [11]. Another reason why the adaptive strategies of processing should be employed to handle the variability of structures and multimodal inconsistency problems is that the information recognition systems evolved to handle complex images are another argument [12]. Although earlier research has been carried out in developing sonar processing, image-based recognition, and neural modeling as discrete entities, there is little research done quantitatively to determine the role of combined signal and image processing on consistent spatial decision accuracy. Most of the available literature concentrates on classification performance but not regression-based precision of localization, and it has a gap in knowledge about the multimodal fusion impacts on coordinate estimations in complicated operational environments.

Although acoustic and visual processing techniques have become extremely developed, there are still a number of limitations. To begin with, unimodal strategies tend to be weak in scenarios: sonar techniques might be low-spatial-resolution, but visual systems can be destroyed by the environment. Second, little literature has assessed the quantifiable effect of multimodal feature-level combination on continuous localization accuracy in a variety of environmental conditions. Third, comparative studies tend to use averaged measures of performance with little analysis of variability carried out in a scenario. It is thus necessary that a research work that explicitly measures the extent to which the incorporation of both signal and image characteristics impacts the accuracy of decisions made in a complex underwater system be undertaken. Specifically, the contextual contingency of the multimodal advantages is not fully described.

This study examines the effect of signal and image processing applications on localization accuracy in underwater conditions. The main aim is to make a comparison of sonar-only, camera-only methods, and multimodal fusion methods when the experimental conditions are the same. The standard regression-based measures of error are used to measure localization performance, and in particular, the variation in performance between the various environmental sets. The paper also analyzes the role of feature-level multimodal integration in spatial estimation accuracy and gives an organized evaluation of the influence of integrated processing on decision making of complex systems.

2. Methodology

2.1 Experimental Framework

The study discusses the effect of signal and image processing, which compete with one another on the precision of diver localization on complex underwater environments. It was formulated as a localized and two-dimensional regression problem, and the acoustic and visual data were simultaneously used to predict the spatial coordinates of a diver. Three strategies of processing were identified to observe the contribution of each sensing modality and these three strategies were sonar-only signal processing, camera-only image processing and multimodal feature-level fusion. The models were trained and tested using the same conditions to ensure that performance differences make modality contribution and not architectural prejudice.

2.2 Dataset Description and Synchronization

The underwater multimodal data were publicly available, and the experiments were conducted on the basis of the publicly available data that is located on Zenodo [13]. Aligned right camera ZED frames, multibeam sonar images, and ground-truth diver positioning of a diver in real underwater conditions are found in the data set. The data are arranged in four sets of experiments in different environmental settings and motion patterns. A visual frame obtained with the help of a stereo camera system (a right channel in the current study), a sonar frame obtained at the same time, and the coordination of ground-truth diver locations in image pixel space are used as the constituents of each sample. Temporal alignment of camera and sonar modalities was accomplished using the mapping file provided with the dataset, and each frame of the camera and sonar was matched using the timestamps. Co-ordinates of ground-truth records were combined with the sonar time being taken as the key synchronization

value. The samples that had full cross-modal correspondence were only kept. The result of the sync and filtering was 1,193 aligned multimodal samples.

2.3 Preprocessing

The images were first converted to grayscale before the extraction of the features to normalize the representations of intensities, thus providing a low-level representation. In order to have similar dimensionality across modalities, each frame was resized to 64 × 64 pixels. This resolution retains the coarse spatial structure but at the expense of significantly dimensionality-reducing features. The intensity of the pixels was scaled to the interval [0,1] so that the model could be trained more numerically. No data augmentation strategies were used because the aim of this research was to separate the role of modality integration instead of opting to maximize predictive performance with the help of artificial transformations.

2.4 Feature Extraction

The extraction of features was devised in such a way that it would extract global intensity features as well as the coarse spatial structure of sonar and visual data. Global statistical descriptors were calculated on each frame of size 64 × 64, which consisted of mean intensity, standard deviation, minimum intensity, and maximum intensity. These characteristics are a summary of total reflectivity and variation in brightness. In order to apply spatial information, the individual frames were subdivided into a 4 ×4 grid of non-overlapping subregions. The average intensity of every sub-area was estimated and added to the feature vector. This process gave the 20-dimensional representation of each modality, which has 4 global statistics and 16 regional mean values. In the case of multimodal fusion, sonar and camera feature vectors were in turn concatenated to create a 40-dimensional representation. The feature-level fusion was applied to retain the feature modality-specific features with the ability to do joint regression learning.

2.5 Regression Model

Diver localization was formulated as a continuous regression model of predicting horizontal and vertical coordinates in image space. The reason why a Random Forest Regressor was used is that it is highly resistant to nonlinear association and that it can be utilized in moderately-sized datasets. This model was set to the default splitting criteria and 200 decision trees. Sonar, camera, and the jointly trained fusion feature vectors were then trained as separate regression models. The same hyperparameters were applied in all three setups so that the results can be compared directly.

2.6 Training and Evaluation Protocol

The data was divided into an 80/20 split for training and testing. The split indices used were the same in all three models, so that modalities are fairly compared. Localization accuracy was evaluated using Euclidean distance between predicted and ground-truth coordinates, defined as:

$$E = \sqrt{(x_{true} - x_{pred})^2 + (y_{true} - y_{pred})^2}$$

To measure the performance, the following methods were used: Root Mean Square Error (RMSE) was used to measure the performance, Mean Absolute Error (MAE), median error, standard deviation error, and maximum observed error. Besides general performance assessment, per-set RMSE was calculated to evaluate the strength in different environmental situations.

3. Results

3.1 Dataset Characteristics and Spatial Distribution

The final multimodal record was 1,193 synchronized samples within four sequences of the experiment (Table 1). The most numerous were the set_3 (425 frames), set_4 (374), set_1 (209) and set_2 (185).

Table 1. Dataset Composition per Experimental Set

Set	Paired Frames
set_1	209
set_2	185
set_3	425
set_4	374
TOTAL	1193

All ZED images have been captured at 1280 × 720 pixels, and all sonar frames at 932 × 514 pixels. There was no variation in dimension between samples (Table 2), which guaranteed that the feature extraction between modalities was similar.

Table 2. Image Dimension Statistics

Statistic	ZED	Sonar
Mean Width	1280	932
Mean Height	720	514
Std Width	0	0
Std Height	0	0
Min Width	1280	932
Min Height	720	514
Max Width	1280	932
Max Height	720	514

Table 3 gives ground-truth statistics of coordinates. The difference between Y means in sets implies that the depth distributions in the image plane are scenario-dependent.

Table 3. Target Coordinate Statistics per Set

Statistic	set_1	set_2	set_3	set_4
Mean X	414.74	462.65	456.78	400.72
Std X	48.57	69.80	50.77	45.67
Min X	354	343	335	340
Max X	518	542	525	505
Mean Y	281.11	286.79	228.27	323.13
Std Y	49.78	47.55	32.42	36.03
Min Y	203	211	187	214
Max Y	347	354	316	367

Representative multimodal samples illustrating temporal alignment between ZED and sonar frames are shown in Figure 1.

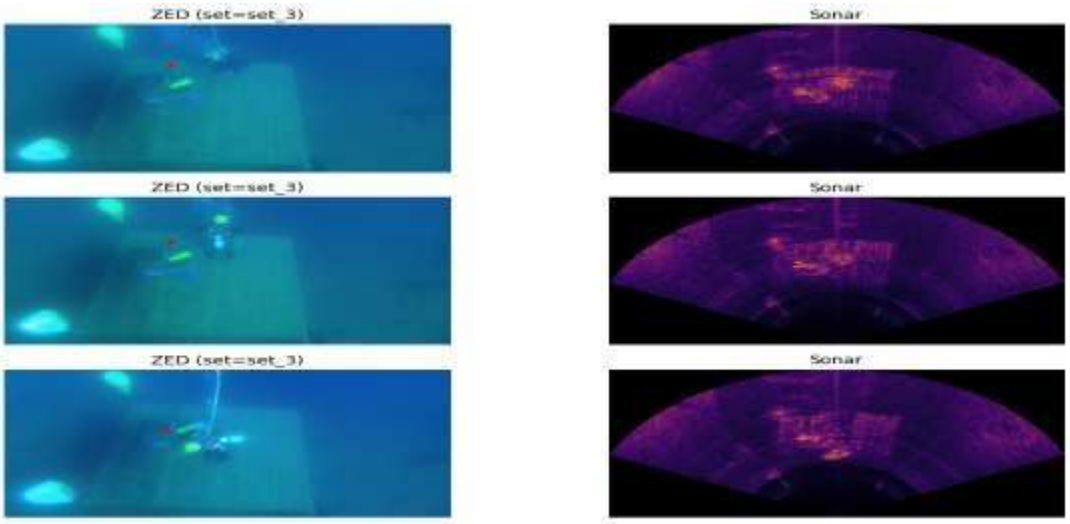


Figure 1. Representative Paired Multimodal Samples

Figure 2 shows the spatial distribution of the positions of divers worldwide. The dissimilar clustering proves the diversity of trajectories and spatial complexity to be evaluated in different environmental conditions.

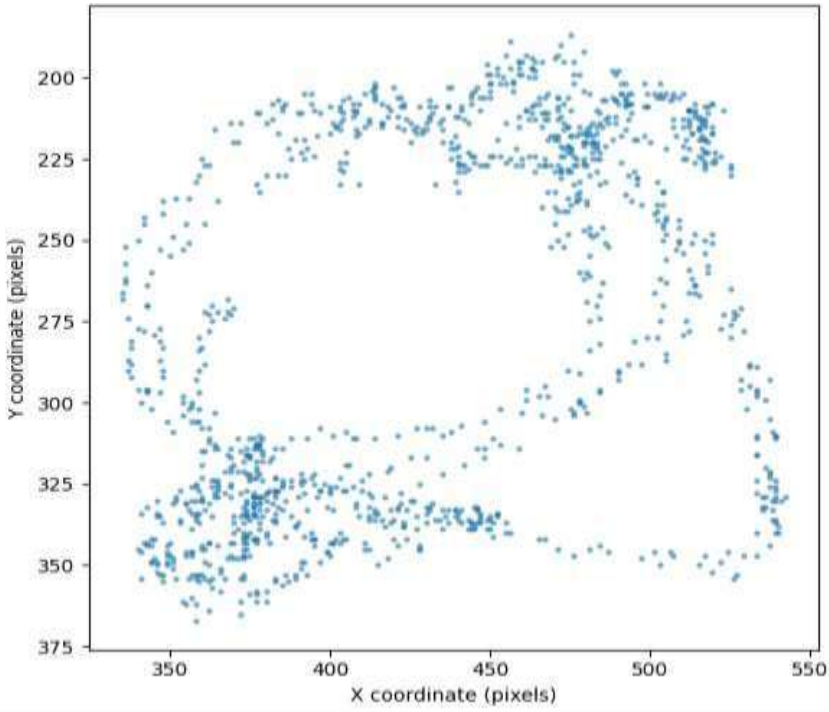


Figure 2. Spatial Distribution of Diver Positions Across All Experimental Sets

3.2 Overall Performance Comparison

Table 4 provides the performance metrics of unimodal and multimodal models. The highest localization error (RMSE = 18.79 pixels) was achieved in the sonar-only model. processing with cameras alone brought the RMSE to 13.31 pixels. The lowest RMSE of 12.84 pixels was obtained on the fusion model, which is equivalent to a reduction of 31.7 and 3.5% as compared to sonar-only processing and camera-only processing, respectively.

Table 4. Extended Performance Metrics (Euclidean Error)

Model	RMSE	MAE	Median Error	Std Error	Max Error
Sonar	18.79	14.81	12.44	11.57	81.76
Camera	13.31	9.13	6.38	9.68	72.97
Fusion	12.84	8.63	6.38	9.50	88.72

The error distributions for all three models are shown in Figure 3. Fusion shifts the distribution toward lower Euclidean errors while maintaining comparable dispersion.

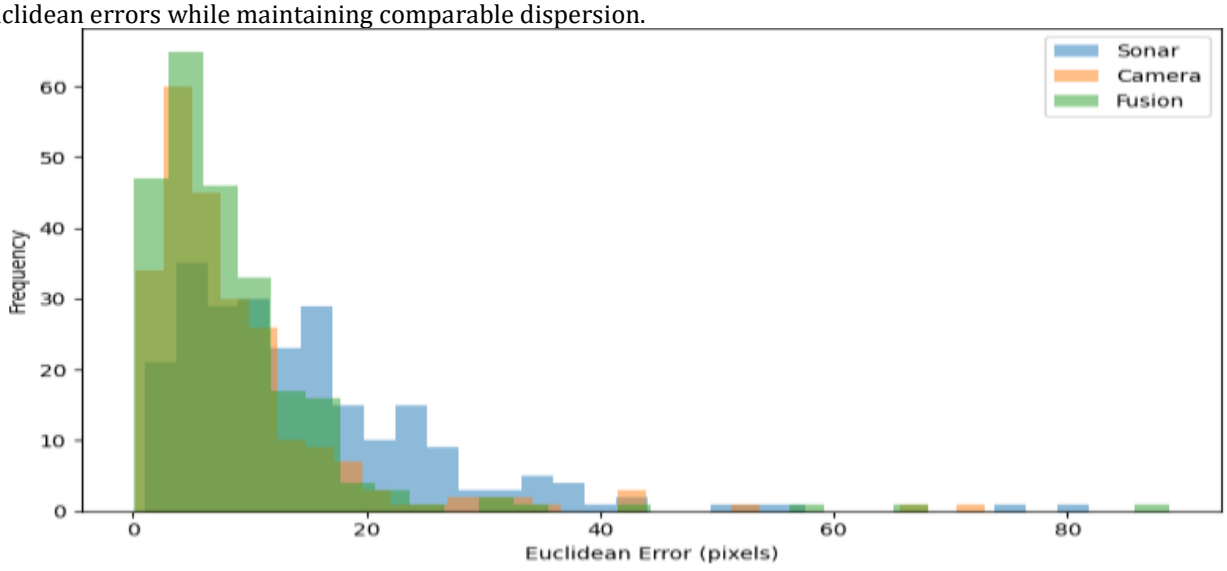


Figure 3. Euclidean Error Distribution for Sonar, Camera, and Fusion Models

3.3 Per-Set Robustness Analysis

Per-set performance results are summarized in Table 5 and illustrated in Figure 4. Fusion demonstrated substantial improvements in set_3 and set_4, where error reductions exceeded 40% relative to sonar-only processing. In set_2, moderate improvements were observed. In set_1, fusion improved over sonar but underperformed relative to camera-only processing, suggesting that sonar contributions may be less beneficial under visually favorable conditions. These trends are visually confirmed in Fig. 4.

Table 5. Per-Set RMSE and Improvement Percentages

Set	Sonar	Camera	Fusion	Fusion vs Sonar (%)	Fusion vs Camera (%)
set_1	24.93	16.95	19.99	19.79%	-17.95%
set_2	16.82	15.79	14.35	14.69%	9.12%
set_3	20.23	12.22	10.98	45.72%	10.11%
set_4	13.56	10.27	7.86	42.00%	23.47%

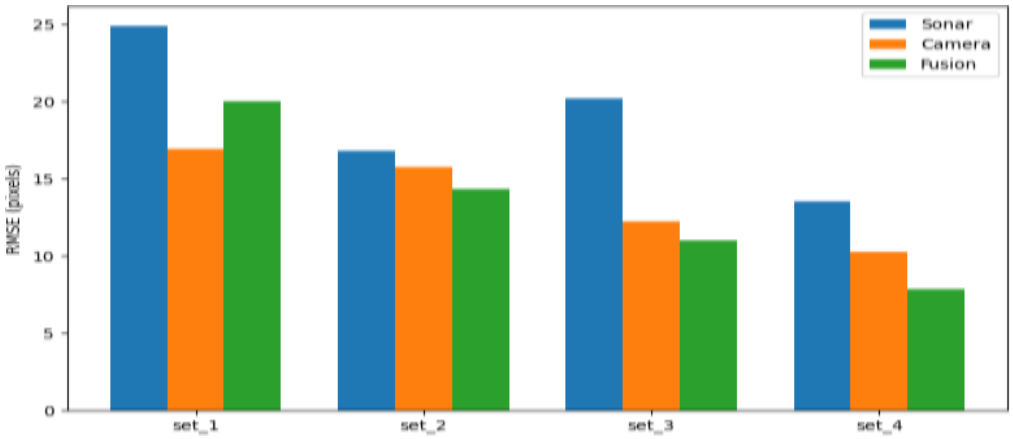


Figure 4. Per-Set RMSE Comparison for Sonar, Camera, and Fusion Approaches

3.4 Qualitative Evaluation

The comparison of ground-truth positions and model predictions is presented in Figure 5(A-C) using qualitative comparisons. Predictions of fusions are more likely to be similar to the ground-truth coordinates than sonar-only predictions and often tend to be more stable than camera-only predictions, especially when working with sets that have a wider range of spatial variability.

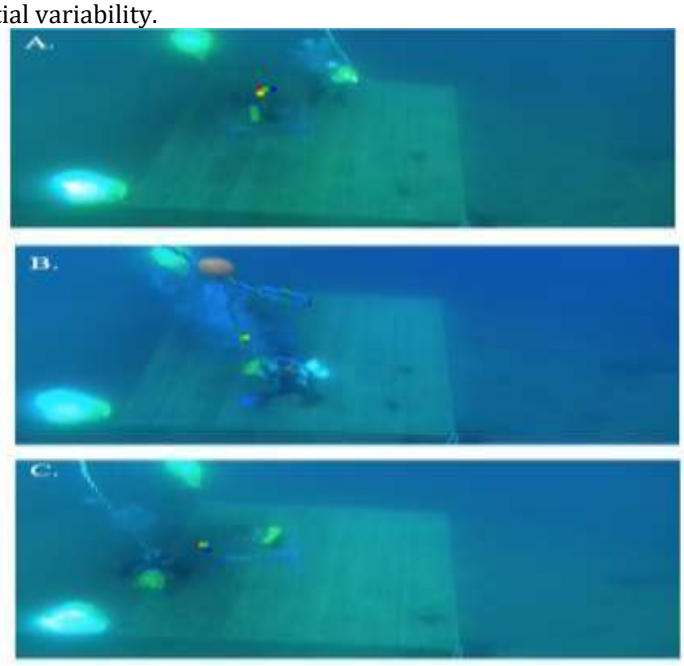


Figure 5(A-C). Qualitative comparison of diver localization predictions under different experimental conditions. Red denotes ground truth; blue denotes sonar prediction; yellow denotes camera prediction; green denotes fusion prediction.

4. Discussion

These findings indicate that multimodal integration enhances the total localization accuracy in comparison with sonar-only processing and, in most instances, with camera-only processing. The decrease in the RMSE of the fusion model shows that acoustic and visual features complement each other in terms of space. Although camera-based estimation showed good performance in conditions with good visual conditions, integration of sonar-obtained features enhanced stability in conditions with an increased level of spatial variability. This interpretation is also supported by the error distribution analysis. This was the case with fusion, where on the one hand, there was a decrease in the average localization error but at the cost of a slight increase in dispersion, indicating that the gains are due to the better typical-case estimation and not the reduction in extreme outliers. The increase in performance did not, however, occur in all the experimental sets. In one instance, fusion failed to do well as compared to camera-only processing, which means that multimodal integration is context-specific. In case visual information is good enough, extra acoustic characteristics can add marginal variance as opposed to enhancement. These results imply that integrated processing leads to increased resilience in complicated environments, as well as the fact that it does not imply a universal advantage over robust unimodal baselines.

The evidence of multimodal fusion improvements is in line with the larger trends in more complex analysis methods of signal and image processing, with integrated representations being more frequently used to improve the reliability of systems in uncertain situations [14]. Signal-domain and image-domain processing used in combination with each other in underwater acoustic applications has been demonstrated to enhance detection performance, especially in cases where environmental noise impacts on individual modalities [15]. As it is observed in recent reviews of sonar imagery analysis, acoustic processing performance is highly influenced by feature representation and the environmental structure [16]. This understanding is consistent with the relatively larger error in sonar-only processing, especially in the case of simple intensity-based descriptors. Meanwhile, new hypercomplex signal processing frameworks are suggesting the use of single mathematical models to describe heterogeneous inputs to the senses, supporting the theoretical justification of using multimodal integration [17].

In non-underwater systems, hybrid systems learning that consists of several streams of signals has been shown to have improved predictive capability. As a case in point, integrated convolutional and sequential models have been used to detect faults in dynamic systems in a better way [18], and pattern recognition frameworks of an ensemble nature have been used to make classification more robust by multi-source aggregation [19]. These results correlate with the current research findings in which the use of both acoustic and visual features decreased the localization error as compared to unimodal strategies. It is also true that the localization flow in this paper, through the application of vision is similar to those implemented in the current computer vision applications where image-based learning provides robust estimates of space under stable conditions [20]. The most recent multimodal systems, one of which is the integration of visual information with other sensory information, such as visual-tactile fusion in autonomous navigation, have demonstrated that in case one of the senses is already invalid, the other modalities complement them [21]. Similarly, the relevance of contextual soundness in scenario application in natural and variable situations is highlighted by detailed reviews of detection methods which rely on image-processing [22]. This general principle is reflected in the gains that are scenario-dependent in this case.

Underwater perception systems have practical implications for the findings. To begin with, Multimodal integration seems to be the most effective in the conditions of diminished visual acuity or increased structural complexity. In these instances, the sonar-acquired information serves as the complementary information that normalizes a spatial estimate. Second, the difference in the effectiveness of fusion between sets indicates that the adaptive integration strategies could improve reliability even further. Instead of using constant fusion weights systems, in the future can be more useful to use dynamically adapted modality contributions depending on the signal as estimated. On a more general level, these findings substantiate the opinion that when complex sensing environments demand such processing architectures, then they need to be able to take advantage of modality complementarity instead of depending on a dominant sensor stream.

There are a number of limitations that should be noted. The representations of features were based on crude statistical features as opposed to rich and task-oriented embeddings. Although this design can make interpretability and controlled comparison possible, features designed to improve the modality can provide more significant enhancements. Another substitution that was put into place was the fusion by direct feature concatenation. More elaborate mechanisms of integration could be more balanced with the modality influence, and minimize the possible interference. The adaptive multimodal frameworks and performance in controlled degradation conditions should be investigated in future research. More expansive validation in other environmental contexts would further explain the context of limitation upon which integrated signal and image processing is capable of improving accurate decisions consistently.

5. Conclusion

Multibeam sonar and monocular imagery processing enhanced the accuracy of localization of divers in a variety of underwater conditions by selectively using the complementary properties of sensing, where fusion produced the

smallest total error, even though it was still affected by modality-dependent scenario requirements. On 1,193 synchronized samples, camera-only estimation significantly outperformed sonar-only estimation, and feature-level fusion resulted in addition of further reduction in aggregate error, showing that acoustic features can provide valuable structural information when visual information is more unstable. The per-set analysis, however, indicated that fusion does not dominate the strongest unimodal baseline uniformly; when the sonar features are added to be more visually appealing, it creates a variance that slightly decreases performance. The results indicate the importance of multimodal integration to provide strength, but not necessarily better performance in all operating conditions. Future directions include focusing efforts on adaptive fusion techniques that rank modalities based on their quality estimations, adding more information into modalities through richer modalities, and performance evaluation through the use of cross-validation and statistical significance tests. Applying the evaluation to other underwater scenes and controlled experimentation of degradation would further explain the limits of operation at which integrated signal and image processing provide the maximum enhancement of the decision accuracy.

References

1. Bhattacharyya, S., Mukherjee, A., Bhaumik, H., Das, S. and Yoshida, K., 2021. *Recent trends in signal and image processing*. Springer Singapore.
2. Monga, V., Li, Y. and Eldar, Y.C., 2021. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Processing Magazine*, *38*(2), pp.18-44.
3. Aicha, M., 2024, April. Techniques and applications of image and signal processing: a theoretical approach. In *2024 8th International Conference on Image and Signal Processing and their Applications (ISPA)* (pp. 1-8). IEEE.
4. Zhang XY, Liu CL, Suen CY. Towards robust pattern recognition: A review. Proceedings of the IEEE. 2020 May 28;108(6):894-922.
5. Nixon, M. and Aguado, A.S., 2025. *Feature extraction and image processing for computer vision*. Academic press.
6. Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Umar, A.M., Linus, O.U., Arshad, H., Kazaure, A.A., Gana, U. and Kiru, M.U., 2019. Comprehensive review of artificial neural network applications to pattern recognition. *IEEE access*, *7*, pp.158820-158846.
7. Beyerer, J., Hagmanns, R. and Stadler, D., 2024. Pattern recognition: introduction, features, classifiers and principles. In *Pattern Recognition*. De Gruyter Oldenbourg.
8. Anitha, U., Malarkkan, S., Jebaselvi, G.A. and Narmadha, R., 2019. Sonar image segmentation and quality assessment using prominent image processing techniques. *Applied Acoustics*, *148*, pp.300-307.
9. Sun, C., Wang, L., Wang, N. and Jin, S., 2021. Image recognition technology in texture identification of marine sediment sonar image. *Complexity*, *2021*(1), p.6646187.
10. Singh, H., 2019. *Practical machine learning and image processing: for facial recognition, object detection, and pattern recognition using Python*. Apress.
11. Zohuri, B. and Rahmani, F.M., 2024. Computational Intelligence for Signal and Image Processing Unleashing the Power of Algorithms. *Journal of Clinical Case Reports and Studies*, *5*(2), pp.1-10.
12. Abdullayeva, G.G. and Alizadeh, U.M., 2020. An information recognition system for complex images. *Adcaij-advances in distributed computing and artificial intelligence journal*, pp.79-93.
13. Rodrí´guez, L., Ridao, P., Ribas, D., Carreras, M. and Mallios, A., 2023. Fusion of Underwater Camera and Multibeam Sonar for Diver Detection and Tracking (Version 1.0) *Dataset*. Zenodo. https://zenodo.org/records/10220989
14. Basheer, M., Ramya, S.K., Durai Ganesh, A., Halambi, B. and Geethanjali, L.S., COMPLEX ANALYSIS METHODS FOR IMAGE AND SIGNAL PROCESSING.
15. Qin, Z., Li, C., Wang, H., Nie, L. and Yin, H., 2025. Active underwater acoustic signal detection using image pattern recognition techniques. *Ocean Engineering*, *341*, p.122417.
16. Chai, Y., Yu, H., Xu, L., Li, D. and Chen, Y., 2023. Deep learning algorithms for sonar imagery analysis and its application in aquaculture: A review. *IEEE Sensors Journal*, *23*(23), pp.28549-28563.
17. Yu, Z., Li, D., Du, P., Luo, W., Kou, K.I., Bhatti, U.A., Benger, W., Lv, G. and Yuan, L., 2024. Hypercomplex signal processing in digital twin of the ocean: theory and application [Hypercomplex Signal and Image Processing]. *IEEE Signal Processing Magazine*, *41*(3), pp.33-48.
18. Huang, T., Zhang, Q., Tang, X., Zhao, S. and Lu, X., 2022. A novel fault diagnosis method based on CNN and LSTM and its application in fault diagnosis for complex systems. *Artificial Intelligence Review*, *55*(2), pp.1289-1315.
19. Nagasubramanian, G., Sakthivel, R.K., Patan, R., Sankayya, M., Daneshmand, M. and Gandomi, A.H., 2021. Ensemble classification and IoT-based pattern recognition for crop disease monitoring system. *IEEE Internet of Things Journal*, *8*(16), pp.12847-12854.
20. Chen, B., Chen, C., Hu, J., Sayeed, Z., Qi, J., Darwiche, H.F., Little, B.E., Lou, S., Darwish, M., Foote, C. and Palacio-Lascano, C., 2022. Computer vision and machine learning-based gait pattern recognition for flat fall prediction. *Sensors*, *22*(20), p.7960.

21. Shi, R., Yang, S., Chen, Y., Wang, R., Zhang, M., Lu, J. and Cao, Y., 2023. CNN-Transformer for visual-tactile fusion applied in road recognition of autonomous vehicles. *Pattern Recognition Letters*, *166*, pp.200-208.
22. Nagar, H. and Sharma, R.S., 2020, May. A comprehensive survey on pest detection techniques using image processing. In *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 43-48). IEEE.