Multi-sensor data capture and fusion for mapping hazardous materials onboard ships

Richard J. D. Moore[1*], Peter Ø. Ivarsen[1], and Ahmed Mohammed[1]

Abstract — **Ship decommissioning requires accurate, spatially grounded knowledge of hazardous materials within complex vessel interiors, yet this information is often fragmented across heterogeneous sources and disconnected from the ship's current physical state. We present a multi-sensor data capture and fusion pipeline that combines mobile 3D scanning, gas sensing, and vision-language perception to generate enriched, semantically annotated 3D ship models. The system integrates lidar–inertial SLAM, RGB imagery, and combustible gas sensing, along with post-processing methods for point-cloud denoising and visibility-aware multi-view fusion. Using the museum ship DS *Hestmanden* as a case study, we demonstrate spatial localisation of gas measurements and 3D semantic mapping of ship-like environments. Our results constitute a first step toward linking unstructured hazard data with spatial models, enabling hazard localisation and safer, more informed ship decommissioning.**

I. Introduction

The safe and efficient decommissioning of ships requires detailed knowledge of both vessel structure and the location of embedded hazardous materials. In practice, this information is fragmented across heterogeneous sources, including textual documentation, tables, diagrams, and images within the Inventory of Hazardous Materials (IHM), as well as 3D scan data capturing the ship's current physical state. The goal of this work is to automatically integrate documentary information with 3D scan data to produce enriched, semantically meaningful 3D models that support downstream tasks such as the automated generation of safer and more informed cutting plans during decommissioning. This research is conducted within the SHEREC Horizon Europe project, which aims to advance AI technologies for safer and more sustainable ship recycling.

Our broader ambition is to develop scalable methods for linking unstructured and semi-structured documents to spatial representations derived from mobile 3D sensing. In this paper, we present initial steps toward this goal, including the design and preliminary evaluation of a mobile sensor system for mapping hazardous environments; the creation of a ship decommissioning dataset based on the DS *Hestmanden*, comprising scan data and documentation including the IHM; post-capture processing techniques to improve the usability of mobile scan data; and early experiments with open-vocabulary segmentation onboard ships. While motivated by ship recycling, the proposed system has broader applicability, including the enrichment of digital building or infrastructure models and the alignment of legacy databases with contemporary sensor data.

Figure 1. (L-R) Our sensor rig and DS Hestmanden.

II. System Design & Methods

A. Sensor rig design

Our sensor rig (Fig. 1) is designed for precise 3D semantic mapping of complex environments. It is compact, lightweight, and modular, allowing hand-carried or drone-mounted operation, and exposes raw sensor data for downstream processing. A key SHEREC goal is to automate ship scanning, therefore, the rig must also provide real-time localisation and guidance to a semi-autonomous drone platform in cluttered, low-visibility environments.

The rig must support detection of hazardous materials, including gases listed in the IHM, as well as previously unrecorded hazards. Relevant gases include flammable species that pose explosion risks during robotic cutting operations and short-term toxic gases such as hydrogen, oxygen deficiency, carbon monoxide, and hydrogen sulphide, based on occupational exposure limits [1].

Based on these requirements, the sensor rig comprises:

- **Livox MID-360 lidar + IMU** [2]: wide-angle 3D sensing with >10 m range for robust localisation.
- **RealSense D455 stereo camera** [3]: RGB and depth data for semantic perception.
- **DVXplorer event camera** [4]: high-dynamic-range sensing for guidance in poor visibility and rapid relative motion estimation.
- **NevadaNano MPS Flammable Gas Sensor 4** [5]: detection of gases hazardous to robotic operations; USB interface. Four-gas monitors such as RKI Instruments GX-3R [6] detect also toxic gases but are designed for handheld use.

[1] SINTEF Digital, Oslo, Norway.

[*] Correspondence: richard.moore@sintef.no

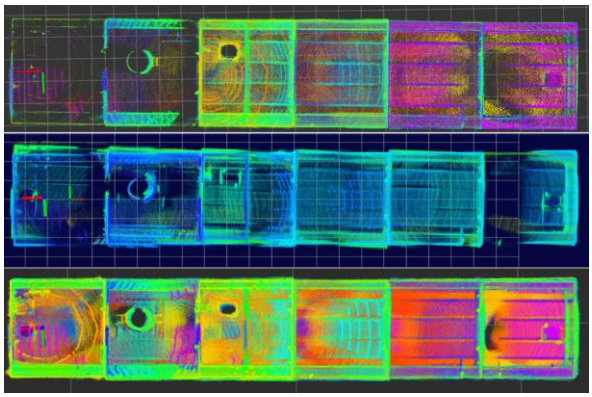

Figure 2. Top-down views of a reconstructed ballast tank showing a comparison of three lidar-inertial SLAM algorithms (top-bottom): GLIM, Fast-LIO2, and COIN-LIO. The ballast tank comprises several compartments separated by narrow openings.

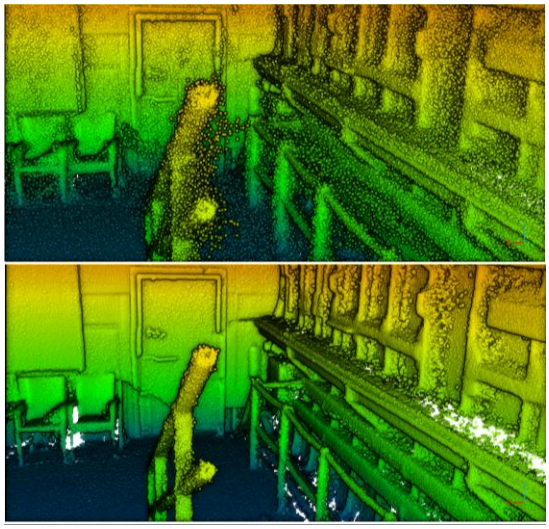

Figure 3. Sections from the 3D reconstruction of DS Hestmanden showing our steps for point cloud densification (top - bottom): GLIM SLAM 3D map following densification and following denoising using density gradients and SOR filter.

All sensors are interfaced via ROS2, with data synchronised in software. Lidar frames are captured at 10 Hz and RGB images at 30 Hz. Event data is recorded for all runs, while depth data is available for a subset. Hardware synchronisation is not yet implemented; data association uses nearest-timestamp matching to each lidar frame. The gas sensor outputs environmental variables and gas concentrations, including hydrogen, methane, and light, medium, and heavy hydrocarbons. Measurements are recorded at 1 Hz with a response time (T90) below 20 s. All data is logged to a laptop for post-processing. Camera intrinsics are taken from the RealSense ROS node, and lidar–camera extrinsics are computed following [7].

B. Dataset capture (DS Hestmanden)

We aim to combine complete ship scan data with heterogeneous data from documentation. This requires physical access to all onboard areas and to sensitive ship records, which are not available in open datasets. To create our own dataset, we selected the museum ship DS Hestmanden (Fig. 1), to which Vestagder Museet has provided access.

The dataset includes scan data acquired using our hand-carried sensor rig, original architectural drawings dating back to 1911, and the ship's Inventory of Hazardous Materials (IHM), updated in 2025. The IHM contains textual descriptions, tables, figures, and images documenting hazardous materials onboard. The scan data cover a meandering 775 m trajectory through the vessel and comprise 19,380 poses, each associated with an RGB image and lidar frame, as well as 390,520 IMU frames and 195,233 event-camera messages. We plan to release a more comprehensive version of this dataset in the near future[2].

C. 3D localisation and mapping (SLAM)

To identify a baseline algorithm for downstream spatial tasks that is compatible with the conditions expected in ship-like environments, we tested several state-of-the-art lidar-inertial SLAM and odometry algorithms including GLIM [8], Fast-LIO2 [9], and COIN-LIO [10]. These approaches were evaluated using the lidar-visual-inertial ballast water tank dataset [11] due to the unavailability of ship data at the time of comparison. From a qualitative comparison of the 3D reconstructions (Fig. 2), we found that GLIM performed robustly, producing more accurate 3D maps in ship-like environments. In particular, GLIM was better able to retain a coherent reference frame through areas with poor scan overlap – i.e. when passing through narrow doors or manholes – due to the use of a GICP factor [12] that enables optimisation of poses with poor overlap using scan data from both before and after that point in the factor graph.

D. Post-capture processing

We use the GLIM framework to compute sensor rig trajectories and to produce an initial sparse point cloud map. We then perform the following post-processing steps.

Point-cloud densification and denoising can substantially improve the robustness of classification networks when operating on sparse or noisy data [13] – typical of mobile 3D lidar + SLAM systems, as opposed to terrestrial scanning. We therefore apply two post-processing steps to improve the quality of the reconstructed point clouds.

First, all individual lidar scans are incorporated into the SLAM map using high-accuracy deskewing based on known relative poses and IMU data, followed by GICP registration of each scan to the SLAM map.

Second, we apply a scalable denoising method suitable for point clouds comprising hundreds of millions of points. Both learning- [14] and normal-based [15] approaches were evaluated but were found to introduce artifacts such as point clumping and over-smoothing, and to be computationally expensive at scale. We therefore propose a lightweight denoising method based on the assumption that lidar depth noise follows a Gaussian-like distribution around true surfaces. Surface noise is reduced by estimating density gradients and iteratively nudging points along these gradients (Fig. 3).

[2] https://github.com/SINTEF/SHEREC---Ship-Decommissioning-Dataset

Our algorithm proceeds as follows:

1. Voxelise the input point cloud using a grid size of 0.5 m.
2. Sub-voxelise each outer voxel with an inner grid size of 1.0 cm and record point counts per inner voxel.
3. Normalise point densities within each outer voxel to account for large density variations, then smooth with a 3D Gaussian kernel (window size 13, $\sigma = 2$) and compute 3D density gradients.
4. Iteratively nudge points along density gradients, using a step size inversely proportional to local density, until they reach a dense region (threshold = 0.4 pts/cm³).

This approach is computationally efficient, as density gradient fields are easily vectorised and computed only once. Compared to normal-based methods, it is also less sensitive to local noise. Finally, a statistical outlier removal (SOR) filter is applied to eliminate isolated points far from surfaces (Fig. 3).

Visible point filtering to recover full resolution visible point clouds for each pose. Unlike terrestrial lidar scanning, which produces dense point clouds per pose, continuous mobile scanning yields only single lidar frames, resulting in low spatial resolution. Although we densify the accumulated point cloud as described above, point visibility from individual poses becomes unknown, complicating projection of 2D image information onto 3D points. We observe that standard hidden point removal methods [16] discard excessive points in noisy mobile scans and may fail to eliminate unobservable points behind sparse regions. We therefore propose a novel method for recovering full-resolution visible point clouds in mobile scanning datasets.

Our algorithm operates as follows:

1. Use a single lidar frame to define the **visible volume** for the current pose.
2. Construct a watertight mesh from the current frame, by:
    a) Compute unit direction vectors for all points.
    b) Compute a convex hull over these directions (a mesh on the unit sphere) to efficiently obtain a watertight surface.
    c) Transform hull vertices back to the local frame and expand the mesh slightly (2.5 cm) to account for noise.
3. Identify points from the accumulated global point cloud that fall inside this mesh volume. An oriented bounding box is first used to prune candidates, followed by ray casting to test inlier points.

Our key novelty is deriving the visible volume solely from points in the current lidar frame, which are by definition visible (Fig. 4). This allows us to retain all accumulated points within the observed volume, achieving much higher density than [16], which keeps only points on the estimated visible surface. A limitation is that visibility can only be computed for poses with lidar frames. Fig. 5 compares our method to [14] (Open3D implementation, radius = 1e4) on a $20 \times 20 \times 20$ m region around pose #5785 in the Hestmanden dataset: from 74.8 M points, we retain 19 M visible points, while [16] retains 4.1 M despite assuming spherical field of view (we are restricted to lidar FoV). Runtime was 14.78 s for our method versus 100.7 s for [16] on a ~3 GHz CPU.

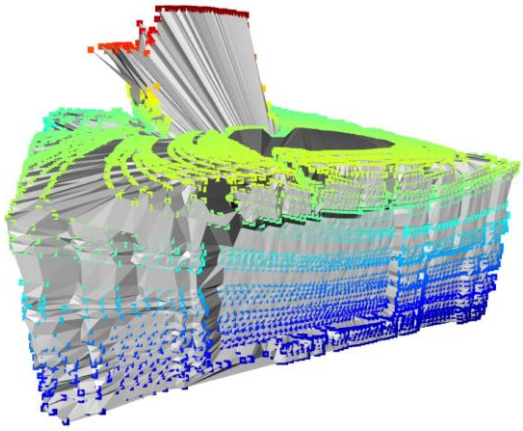

Figure 4. Points from a single lidar frame coloured by height and overlaid with our computed "visible volume" mesh.

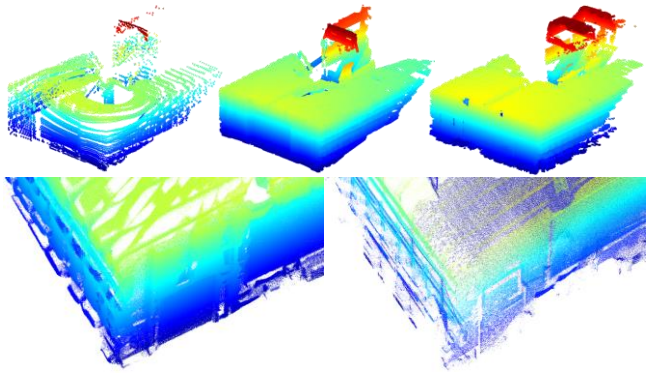

Figure 5. Top row (L-R): lidar points from a single frame only; all visible points computed by our approach; and all visible points computed by the standard method [16]. Bottom row (L-R): detail in our visible point cloud compared to [16] (identical point size).

Figure 6. Details from two locations in the DS Hestmanden reconstruction showing comparison between three methods for multi-view fusion: (top row) simple averaging, (middle row) median colour, (bottom row) weighted averaging. Weighted averaging gives best contrast and detail.

E. *Multi-view fusion*

Our long-term goal is to build a language-aligned 3D foundation model that preserves spatial relationships. As an initial baseline, we instead leverage state-of-the-art 2D open-vocabulary image segmentation methods (e.g. SAM3 [17]) and project their outputs into the 3D point cloud.

We use back-projection to associate 2D and 3D data but must solve the multi-view fusion problem. While implicit neural representations such as NeRF [18] achieve state-of-the-art multi-view fusion for low-dimensional attributes like colour, they scale poorly to large point clouds or higher-dimensional representations such as vision–language features. We therefore evaluate three simple multi-view fusion strategies: unweighted averaging; component-wise median filtering, which suppresses outliers; and weighted averaging, which prioritises observations from images expected to yield higher-quality measurements.

We compute our weights for the weighted average as a linear combination of two terms (distance and orientation):

$$w_p^{dist} = \left(1.0 - \frac{|d_p - 2.5|}{5.0}\right) \cdot 0.9 + 0.1, \text{ clamped to } [0,1] \text{ and}$$

$$w_p^{ang} = \left|\hat{\boldsymbol{n}}_p \cdot \hat{\boldsymbol{z}}_{\boldsymbol{cam}}^{\boldsymbol{world}}\right|,$$

where d_p is the distance from point p to the camera, $\hat{\boldsymbol{n}}_p$ is the estimated surface normal, and $\hat{\boldsymbol{z}}_{\boldsymbol{cam}}^{\boldsymbol{world}}$ is the camera's central view direction expressed in the world frame. The distance term favours observations from around 2.5 m, while the orientation term prefers near-orthogonal views over oblique ones. We find that weighted averaging provides the best contrast and detail (Fig. 6). Although illustrated with point colour, we consider this a reasonable proxy for other attributes such as visual-language features.

III. Preliminary Results

To evaluate combustible gas detection and mapping, we conducted a small-scale test in an office-like environment with a single known gas source (lighter fluid). For poses where the gas sensor detected gas, we proportionally increased the weights of surrounding voxels in a gas-concentration heatmap (Figure 9). The results show that we can detect and localise combustible gas sources to approximately room-level accuracy. Future work will focus on improved plume characterisation and more precise source localisation.

To assess the suitability of 2D open-vocabulary segmentation for semantic labelling of ship-scale 3D data, we applied SAM3 to all images in the DS *Hestmanden* dataset, segmenting them into 22 predefined classes (Fig. 8). Where masks overlapped, we preferred the higher-confidence label. Labels were projected into the point cloud using the above method, with one modification: since labels are discrete, we applied a weighted mode rather than a weighted average for multi-view fusion, using the same weighting scheme as before.

The results (Fig. 8) show that this approach yields usable semantic labels even in specialised environments such as ships. However, the need to predefine object classes is limiting. As future work, we aim to embed visual-language features directly into the point cloud to enable open-vocabulary querying, and to preserve meaningful 3D spatial relationships to support queries with spatial constraints.

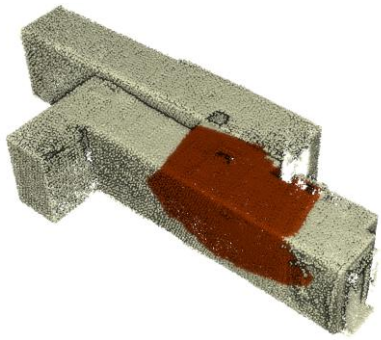

Figure 7. 3D heatmap of a corridor environment created with our sensor rig. Point colour reflects flammable gas concentration (grey = no gas, dark – light red = increasing concentration), showing localisation of the source.

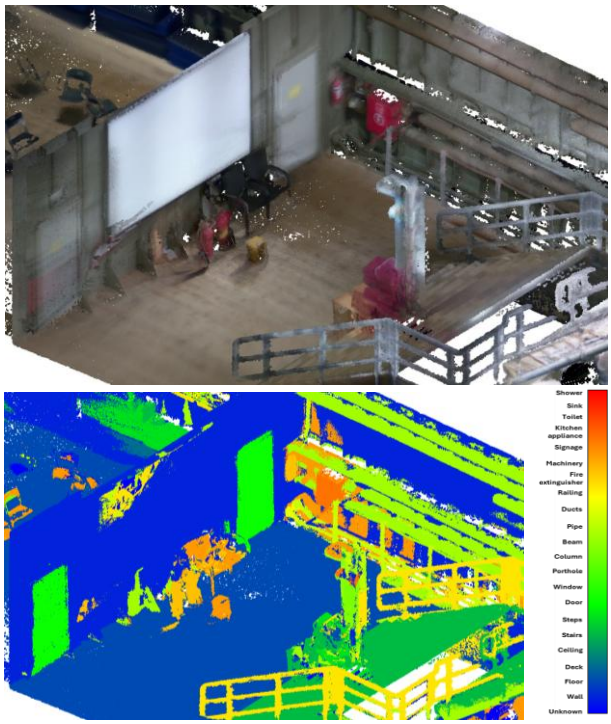

Figure 8. Detail sections from our 3D reconstruction of DS Hestmanden, showing (top – bottom) true colour and semantic segmentation labels from SAM3.

IV. Conclusion

We have presented a system design and methodology for creating a 3D semantic map of the ship DS *Hestmanden*. The approach introduces novel components, including new algorithms for point-cloud densification and visible-point filtering. We plan to release a more comprehensive version of the ship decommissioning dataset in the near future. Ongoing work will focus on fusing semantic information from 3D reconstructions with unstructured IHM data to enable automatic localisation of hazardous objects and materials onboard ships.

Acknowledgment

Sincere thanks to Vestagder Museet for providing access to DS Hestmanden and making the ship documents available.

References

[1] "International Chemical Safety Cards." Accessed: Apr. 17, 2026. [Online]. Available: https://chemicalsafety.ilo.org/dyn/icsc/showcard.home

[2] "MID-360", Livox. Accessed Apr. 21, 2026. [Online]. Available: https://www.livoxtech.com/mid-360

[3] "Real-Sense Depth Camera D455", Intel. Accessed Apr. 21, 2026. [Online]. Available: https://www.intel.com/content/www/us/en/products/sku/205847/intel-realsense-depth-camera-d455/specifications.html

[4] "DVXplorer Event Camera", Inivation. Accessed Apr. 21, 2026. [Online]. Available: https://docs.inivation.com/hardware/current-products/dvxplorer.html

[5] "MPS 4.0 Flammable Gas Sensor," NevadaNano. Accessed: Apr. 17, 2026. [Online]. Available: https://nevadanano.com/products/mps-flammable-gas-sensor/mps-flammable-gas-sensor-4/

[6] "GX-3R Four Gas Monitor, confined space detector for LEL, O2, H2S & CO. World's smallest and lightest portable Multi-Gas Detector. RKI's high quality safety instrument.," RKI Instruments. Accessed: Apr. 17, 2026. [Online]. Available: https://www.rkiinstruments.com/product/gx-3r-gas-detector/

[7] K. Koide, S. Oishi, M. Yokozuka, and A. Banno, "General, Single-shot, Target-less, and Automatic LiDAR-Camera Extrinsic Calibration Toolbox," in 2023 IEEE International Conference on Robotics and Automation (ICRA), London, United Kingdom: IEEE, May 2023, pp. 11301–11307. doi: 10.1109/ICRA48891.2023.10160691.

[8] K. Koide, M. Yokozuka, S. Oishi, and A. Banno, "GLIM: 3D Range-Inertial Localization and Mapping with GPU-Accelerated Scan Matching Factors," Robot. Auton. Syst., vol. 179, p. 104750, Sep. 2024, doi: 10.1016/j.robot.2024.104750.

[9] W. Xu, Y. Cai, D. He, J. Lin, and F. Zhang, "FAST-LIO2: Fast Direct LiDAR-Inertial Odometry," IEEE Trans. Robot., vol. 38, no. 4, pp. 2053–2073, Aug. 2022, doi: 10.1109/TRO.2022.3141876.

[10] P. Pfreundschuh, H. Oleynikova, C. Cadena, R. Siegwart, and O. Andersson, "COIN-LIO: Complementary Intensity-Augmented LiDAR Inertial Odometry," May 30, 2024, arXiv: arXiv:2310.01235. Accessed: Oct. 30, 2024. [Online]. Available: http://arxiv.org/abs/2310.01235

[11] M. Dharmadhikari et al., "Autonomous Exploration and General Visual Inspection of Ship Ballast Water Tanks using Aerial Robots," Nov. 07, 2023, arXiv: arXiv:2311.03838. Accessed: Oct. 25, 2024. [Online]. Available: http://arxiv.org/abs/2311.03838

[12] koide3, koide3/gtsam_points. (Apr. 14, 2026). C++. Accessed: Apr. 16, 2026. [Online]. Available: https://github.com/koide3/gtsam_points

[13] H. Zhang et al., "Deep learning-based 3D point cloud classification: A systematic survey and outlook," Displays, vol. 79, p. 102456, Sep. 2023, doi: 10.1016/j.displa.2023.102456.

[14] D. de S. Edirimuni, X. Lu, G. Li, L. Wei, A. Robles-Kelly, and H. Li, "StraightPCF: Straight Point Cloud Filtering," May 14, 2024, arXiv: arXiv:2405.08322. doi: 10.48550/arXiv.2405.08322.

[15] Z.-A. Feng and X.-F. Han, "Guided normal filter for 3D point clouds," Multimed. Tools Appl., vol. 82, no. 9, pp. 13797–13810, Apr. 2023, doi: 10.1007/s11042-022-13751-w.

[16] S. Katz, A. Tal, and R. Basri, Direct visibility of point sets, vol. 26. 2007. doi: 10.1145/1275808.1276407.

[17] "SAM 3: Segment Anything with Concepts | Research - AI at Meta." Accessed: Apr. 17, 2026. [Online]. Available: https://ai.meta.com/research/publications/sam-3-segment-anything-with-concepts/

[18] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis," Aug. 03, 2020, arXiv: arXiv:2003.08934. doi: 10.48550/arXiv.2003.08934.