# A Complete System for Automated 3D Semantic-Geometric Mapping of Corrosion in Industrial Environments

Rui Pimentel de Figueiredo[a], Stefan Nordborg Eriksen[b], Ignacio Rodriguez[c], Simon Bøgh[a]

[a]*Department of Materials and Production, Aalborg University, Aalborg Øst, 9220, Denmark*
[b]*Department of Electronic Systems, Aalborg University, Aalborg Øst, 9220, Denmark*
[c]*Department of Electrical Engineering, University of Oviedo, Gijon, 33203, Spain*

**Abstract**

Corrosion, a naturally occurring process leading to the deterioration of metallic materials, demands diligent detection for quality control and the preservation of metal-based objects, especially within industrial contexts. Traditional techniques for corrosion identification, including ultrasonic testing, radio-graphic testing, and magnetic flux leakage, necessitate the deployment of expensive and bulky equipment on-site for effective data acquisition. An unexplored alternative involves employing lightweight, conventional camera systems, and state-of-the-art computer vision methods for its identification.

In this work, we propose a complete system for semi-automated corrosion identification and mapping in industrial environments. We leverage recent advances in LiDAR-based methods for localization and mapping, with vision-based semantic segmentation deep learning techniques, in order to build semantic-geometric maps of industrial environments. Unlike previous corrosion identification systems available in the literature, our designed multimodal system is low-cost, portable, semi-autonomous and allows collecting large datasets by untrained personnel.

A set of experiments in an indoor laboratory environment, demonstrate quantitatively the high accuracy of the employed LiDAR based 3D mapping and localization system, with less then $0.05m$ and $0.02m$ average absolute and relative pose errors. Also, our data-driven semantic segmentation model, achieves around 70% precision when trained with our pixel-wise manually

annotated dataset.

*Keywords:* Computer vision, Corrosion detection, Semantic segmentation, Simultaneous localization and mapping

## 1. Introduction

Corrosion is a major problem that degrades metallic surfaces, which are ubiquitous in man-made constructions. Its identification and mitigation have significant socio-economic impacts across various industries, contributing to the maintenance and longevity of infrastructures, such as bridges, pipelines, and buildings, to name a few, reducing the frequency of reconstruction projects and associated costs [1].

Early and accurate identification of corrosion allows for timely intervention and maintenance, preventing costly repairs and replacements, reduced downtime in industries such as oil and gas, aerospace, and manufacturing leading to increased productivity and financial savings [2].

Corrosion in metal-made infrastructures can lead to leaks and spills, causing environmental pollution. Timely identification helps prevent such incidents, preserving ecosystems and minimizing the environmental impact [3]. Also, corrosion-related failures can compromise the safety of structures and equipment. Identifying and addressing corrosion hazards enhances workplace safety and reduces the risk of accidents and injuries, and may lead to lower insurance costs, as the risk of failures and subsequent claims is reduced. Insurers may offer lower premiums for well-maintained and corrosion-free assets [1].

Among the leading-edge technologies for identifying it, magnetic flux leakage, ultrasonic testing, and remote visual inspection stand out. Although magnetic flux leakage provides highly accurate solutions, its deployment is expensive, challenging, and necessitates skilled and trained human operators for successful operation. However, the current process of inspecting offshore assets using visual inspection technologies and evaluating where to conduct maintenance and repairs is a highly manual and laborious task. First, inspectors are required to traverse the asset and record and track a large number of images of corroded elements, which are manually labeled in a coarse manner. Also, industrial assets can be multi-story, multi-platform metal structures, naturally resulting in the collection of large datasets, which may be unfeasible to transmit via satellite connections, for instance, in remote offshore

locations. Additionally, the nature of industrial assets, as well as various safety and security concerns surrounding them, add challenges to performing the inspection task.

After the data collection procedure, the collected datasets undergo manual review by experts, evaluating the severity of the corrosion, and thus which structures require maintenance, and for which maintenance can be postponed. Some structures are more critical than others to maintain, for the sake of asset operation and safety. As part of this review, most of the time is spent resolving where a given image was recorded, and identifying the specific location of the corroded structure. This process involves the development of maintenance work packages. Within these packages, structures identified as critically corroded are designated for maintenance, repair, or replacement. These tasks may extend over several months, whether carried out on-site or, in the case of mobile offshore platforms, at a dry dock.

To alleviate some of the challenges posed by the manual nature of the current process, we envision a data collection device with the following general properties:

- **Portability**: Inspectors should be able to easily carry and operate the data collection device in industrial assets during long periods.
- **Accuracy**: High precision imaging sensors for accurate 3D semantic mapping of corroded assets.
- **Autonomy**: Intelligent software that should be able to perform mapping, localization and semantic localization and categorization of industrial assets, from sensory data. However, the software may not strictly run online, on the data collection device, but instead offline in a more powerful computational device.

Taking into account the previous properties, we leverage current advances in vision-based semantic segmentation and simultaneous localization and mapping (SLAM) approaches, and the availability of low-cost consumer grade LiDAR and visual camera systems, to propose a complete geometric-semantic mapping and localization system. We combine highly accurate 3D point cloud data provided by LiDAR, with monocular color vision for robust and precise localization of corrosion spots in 3D.

The rest of this article is organized as follows. First, in Section 2, we review the state-of-the-art technologies for positioning and corrosion identification, with a focus on vision-based 3D localization and mapping and

semantic segmentation techniques. Second, in Section 3 we describe in detail the proposed system, including design and sensor choices, as well as algorithmic choices. Then, in Section 4 we validate our system and demonstrate its applicability by reporting the results from a set of experiments performed both in an indoor laboratory scenario and in an outdoor offshore environment. Finally, we draw our main conclusions and propose extensions for future work in Section 5.

## 2. Related work

In this section we overview the state-of-the-art methods utilized in our approach, including existing technologies for positioning, corrosion detection, vision-based sensor calibration, localization and mapping, and semantic segmentation techniques.

### *2.1. Positioning Systems*

Multiple positioning technologies exist and have been reported in the literature for multiple purposes. With a focus on offshore positioning cases, industrial positioning, emergency systems, and positioning systems for autonomous systems, the following was concluded.

Technologies such as Ultra-Wide Band (UWB) [4], Ultra-sound [5], Wi-Fi [6], and Bluetooth Low-Energy (BLE) [7] can work in indoor settings and are simple to deploy as end user devices are typically based on tags or small battery-power devices. However, these technologies present also some drawbacks. They normally require a dedicated and complex infrastructure deployment with anchors/routers, wires, and centralized control computers; as well as an accurate calibration process. While in typical industrial settings such as factory halls these type of deployments might be feasible, the situation in offshore is more complicated (e.g., in oil/gas platforms). In general, these technologies require clear line-of-sight between infrastructure and localization devices to guarantee a correct operation (i.e., at least three visible routers/anchors at all times to be able to perform triangulation), which is difficult in both industrial and offshore scenarios. Moreover, localization based on these technologies does typically not provide heading/orientation information and does not perform well when the localization devices are close to large metal surfaces. The accuracy of these systems in industrial settings

typically range from approximately 20 cm for UWB (with some sporadic deviations to more than 1 m in certain conditions), to approximately 1.5 m for Wi-Fi and BLE.

Technologies based on satellite systems such as Global Navigation Satellite System (GNSS)/Real Time Kinematic (RTK) [8], and Differential Global Positioning Systems (DGPS) [9] can offer up to cm-level accuracy in industrial and offshore environments. However, their applicability is limited to areas with a full or partial view of the sky, hence inapplicable in indoor settings [10]. In general, these technologies rely on no infrastructure (e.g., GNSS, where only an end user device is required) or simple infrastructure (e.g., RTK, DGPS), where a single unit or just a few independent units are typically needed for automatic systems calibration/correction.

As an alternative, it is possible to use novel technologies or combinations of technologies that do not rely on external infrastructure [11]. Here, recent advances in Computer Vision (CV) [12] methods for Simultaneous Localization and Mapping (SLAM) [13], and Visual-Inertial Odometry (VIO) [14] are at hand, allowing for simultaneous high-precision, real-time and low-cost state-estimation (i.e. tracking), provisioning of 6 degrees of freedom (6 DoF) position and orientation information and mapping. Of course, all these benefits have some associated limitations/drawbacks that need to be overcome such as the unavailability of unified commercial solutions, and more complicated integration setups. Other challenges may include occasional poor performance in featureless areas, and the potential need for high computational resources and stability requirements for mobile setups [15]. Sensor fusion strategies (e.g., integration with Inertial Measurement Units (IMU)) might be of utility to overcome some of the said challenges.

Therefore, to make our proposed solution as universal as possible, relying on external infrastructure or specific conditions is avoided, making it compatible with most of the potential industrial and offshore scenarios. Hence, the last group of novel technologies, and its combination, will be explored.

### *2.2. Technologies for corrosion identification*

Identifying corrosion relies on diverse technologies [16] capable of detecting, quantifying, and characterizing corrosion processes in various materials and environments. Commonly employed methods include Ultrasonic Testing (UT) [17], which utilizes high-frequency sound waves for internal and surface defect detection in industries like oil and gas, aerospace, and manufacturing. Eddy Current Testing (ECT) [18] is a method that utilizes electromagnetic

induction, being particularly useful for inspecting non-ferrous metals and detecting corrosion under protective coatings. Radiographic Testing (RT) [19] involves X-rays or gamma rays to inspect internal structures, commonly applied in aerospace, construction, and automotive industries.

Electrochemical Techniques (ET) [20], encompassing methods like impedance spectroscopy and potentiodynamic polarization, study the electrochemical behavior of metals in corrosive environments, providing vital information on corrosion rates and protection effectiveness. Infrared Thermography (IRT) [21], a non-contact method using infrared cameras, is beneficial for inspecting large areas in industries such as building inspection, aerospace, and electrical utilities. Visual Inspection [22], a basic method involving the visual examination of surfaces, is often used alongside other testing methods to assess corrosion severity. Additional technologies like magnetic particle testing [23], acoustic emission testing [24], and scanning electron microscopy [25] also contribute significantly to corrosion identification.

These technologies play a crucial role in preventing material degradation, structural failure, and costly repairs. Our proposed pipeline employs a vision-based deep-learning method, leveraging low-cost camera sensors for accurate and easily deployable results in corrosion assessment.

### *2.3. Camera and LiDAR calibration*

Camera, LiDAR and inertial (i.e. IMU) calibration is a crucial step in sensor fusion for robotics, autonomous vehicles, and augmented reality applications [26]. This process ensures accurate alignment and synchronization between these sensors. Calibration enables the transformation of measurements from different sensors into a common reference frame, allowing seamless integration of data for robust perception and navigation.

Camera calibration involves determining the intrinsic and extrinsic parameters of a camera, which include focal length, principal point, distortion coefficients, and the transformation (rotation and translation) between the camera and the world coordinate system. The calibration is often performed using a calibration target with known geometric features (typically a checkerboard pattern with known dimensions), and the parameters are estimated by minimizing the re-projection error [27].

A set of algorithms for the calibration of various camera sensors, commonly used in robotics applications (e.g. for precise sensor fusion and localization), and can be found in a popular open-source toolbox named Kalibr [28]. These algorithms are not limited to single camera systems and may

be used to calibrate multiple cameras (stereo), of various types, and IMUs (Inertial Measurement Units).

LiDAR-camera calibration involves aligning LiDAR data, with camera to ensure accurate fusion of LiDAR point clouds with camera. The calibration process determines the transformation matrices between the LiDAR and camera coordinate systems [29]. Calibration, in this case, is typically performed by capturing synchronized data from both sensors while the system undergoes controlled motion.

In all cases, optimization techniques, such as nonlinear least squares, are used to estimate the transformation matrices holding the sensors' intrinsic and extrinsic parameters.

### *2.4. LiDAR-based localization and mapping*

LiDAR-based SLAM is an advanced technology that combines data from LiDAR sensors and optionally inertial measurement units (IMUs) to achieve real-time localization and mapping in dynamic environments. LiDAR sensors employ laser beams to measure distances and create detailed three-dimensional (3D) maps of the surroundings, while IMUs, equipped with accelerometers and gyroscopes, capture information about the platform's acceleration and angular rate. The integration of LiDAR and IMU data enhances the accuracy and robustness of the system, especially in scenarios where Global Navigation Satellite Systems (GNSS) may be unreliable, such as indoor or urban environments.

Approaches for LiDAR based SLAM, involve estimating the precise position and orientation of the sensor apparatus in relation to its surroundings, while building a map representation of the environment (typically a point cloud), without prior knowledge of the surroundings. This is achieved through continuous fusion and optimization of LiDAR and, if available, IMU measurements, enabling the system to maintain an accurate and up-to-date understanding of its pose. The integration of these sensors facilitates overcoming challenges like dynamic movements, changes in orientation, and variations in the environment. Loop closure detection is crucial technique employed in SLAM for correcting accumulated errors in both localization and mapping, and typically occurs when the sensing agent revisits a previously seem location [30].

When a map of the environment is known a priori, SLAM simplifies to a localization problem, in which the goal is to estimate the sensor apparatus pose with respect to the map coordinate system. LiDAR based SLAM and

localization systems are crucial for various applications, including robotics, autonomous vehicles, and augmented reality, where high-precision spatial awareness and real-time localization are paramount. Researchers and engineers continue to refine and develop LiDAR SLAM algorithms to enhance performance, robustness, and adaptability across diverse and challenging scenarios. We overview the most relevant state-of-the-art approaches below.

*2.4.1. LiDAR-based SLAM approaches*

A similar work to ours is the one of [31] that proposes a portable LiDAR system for long-term and wide-area people behavior tracking. The authors utilize an optimization-based Simultaneous Localization and Mapping (Graph-SLAM) [32] approach for mapping the environment, while estimating the system's pose and concurrently tracking target individuals. The system operates in two distinct phases: 1) offline environmental mapping and 2) online sensor localization and people detection/tracking. During the offline mapping phase, a comprehensive 3D environmental map covering the entire measurement area is generated. The mapping process employs a Graph-SLAM approach. To address accumulated rotational errors in scan matching, ground plane and GPS position constraints are introduced for indoor and outdoor environments, respectively. The proposed system thus employs a sophisticated mapping strategy to create accurate and comprehensive environmental 3D point cloud representations.

A recent approach entitled Fast-LIO [33], proposed a LiDAR-Inertial Odometry (LIO) framework featuring a Tightly-Coupled Iterated Kalman Filter for robust and efficient motion estimation. The pipeline integraties LiDAR and inertial sensor measurements through a Kalman filter framework. Fast-LIO is designed for real-time applications, such as robotics and autonomous navigation, offering fast and accurate odometry estimation. The key contributions of the approach lie in the technical advancements of the Tightly-Coupled Iterated Kalman Filter, enhancing the reliability and speed of LiDAR-Inertial Odometry for precise motion tracking. In a second iteration [34], the authors propose two new techniques to further improve the performance of the previous algorithms. The first involves direct registration of raw points to the map, eliminating the need for feature extraction. This approach enhances accuracy by exploiting subtle environmental features and is adaptable to various LiDARs. The second novelty lies in maintaining a map using the ikd-Tree, a new incremental k-d tree data structure. This structure facilitates incremental updates, point insertion/deletion, and dy-

namic re-balancing, demonstrating superior overall performance compared to existing dynamic data structures while supporting down-sampling on the tree.

*2.4.2. LiDAR-based localization approaches*

In the work of [35], the authors propose a 3D Monte Carlo approach for localization in dynamic environments in the context of automated driving, utilizing efficient distance field representations. Their approach enhances accuracy in estimating the vehicle's position, especially in scenarios with dynamic elements. The study employs Monte Carlo methods for three-dimensional localization while optimizing the representation of the surrounding environment through efficient distance fields. The research has implications for improving the precision of autonomous vehicles navigating through dynamic settings.

In the work of [36], the authors propose a multi-sensor three-dimensional Monte Carlo localization method designed for long-term aerial robot navigation. The approach integrates information from multiple sensors to enhance the precision and reliability of localization. Utilizing Monte Carlo methods, the system estimates the aerial robot's three-dimensional position over time. The multi-sensor setup aims to address the challenges associated with long-term navigation, offering improved adaptability and robustness. The article discusses the technical details of the proposed method, emphasizing its applicability for sustained and accurate aerial robot navigation in dynamic environments.

In the work of [37], the authors advocate that for the utilization of a simplified and refined Point-to-Point Iterative Closest Point (ICP) registration method, referred to as KISS-ICP. The authors assert that when implemented correctly, this approach offers simplicity, accuracy, and robustness in the context of point cloud registration. The article emphasizes the importance of proper execution to achieve optimal results. KISS-ICP is presented as an efficient and effective solution for point-to-point registration tasks, contributing to improved performance in various applications such as 3D mapping, computer vision, and robotics.

In [38], the authors propose an effective approach for 3D LiDAR-based Monte Carlo based localization. The method involves the fusion of measurement models optimized through importance sampling, aiming to enhance the efficiency and accuracy of the localization process. The article presents a detailed analysis of the proposed solution, highlighting its advantages in

terms of computational efficiency and robustness in handling complex 3D environments. The fusion of optimized measurement models contributes to improved localization performance, making the approach a valuable contribution to the field of autonomous navigation and robotics using 3D LiDAR sensor data.

### *2.5. Image-based semantic segmentation*

The comprehension of scenes through vision-based analysis holds paramount significance across diverse domains, including robotics, manufacturing [39], medical imaging [40], inspection [41, 42, 43, 44], and surveillance [45]. Scene understanding encompasses varied tasks, including image classification, object detection, and semantic segmentation. Vision-based semantic segmentation, addresses the intricate challenge of assigning object class labels to individual pixels within images.

This section undertakes a comprehensive revision and explanation of primary techniques aimed at addressing the aforementioned semantic segmentation challenge. While image classification and object detection respectively deal with classifying images and localizing regions of interest (bounding boxes), semantic segmentation tackles the more intricate task of assigning a class label to each pixel in an image. Approaches documented in the literature can be classified into two distinct paradigms: model-based and data-driven.

Classical computer vision methodologies are based on theoretically principled methods striving to analytically resolve geometric and physical aspects of image formation. In contrast, data-driven methodologies seek to learn statistical properties directly from visual data using machine learning techniques. The ascendancy of deep learning, coupled with the availability of extensive publicly annotated datasets, e.g., Ms COCO [46], CityScapes [47], AED20K [48], to name a few, has resulted in data-driven approaches consistently outperforming their model-based counterparts, particularly in addressing increasingly complex tasks.

One of the initial efficacious forays into deep learning-based semantic segmentation was Mask R-CNN [49]. The architecture exhibits conceptual simplicity, comprising a Convolutional Neural Network (CNN) backbone for robust feature extraction, succeeded by a meticulously optimized Region Proposal Network (RPN) tasked with generating candidate regions of interest. Subsequently, three parallel branches undertake the tasks of classification, bounding box regression, and pixel-level mask predictions. Mask R-CNN, along with subsequent architectures sharing analogous design principles, has

demonstrated state-of-the-art performance across diverse semantic segmentation datasets, notably showcasing excellence on the Microsoft COCO dataset [50].

U-net, a fully convolutional neural network introduced in [51], represents a step forward from previous deep learning based semantic segmentation approaches. This architecture hinges on the novel concept of substituting fully connected layers with up-sampling layers, thereby enabling precise pixel-level predictions. More specifically, U-net adopts an encoder-decoder architecture devoid of fully connected layers. The CNN encoder, operational along the contracting path, strategically downscale the input image to a low-dimensional feature space. Simultaneously, the expansive path of the decoder utilizes de-convolutional layers to up-sample the feature space. U-net has achieved notable success, particularly in the realm of biomedical imaging applications, where it has proven effective in the segmentation of tumor cells. Due to its top performance and popularity, and based on our previous experimental work [39], we utilize U-net model as our corrosion semantic segmentation approach.

## 3. Methodologies

In this section we describe in the detail the design choices behind the proposed portable hand-held LiDAR-inertial camera and data collection system, and our approaches for 3D mapping and localization of corrosion in industrial environments.

### *3.1. Problem formulation*

Offshore assets are not the only locations where corrosion must be managed, but provides an extreme case, where even gaining access to the site is prohibitive. The person conducting the inspection is also unlikely to be the same person evaluating the severity of corrosion. As such, the latter cannot to the same degree rely on prior knowledge of the site when reviewing the dataset.

Considering the presented issues with offshore corrosion management, the overall properties of the envisioned system, while simultaneously acknowledging that such a system may prove beneficial in onshore settings as well, we present a portable LiDAR-inertial camera system that can automatically localize and identify corroded structures, in industrial sites.

In particular, we address several key challenges:

- **Perception and data collection system design:** To support corrosion inspections at various environments, without the need for any external permanent fixtures, a portable platform for hosting sensors and computational resources is required. The platform should be lightweight, portable, and allow recording RGB and point cloud data streams in outdoor environments for prolonged periods of time. We envision a handheld sensor apparatus containing a camera and LiDAR, and a backpack computer and power-system.

- **Sensor calibration:** With the purpose of projecting image-based data accurately to the LiDAR point cloud, we need to find the pose of the *camera system*, relative to the LiDAR, i.e., the extrinsic camera matrix parameters, encoding a transformation between the camera and the LiDAR. This way, one can accurately map RGB and semantic information to LiDAR point clouds, in order to build detailed colored point clouds, representing the surrounding environment.

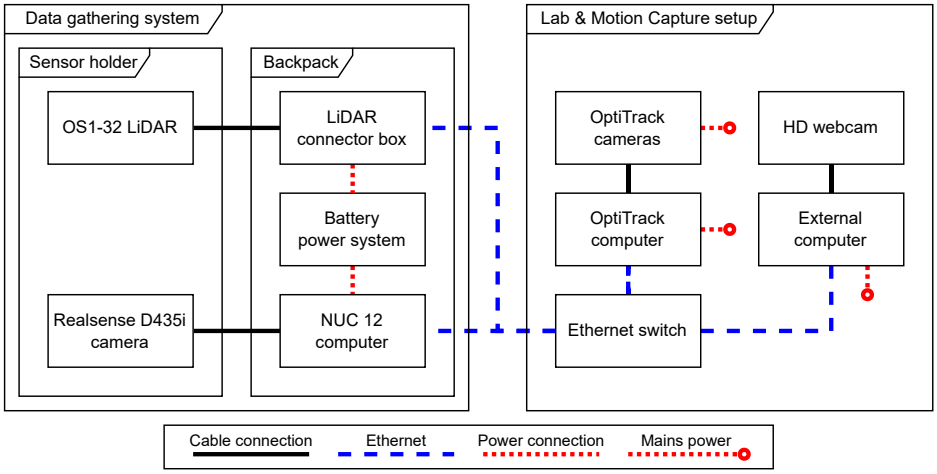
- **3D localization and mapping:** Offshore assets can be geometrically complex and multiple stories tall, and images are also likely to be captured from varied poses. As such, the localization system must be able to operate in 3D, i.e., a detailed map of the environment is required for localization, and must be generated using a LiDAR and a camera system.

- **Corrosion detection:** images may contain numerous small spots of corrosion that can be missed, and some images may feature corrosion runoff or other visual blemishes that are not relevant on their own. Corrosion is also quite varied, depending on the environment and type of metal. A well-defined semantic segmentation system is needed for this task.

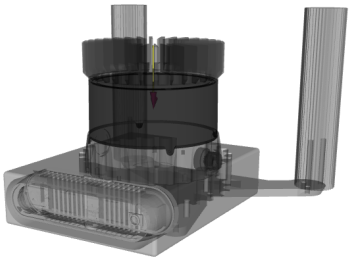
In the rest of this section we describe in detail our proposed solution that addresses the former challenges, through the use of state-of-the-art SLAM and image-based semantic segmentation techniques, for geometric-semantic mapping of corrosion in industrial environments (see Figure 2).

### *3.2. The Design of a Portable Corrosion Identification System*

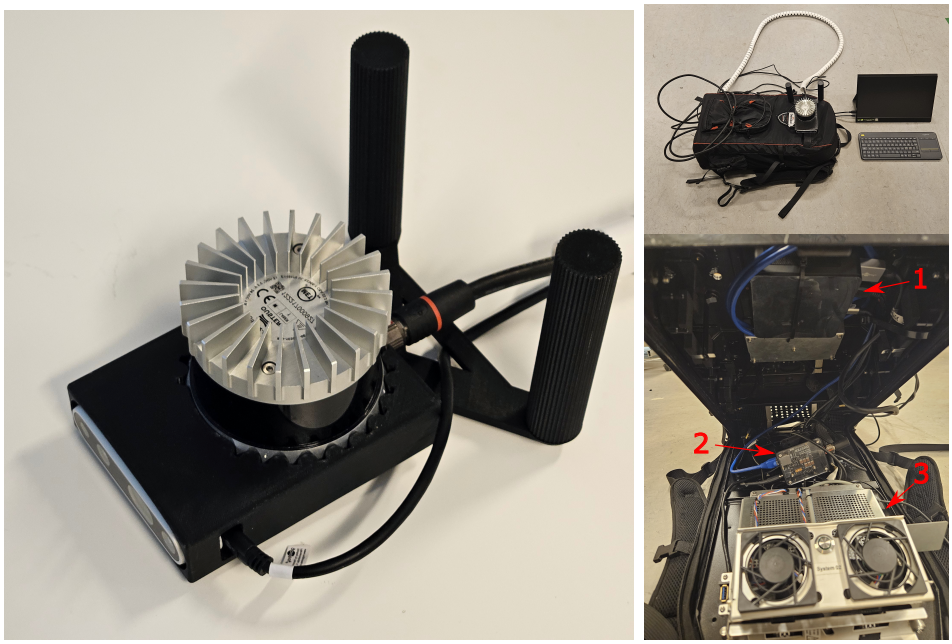
The design of a portable corrosion identification system leveraging LiDAR-inertial camera sensors represents a novel approach in corrosion assessment technology.




(a) Diagram of envisioned system.



(b) Sensor-holder CAD model.




(c) Left: Assembled sensor-holder apparatus. Top right: Full portable system, with optional portable monitor and wireless keyboard. Bottom right: Internals of the backpack, showing computer (1), LiDAR connector box (2) and battery power system (3).

Figure 1: Detailed design view of our portable data capture system for 3D semantic-geometric mapping.

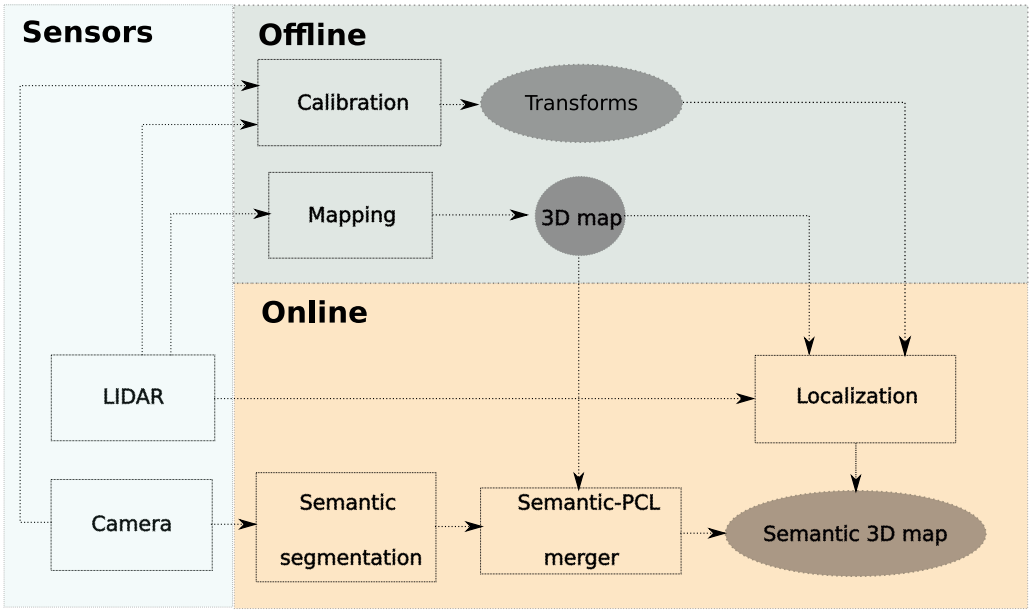



Figure 2: Proposed modular pipeline for semantic-geometric mapping of corrosion in metallic surfaces.

Our proposed system integrates a LiDAR and inertial measurement units (IMUs) for capturing motion and orientation data, for precise 3D localization and mapping, and camera sensors for visual inspection of corrosion spots on metallic surfaces. The fusion of these technologies enables a comprehensive and accurate evaluation of corrosion in diverse environments, in particular, industrial environments. The LiDAR-inertial camera sensors provide the system with the capability to create high-resolution 3D models of the environment, and detailed corrosion mapping.

The portability aspect ensures the system's adaptability for field applications, facilitating on-site inspections in real-world conditions. The design prioritizes the development of a lightweight and user-friendly system that combines the strengths of LiDAR, inertial, and camera sensors to deliver reliable corrosion identification in a portable and efficient manner, catering to the evolving needs of corrosion inspection in various industries.

### *3.3. LiDAR-Camera Geometry*

In order to be able to map 2D camera image pixels $(u, v) \in \mathbb{N}^2$ to LiDAR 3D coordinates $(x_l, y_l, z_l) \in \mathbb{R}^3$ and vice-versa, one needs to know the intrin-

sic parameters and the relative transformation between both sensors. This mapping can be expressed in linear homogeneous coordinates, according to

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{K}\mathbf{T}_{l\to c} \begin{bmatrix} x_l \\ y_l \\ z_l \\ 1 \end{bmatrix}$$

where $\mathbf{K}$ denotes the intrinsic camera parameters matrix, encoding the transformation from the camera coordinate system $\mathcal{C}$ to the pixel coordinate system $\mathcal{I}$, and $\mathbf{T}_{l\to c}$ the transformation from the LiDAR coordinate system $\mathcal{L}$ to the camera coordinate system $\mathcal{C}$. Transformations and 2D and 3D coordinates are, through the rest of the article, expressed in homogeneous form.

*3.3.1. Intrinsic camera parameters*

To know exactly how points in the scene map to the image plane for a given camera, it is necessary to find the intrinsic camera matrix $\mathbf{K}$,

$$\mathbf{K} = \begin{bmatrix} \alpha_x & \gamma & u_0 & 0 \\ 0 & \alpha_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{1}$$

where $\alpha_x = fm_x$ and $\alpha_y = fm_y$, represent the focal distance in pixels, with $f$ denoting the focal distance in camera (metric) coordinates, and $m_x$ and $m_y$ the inverses of the width and height of a pixel in the image plane.

In order to model image distortion due to lens characteristics, we use the Zhang model [27], a widely used lens distortion model in camera calibration. The model introduces rectification coefficients to correct for the non-ideal behavior of lenses, especially radial distortions, incorporating both radial and tangential distortion effects in a camera lens. The model expresses the distorted image coordinates $(u', v')$ in terms of the undistorted coordinates $(u, v)$ using the following polynomial forms,

$$\begin{aligned} u' &= u\left(1 + k_1 r^2 + k_2 r^4 k_3 r^6\right) + 2p_1 uv + p_2(r^2 + 2u^2) \\ v' &= v\left(1 + k_1 r^2 + k_2 r^4 k_3 r^6\right) + p_1(r^2 + 2v^2) + 2p_2 uv \end{aligned}$$

where $k_1, k_2, k_3$ denote the radial distortion coefficients, $p_1, p_2$, the tangential distortion coefficients, and $r$ the radial distance from the principal point.

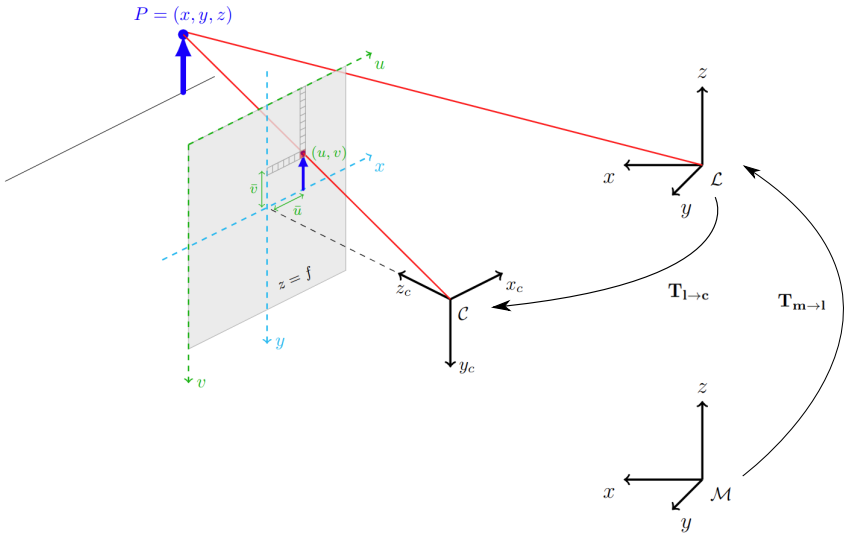



Figure 3: Main coordinate systems and intrinsic parameters considered by our LiDAR camera system, including the camera coordinate system $\mathcal{C}$, the LiDAR coordinate system $\mathcal{L}$ and the map coordinate system $\mathcal{M}$

.

*3.3.2. Extrinsic camera parameters:*

With the purpose of projecting image-based data accurately to the LiDAR point cloud, we need to find the pose of the camera relative to the LiDAR, i.e., the extrinsic camera matrix parameters, which encode a 6D transformation $\mathbf{T_{l\to c}}$ between the camera and the LiDAR, in homogeneous matrix form:

$$\mathbf{T_{l\to c}} = \begin{bmatrix} \mathbf{R} & \mathbf{T} \end{bmatrix} = \begin{bmatrix} r_{xx} & r_{xy} & r_{xz} & t_x \\ r_{yx} & r_{yy} & r_{yz} & t_y \\ r_{zx} & r_{zy} & r_{zz} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2}$$

*3.4. Camera and LiDAR calibration*

In order to estimate the intrinsic camera parameters and the LiDAR to camera rotation and translation (i.e. extrinsic parameters), we utilize two widely adopted algorithms, described below.

*3.4.1. Camera intrinsic calibration*

In this work we rely on the camera calibration procedure of [52], which is a versatile and accurate method for calibrating multiple cameras, and (optionally) multiple IMU sensor suites. The method calibrates both intrinsics and extrinsics and provides a robust parameter identification (i.e. calibration) even in the presence of dynamic motion, in a unified manner. In case of the presence of IMU(s) the procedure incorporates, as well, the IMU biases. The full measurement model includes both camera and IMU measurements, but for the sake of simplicity, we consider only a single camera.

The calibration parameters (intrinsics and extrinsics) are estimated by capturing multiple images of a calibration checkerboard, with known dimensions, from different viewpoints, i.e., by locating and extracting the image coordinates of the checkerboard corners in each image, using a corner detection algorithm. The 3D points coordinates in world coordinates can be mapped by minimizing the re-projection error $E$, i.e., the discrepancy between the observed image points $(u, v)$ and their corresponding projections $(u', v')$ based on the calibrated camera model, according to

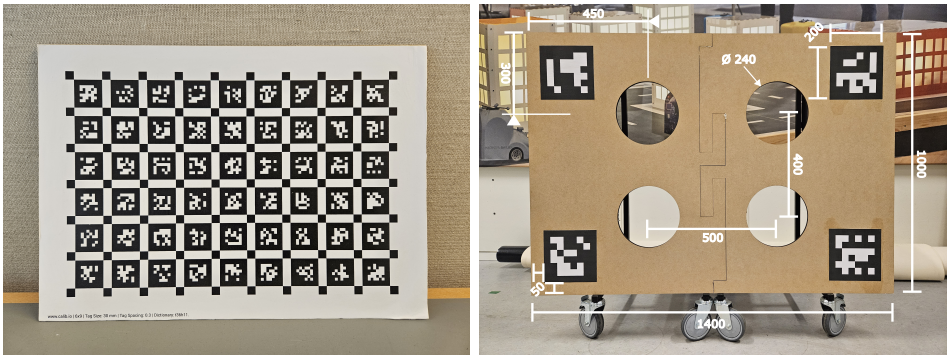
$$E = \sum_i E_i^2 \quad \text{with} \quad E_i = \sqrt{(u' - u)^2 + (v' - v)^2} \tag{3}$$

The parameters are obtained using a nonlinear optimization algorithm (e.g., Levenberg-Marquardt [53]) to minimize the re-projection error. The optimization involves iteratively adjusting both intrinsic parameters and distortion coefficients.

*3.4.2. LiDAR-camera extrinsic calibration*

To find the relative transformation $\mathbf{T}_{l \to c}$ between the LiDAR and camera when sensors are not configured in a predetermined rigid structure, the method described by Beltran et al. [54] is utilised. It is a two-stage solution for calibrating between pairs of monocular cameras, stereo cameras, and LiDARs, using a custom calibration pattern featuring four circular holes and four ArUco markers [55], which can be seen in figure 4b.

In short, the method is based on localizing the centroids of the holes in the target across the two sensors being calibrated, then aligning the two sets of 3D reference points. The following is an overview of the process described by Beltran et al. [54]. Note that details on calibrating with stereoscopic cameras are briefly covered solely for the sake of completeness.




(a) A3-sized AprilGrid camera calibration target, with 8 rows, 12 columns, 22mm tags, and 6.6mm spacing.

(b) Fabricated extrinsic calibration target based on [54], using the default recommended dimensions. All values are in millimeters.

Figure 4: Calibration targets used for intrinsic and extrinsic calibration.

*Reference point estimation.* In the first stage, the goal is to localize the calibration target, and estimate the center points of the holes within, relative to the individual sensors. This is done frame-by-frame, per sensor. Here, there are two pathways: one for 3D input, and one for monocular input.

Stereo-camera feeds are pre-processed into a 3D point cloud, then later used similarly to the LiDAR feed. In both cases, user-adjustable pass-through filters are applied to limit the search space. This requires user intervention, specific to the current physical calibration setup.

Edge points are then identified. For LiDARs, depth discontinuities in the point cloud are used to detect edges [56]. This assumes that points resulting from LiDAR rays passing through the calibration target holes exist, so care must be taken to not leave excessive open space behind the target, nor filter these points out in the preceding step. A visualization of this can be seen in figure 5. For stereo cameras, a Sobel filter [57] is applied to one of the two stereo images, and points are mapped to the resulting pixel values. Points with sufficiently low values are discarded.

It is at this point the two modalities are treated the same. A RANdom Sample Consensus (RANSAC) algorithm [58] is applied to the filtered point cloud data, in order to fit a plane model, representing the target plane surface. The edge-points from the previous step are filtered using the obtained plane model, such that the resulting point cloud only contains points belonging to the target.

Next, the remaining points are projected on the planar model, resulting in a 2D point cloud. The circular holes are then extracted using a 2D circle segmentation. This is performed iteratively, removing inliers, until the remaining points are insufficient to describe further circles. At least four circles must be found to proceed, otherwise the current frame is discarded. Monocular cameras rely on using the four ArUco markers as an ArUco board to estimate the relative position of the target. An example of this can be seen in figure 5.

Once circle centers have been found, the processing of sensor data across modalities are the same. To rule out incorrect detections, a geometric consistency check is performed. The circle center points are grouped in sets of four, and the dimensions of the rectangle they form together are compared against what is theoretically expected. The assumption is that there should only be one valid, geometrically consistent set. If more than one, or zero, valid sets are found, the frame is discarded. Otherwise, the center-points are converted back into 3D space, and the resulting cloud of four points is considered to be valid reference points.

Having a single set of reference points per sensor is technically sufficient, but to reduce potential errors, such as from sensor noise, multiple (i.e. 30) sets are accumulated per sensor, and verified using Euclidean clustering [59]. In case of detecting more than the expected four clusters, the data is considered unreliable, else cluster centroids are used as reference points for the second stage. Furthermore, Beltrán et al [54] allow for (and recommend) accumulating reference points over $M$ target poses, as to generate $4 \times M$ reference points, adding further constraints in order to improve the final results.

*Registration procedure.* In the second stage, the goal is to find the rigid transformation that best aligns the reference point sets with each other. This is handled as a multi-objective optimization, involving $4 \times M$ objective functions.

First, the reference points must be paired, so that each reference point from one sensor is associated with the corresponding reference point from the other sensor. There is no guarantee that reference points in each set are in the same order, so to ensure correct association, the four reference points per target are converted to polar coordinates, and the top-most point (i.e. lowest inclination) is identified. From hereon, the remaining points can then be identified by comparing distances to them, and thus paired correctly.

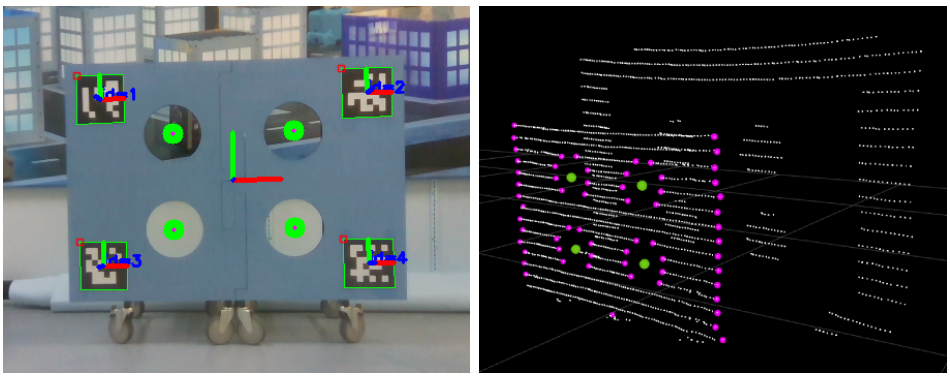


Figure 5: Reference point estimation using ArUco markers for monocular cameras (left) and edge detection for LiDARs (right).

### *3.5. Image-based corrosion identification*

In this work we rely on the popular Deep Neural Network architecture, U-net [51], for image-based semantic segmentation. U-net derives from the fully convolutional network introduced in [60], with modifications tailored to facilitate training with small image samples while surpassing its predecessor in accuracy. The architecture of U-net comprises a conventional convolutional network, coupled with two pathways. The contracting path initiates with two 3x3 unpadded convolutions, succeeded by rectified linear units (ReLU) and downsampling through 2x2 max pooling operations.

In the expansive path, the feature map undergoes up-sampling through 2x2 convolutions, followed by cropping to address the loss of border pixels during convolutions. Subsequently, the cropped feature map from the contracting path is concatenated, and $3 \times 3$ convolutions are applied, followed by rectified linear units (ReLU) as illustrated in Figure 6.

U-net's use of convolutional layers with small receptive fields promotes the extraction of intricate hierarchical features, contributing to its robust representation learning capabilities. The architecture's adaptability to limited training data sets it apart, making it suitable for applications where obtaining extensive labeled samples poses challenges.

The symmetric structure of U-net, with its balanced encoding and decoding paths, plays a pivotal role in maintaining spatial relationships, enhancing feature retention, and boosting its overall performance. A clever combination of skip connections, convolutional layers, adaptability to varied data sizes,

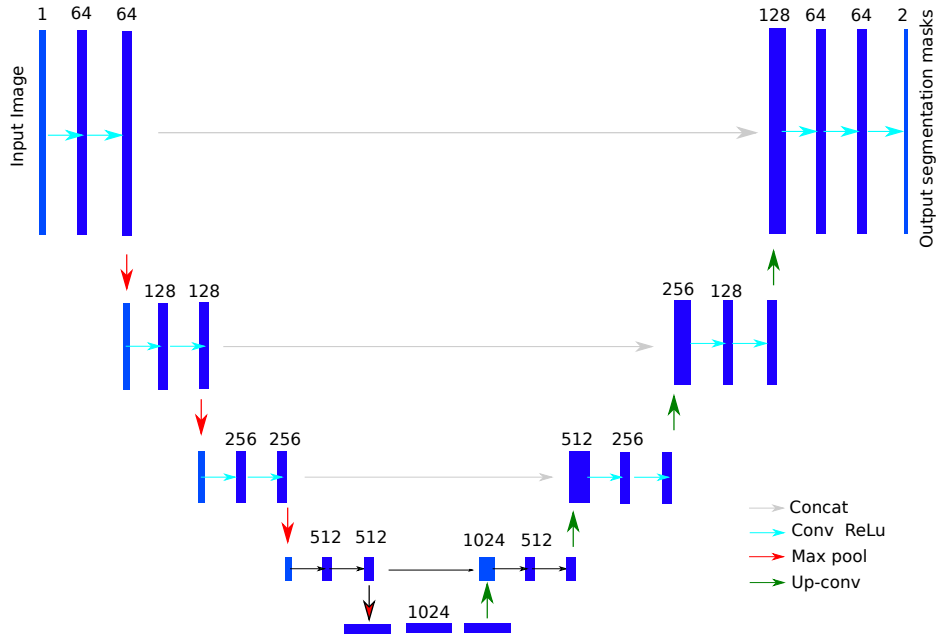



Figure 6: U-Net architecture for a 1024 pixel input (i.e. 32x32) at the lowest resolution. Blue boxes represent multi-channel feature maps, with the number of channels indicated at the top, and the arrows indicate various operations in the network (figure adapted from [51]).

allows achieving high accuracy with limited training data.

### *3.6. LiDAR-based localization and mapping*

Our pipeline for mapping and localization combines two methods. The first, offline, employs a graph-based SLAM approach to build a point cloud map of the environment. The second, given a map, employs a unscented Kalman filter (UKF) [61] based approach to estimate in real-time the location of the sensor apparatus in the map.

#### *3.6.1. Graph-based LiDAR SLAM*

LiDAR-based Graph-SLAM is a sensor fusion technique that utilizes a graph structure to represent the relationships between different poses (positions and orientations) of a LiDAR system over time, along with the associated LiDAR measurements. The optimization process seeks to simultaneously estimate the robot's trajectory and create a consistent map of the environment.

A graph $G = (V, E)$ is built over time, where $V$ is the set of vertices representing sensor poses, and $E$ is the set of edges representing constraints between these poses. The optimal node configuration in the graph, is found using non-linear optimization techniques to minimize the error introduced by the constraints. Let $\mathbf{p}_t$ be the sensors poses at $t$, and $\mathbf{r}_{t,t+1}$ be the relative sensor poses between $t$ and $t+1$ estimated via scan matching [62]. We add them to the pose graph as nodes $[\mathbf{p}_0, ..., \mathbf{p}_N]$ and edges $[\mathbf{r}_{0,1}, ..., \mathbf{r}_{N-1,N}]$. The graph is optimized using the g2o [63] framework, to build a globally consistent 3D map of the environment. To compensate for accumulated pose errors during mapping, the approach allows incorporating ground plane constraints in the graph pose optimization.

The final output of the algorithm is a map $\mathcal{M}$ of size $M$, comprising a finite set of points in $\mathbb{R}^3$. Let us denote by $\mathbf{m}_i \in \mathcal{M}, i \in \{1, ..., M\}$, each point belonging to $\mathcal{M}$.

*3.6.2. Real-time LiDAR-based localization*

During the online localization phase, the localization system [31] estimates its own pose on the previously created 3D point cloud map, by combining a scan-matching algorithm with an angular velocity-based pose predictor. An UKF [61] is used for sequential Bayesian estimation of the pose of the LiDAR-camera sensor apparatus. The scan matching utilizes a Normal Distributions Transform (NDT) [62] for estimating the LiDAR motion between consecutive frames, which outperforms other scan matching approaches, such as the iterative closest point (ICP) [64].

*3.7. Semantic-geometric mapping*

In order to fuse semantic information with the 3D map of the environment, we utilize the known camera intrinsics, extrinsics between the lidar and the camera, and the pose between the LiDAR and the map, given by the localization system.

Given an image $\mathcal{I}$ provided by the system's semantic segmentation algorithm, the semantic color information is projected into the map by combining the transformation between the map and the LiDAR, computed by the localization system $\mathbf{T}_{\mathbf{m}\to\mathbf{l}}$, the transformation between the LiDAR and the camera optical center $\mathbf{T}_{\mathbf{l}\to\mathbf{c}}$. More specifically, for each point belonging to the map, we compute the corresponding image pixel, using homogeneous coordinates according to,

$$\mathbf{u}_i = \mathbf{K}\mathbf{T}_{l\to c}\mathbf{T}_{c\to m}\mathbf{m}_i \quad \text{with} \quad \mathbf{u}_i = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}, \mathbf{m}_i = \begin{bmatrix} x_m \\ y_m \\ z_m \\ 1 \end{bmatrix} \tag{4}$$

and if the point falls within the image plane, its corresponding color is updated, by averaging its current RGB values with the corresponding image pixel one.

## 4. Experiments

In this section we describe in detail the selected sensor apparatus as well as indoor and outdoor experiments to assess the performance and applicability of our methods in a real use-case scenario.

### *4.1. Hardware design and sensor selection*

The design of the portable sensor-holder was a critical aspect behind our efficient data collection in various applications, such as environmental monitoring or industrial inspections. The sensor-holder is versatile, providing a stable platform for mounting sensors of different sizes and types. In our experiments we utilized an Ouster Rev 06 LiDAR, and a RealSense d435i camera during our experiments. The design was selected to offer ease of mobility to capture diverse environments and facilitate data capture. We selected 3D printed material, to ensure easy adaptations during prototyping (e.g. change of sensors), ensure durability in challenging conditions while being lightweight for comfortable use. An ergonomic design that prioritizes user comfort and ease of handling is crucial for prolonged data collection tasks. Additionally, the sensor-holder incorporates features such as screw holes to facilitate the deployment and interchange of sensors.

Complementing the sensor-holder, the design of a backpack for data collection plays a pivotal role in ensuring seamless mobility and accessibility of necessary equipment. The backpack was engineered to be lightweight, and easy to carry for long-periods of time, distributing the weight evenly, reducing strain on the user during extended periods of data collection. Compartments within the backpack should be strategically arranged to accommodate the sensor-holder, ensuring secure storage and easy access. Integration of a power supply system for sensors, along with cable management solutions, is vital to sustain prolonged data collection sessions. The backpack design

Table 1: Hardware used in the experiments.

| Hardware | Used in |
|---|---|
| Ouster OS1-32 Rev 06 LiDAR with uniform beam spacing | All experiments |
| Intel Realsense D435i depth camera | All experiments |
| Intel NUC 12 Pro computer with Intel Core i7-1260P processor | All experiments |
| Portable backpack power system with 4 lithium batteries | All experiments |
| OptiTrack motion capture system with dedicated computer | Indoor experiments |
| High definition webcam with dedicated computer | Indoor experiments |
| Gigabit ethernet switch | Indoor experiments |

should prioritize user comfort with padded shoulder straps and a ventilated back panel to mitigate discomfort caused by extended wear. Furthermore, the incorporation of smart features such as ground truth markers in a controlled laboratory environments allows assessing the overall capabilities of the data collection process, making the backpack an integral component of a well-rounded field data collection system. Please see Table 1 for details on the selected hardware.

### *4.2. Indoor experimental setup*

In order to test the proposed system, and compare our solutions for LiDAR-based localization, experimental data recording sessions in an indoor laboratory setting were conducted. The purpose for these recordings were to create datasets for assessing the performance of the localization algorithms, using a motion capture system. Furthermore, uncorroded and corroded metal parts were placed in the environment, to allow testing the performance of the corrosion detection algorithms.

The setup was as follows: LiDAR and depth camera sensors were affixed to the custom 3D print with handles, as shown in Figure 1c. The NUC 12 computer, along with LiDAR connector box, were installed into and powered by the backpack. Prior to recording, the motion capture system was calibrated, and the tracked area was marked with tape. A rigid-body tracking marker was attached to the top of the LiDAR.

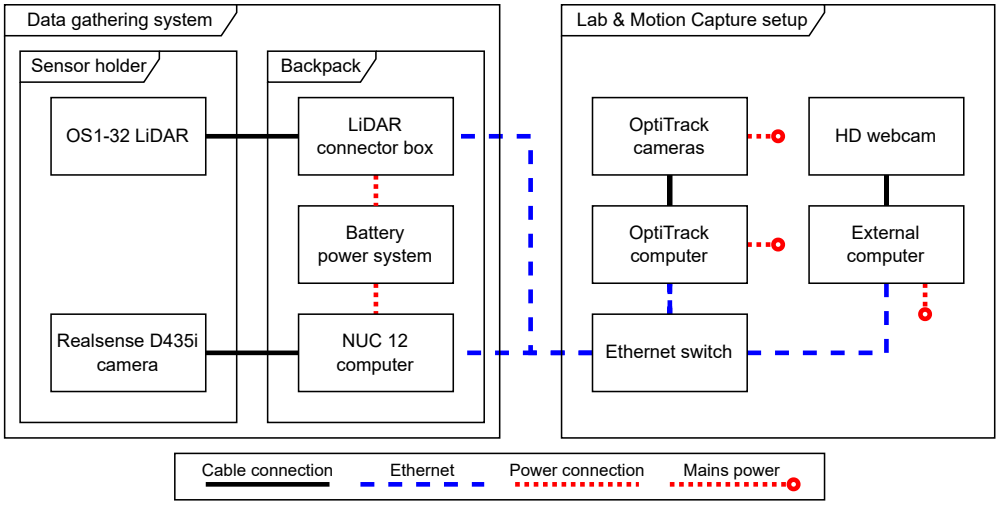
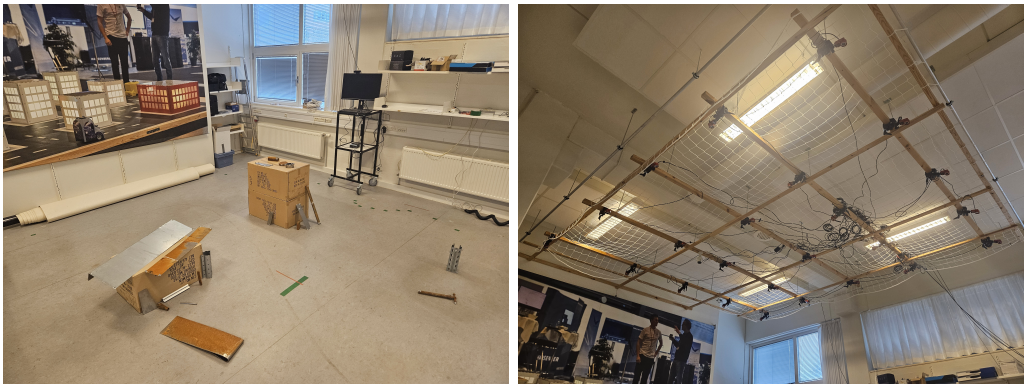



Figure 7: Hardware setup. The Lab & Motion Capture setup is exclusively used in indoor experiments.



(a) General view of the indoor experimental area.

(b) Motion capture system installed on the lab ceiling.

Figure 8: Pictures from indoor experimental setup.

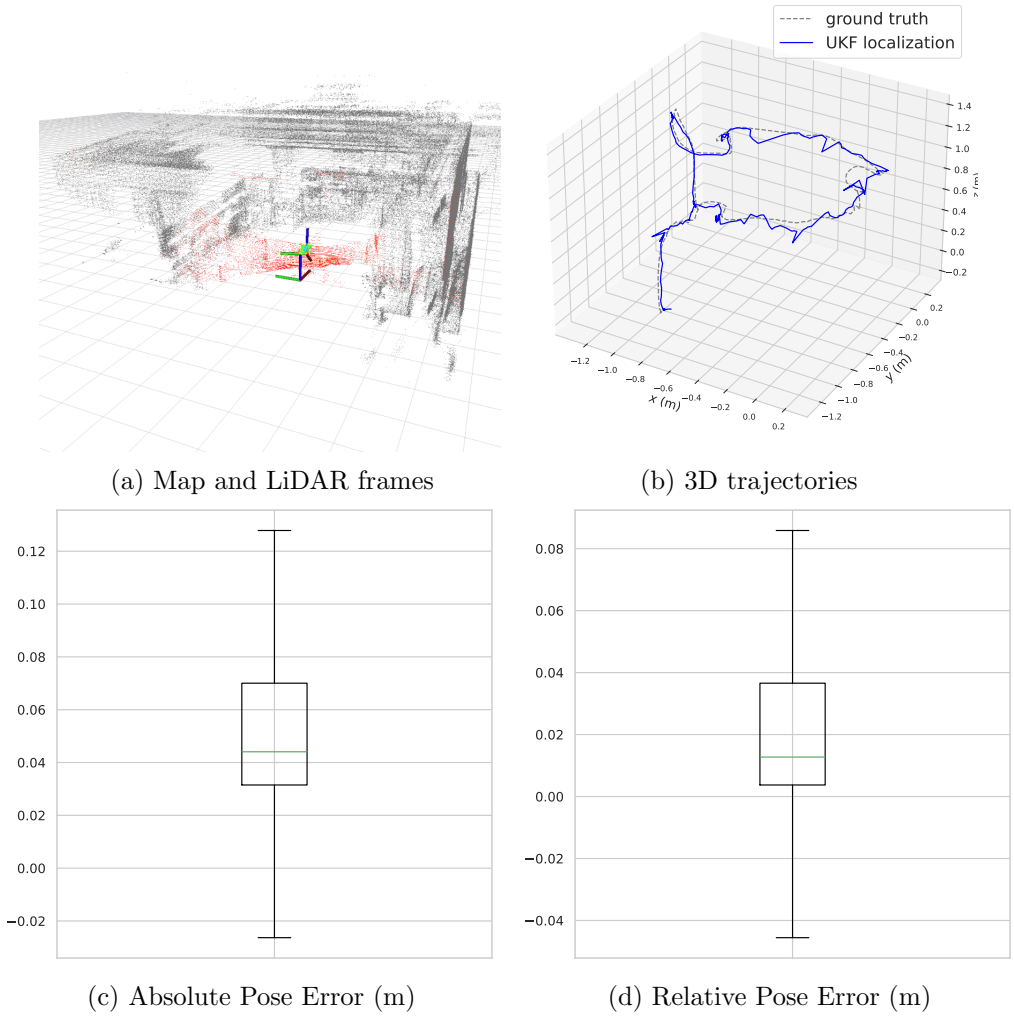
*4.2.1. Localization and mapping performance*

In order to assess the performance of the indoor localization and mapping algorithms, we integrated a high precision optitrack motion capture system, with $0.1mm$ precision, for ground truth acquisition [65], installed in the ceiling of the lab (see Figure 8). A set of markers were placed on top of the LiDAR-camera system in order to build a trackable reference frame. To

quantitatively evaluate the performance of the algorithms, we compared the trajectory given by the ground truth system with the computed trajectory UKF localization algorithms. More specifically, trajectories were temporally and spatially aligned using the Umeyama alignment algorithm [66], and the absolute and relative pose errors were used for quantitative evaluation, i.e., to numerically assess the consistency of the trajectory.

While the absolute pose error (APE) focuses on the accuracy of the overall position and orientation of the LiDAR-camera system in the ground truth frame, i.e., global trajectory consistency, the relative pose error (RPE) is concerned with the accuracy of the changes in pose between consecutive measurements in a local coordinate frame, hence suitable for measuring drift in odometry systems.

Figure 9 (a) depicts the map point cloud obtained with the SLAM algorithm, and the map and LiDAR reference frames during online localization. Figure 9 (b), (c) and (d) show the 3D trajectories of the ground truth and UKF localization algorithm, and the absolute and relative pose errors, respectively. The results demonstrate the high-accuracy of the used LiDAR-based localization system with less then $0.05m$ and $0.02m$ average absolute and relative position errors, respectively. One setback, though, of the employed method resides on the fact that it needs a good pose initialization. However, this limitation can be easily surpassed with the use of Monte Carlo Localization (MCL) based methods (e.g. [67]), at the cost of increased computation effort.




(a) Map and LiDAR frames

(b) 3D trajectories

(c) Absolute Pose Error (m)

(d) Relative Pose Error (m)

Figure 9: LiDAR-based localization performance in indoor experimental scenario.

*4.3. Semantic segmentation*

In this section, we conduct a comparative analysis of advanced semantic segmentation networks outlined in the preceding section, specifically focusing on their effectiveness in segmenting corrosion in metallic structures. Throughout our experiments, we employed a 12th Gen Intel® Core™ i9-12900KF processor (24 cores) and a GeForce RTX 3090ti graphics card for both training and testing.

The dataset utilized for training and testing comprises a total of 14,265 labeled images. These images were captured in a high-definition offshore

Table 2: U-net performance results on corrosion segmentation in offshore environments.

| model | IoU score | Precision | Recall | F-score | Avg. inference time (s) |
|---|---|---|---|---|---|
| UNET-resnet34 | 0.4519 | 0.7112 | 0.5508 | 0.5574 | 0.0445 |

Table 3: Dataset used for training and validating the semantic segmentation networks.

| | total images in dataset |
|---|---|
| train | 8559 (60%) |
| val | 2853 (20%) |
| test | 2853 (20%) |

environment using a DSLR camera and were manually annotated using an online labeling tool [68].

In our experimental approach, we trained a U-net semantic segmentation model, using random crops of size $1024 \times 1024$ extracted from the original input images. The training batch size was set to 8. The dataset was partitioned into training (60%), validation (20%), and testing (20%) sample sizes. The model underwent pre-training on the ImageNet [69] dataset. We use a ResNet-34 as our U-net backbone [70], that consists of 34 layers, being popular in image recognition tasks, due to its clever balance between depth and computational efficiency.

Figure 10 shows qualitatively, the performance of U-net on different scenes. U-net is capable of identifying most ground truth spots, while correctly identifying corrosion spots missed by the labeler. Given the difficulty of accurately labeling corrosion, especially when small spots are easily overlooked by human annotators, precision may not be the most suitable metric for evaluating task effectiveness in this context. As quantitatively shown in Table 2, the method exhibits fast average inference times ($0.0445s$), being suitable for real time application. Although our model achieves high precision when compared with similar corrosion identification approaches, e.g. [71], being able to recognize most ground truth corrosion pixels, without many false positives, recall is relatively low, due to a high rate of false negatives, i.e., missed corrosion spots in the ground truth. We hypothesize that this is due to a relatively high number of missed corrosion spots by the labeler.

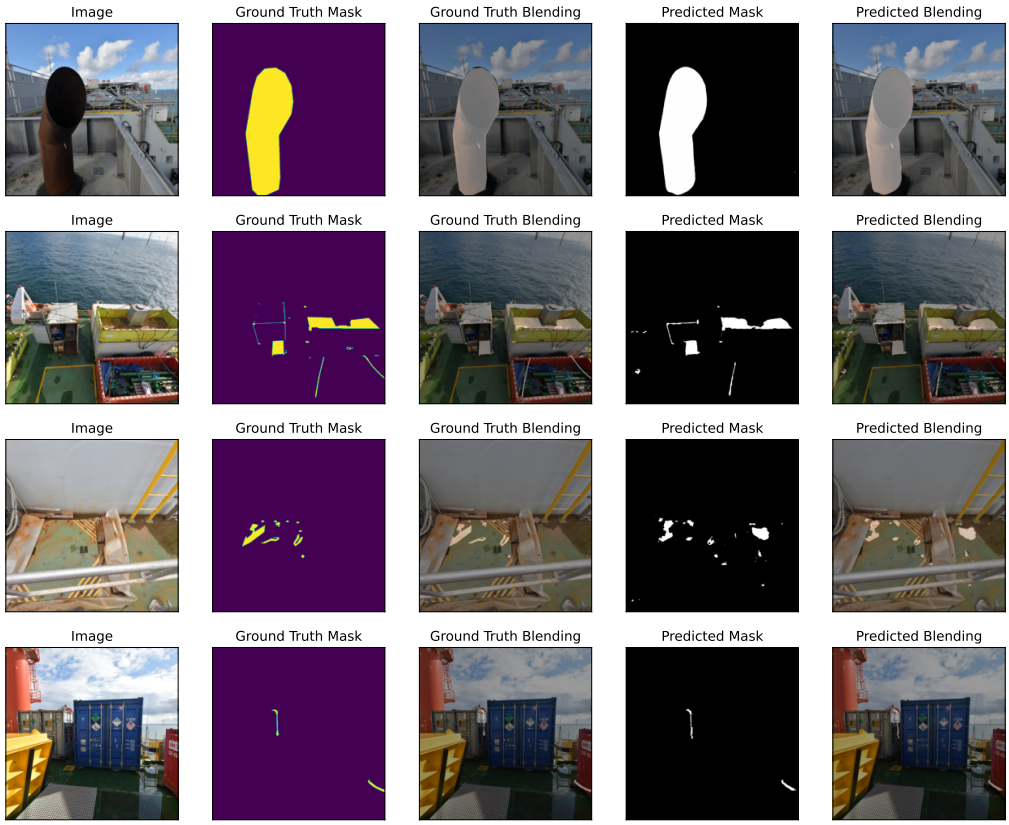



Figure 10: Semantic segmentation masks generated for different example scenes utilizing the U-net architecture featuring a ResNet34 backbone.

## 5. Conclusions and future work

In this work, we proposed a complete system for corrosion identification and 3D mapping of industrial environments. The proposed designed system comprises a portable sensor-holder for a 3D LiDAR and an RGB camera. The system leverages the accuracy of LiDAR data for 3D localization and mapping, with RGB camera information for semantic segmentation of corroded metal structures. Our proposed framework allows building a 3D geometric-semantic map, by fusing semantic data with a prior known 3D map, via known camera and lidar calibration parameters, and a real-time UKF LiDAR-based localization system.

A set of indoor experiments, assessed the accuracy of the individual parts of the mapping and localization system in a laboratory environment.

For future work we intend to replace the current UKF-based localization system with a monte-carlo one, to avoid the need of knowing in advance, the approximate initial location of the system. Also, we intend to improve our semantic map representation, with one that allows fusing semantic and geometric data over time in a probabilistic fashion (e.g. using a probabilistic semantic occupancy grid). Ultimately, we plan to enhance our existing semantic segmentation algorithm by training with large amounts of synthetically generated data.

## Acknowledgement

This work was supported in part by the Danish Energy Agency through the project "Predictive Automatic Corrosion Management" (EUDP 2021-II PACMAN, project no.: 64021-2072); and in part by the Spanish Ministry of Science and Innovation under Ramon y Cajal Fellowship number RYC-2020-030676-I funded by MCIN/AEI /10.13039/501100011033 and by the European Social Fund "Investing in your future". The authors would further like to thank Semco Maritime for bringing up use-case challenges, and to Martin Bieber Jensen for designing and 3D printing the sensor suite holder.

## References

[1] G. H. Koch, M. P. Brongers, N. G. Thompson, Y. P. Virmani, J. H. Payer, et al., Corrosion cost and preventive strategies in the United States, Technical Report, United States. Federal Highway Administration, 2002.

[2] G. Koch, Cost of corrosion, Trends in oil and gas corrosion research and technologies (2017) 3–30.

[3] R. W. Revie, Corrosion and corrosion control: an introduction to corrosion science and engineering, John Wiley & Sons, 2008.

[4] A. Schjørring, A. L. Cretu-Sircu, I. Rodriguez, P. Cederholm, G. Berardinelli, P. Mogensen, Performance evaluation of a uwb positioning system applied to static and mobile use cases in industrial scenarios, Electronics 11 (2022). URL: https://www.mdpi.com/2079-9292/11/20/3294. doi:10.3390/electronics11203294.

[5] A. L. Creţu-Sîrcu, H. Schiøler, J. P. Cederholm, I. Sîrcu, A. Schjørring, I. R. Larrad, G. Berardinelli, O. Madsen, Evaluation and comparison of ultrasonic and uwb technology for indoor localization in an industrial environment, Sensors 22 (2022). URL: https://www.mdpi.com/1424-8220/22/8/2927. doi:10.3390/s22082927.

[6] D.-I. Nastac, E.-S. Lehan, F. A. Iftimie, O. Arsene, B. Cramariuc, Automatic data acquisition with robots for indoor fingerprinting, in: 2018 International Conference on Communications (COMM), 2018, pp. 321–326. doi:10.1109/ICComm.2018.8484796.

[7] S. N. Fatihah, G. R. R. Dewa, C. Park, I. Sohn, Self-optimizing bluetooth low energy networks for industrial iot applications, IEEE Communications Letters 27 (2023) 386–390. doi:10.1109/LCOMM.2022.3215158.

[8] F. Wang, Y. Yu, D. Yang, L. Wang, J. Xing, First potential demonstrations and assessments of monitoring offshore oil rigs states using gnss technologies, IEEE Transactions on Instrumentation and Measurement 71 (2022) 1–15. doi:10.1109/TIM.2022.3208660.

[9] C. Specht, J. Pawelski, L. Smolarek, M. Specht, P. Dabrowski, Assessment of the positioning accuracy of dgps and egnos systems in the bay of gdansk using maritime dynamic measurements, Journal of Navigation 72 (2019) 575–587. doi:10.1017/S0373463318000838.

[10] E. Angelats, P. F. Espín-López, J. A. Navarro, M. E. Parés, Performance analysis of the iopes seamless indoor-outdoor positioning approach, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLIII-B4-2021 (2021) 229–235. doi:10.5194/isprs-archives-XLIII-B4-2021-229-2021.

[11] L. Liu, E. Tan, Z. Q. Cai, Y. Zhen, X. J. Yin, An integrated coating inspection system for marine and offshore corrosion management, in: 2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV), 2018, pp. 1531–1536. doi:10.1109/ICARCV.2018.8581327.

[12] O. C. J. Ricardo Luhm Silva, M. Rudek, A road map for planning-deploying machine vision artifacts in the context of industry 4.0, Journal of Industrial and Production Engineering 39 (2022) 167–180. URL: https://doi.org/10.1080/21681015.2021.1965665. doi:10.1080/21681015.2021.1965665.

[13] J. Choi, M. G. Son, Y. Y. Lee, K. H. Lee, J. P. Park, C. H. Yeo, J. Park, S. Choi, W. D. Kim, T. W. Kang, K. Ko, Position-based augmented reality platform for aiding construction and inspection of offshore plants, The Visual Computer 36 (2020) 2039–2049. URL: https://doi.org/10.1007/s00371-020-01902-9. doi:10.1007/s00371-020-01902-9.

[14] P. Irmisch, D. Baumbach, I. Ernst, Robust visual-inertial odometry in dynamic environments using semantic segmentation for feature selection, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences V-2-2020 (2020) 435–442. doi:10.5194/isprs-annals-V-2-2020-435-2020.

[15] Emerging Inspection Technologies – Enabling Remote Surveys/Inspections, volume Day 1 Mon, May 06, 2019 of *OTC Offshore Technology Conference*, 2019. URL: https://doi.org/10.4043/29450-MS. doi:10.4043/29450-MS.

[16] P. R. Roberge, Corrosion inspection and monitoring, John Wiley & Sons, 2007.

[17] K. Kapoor, K. S. Krishna, S. A. Bakshu, On parameters affecting the sensitivity of ultrasonic testing of tubes: experimental and simulation, Journal of Nondestructive Evaluation 35 (2016) 56.

[18] A. Sophian, G. Tian, M. Fan, Pulsed eddy current non-destructive testing and evaluation: A review, Chinese Journal of Mechanical Engineering 30 (2017) 500–514.

[19] T. N. Bortak, Guide to Protective Coatings Inspection and Maintenance, US Department of the Interior, Bureau of Reclamation, 2002.

[20] R. G. Kelly, J. R. Scully, D. Shoesmith, R. G. Buchheit, Electrochemical techniques in corrosion science and engineering, CRC Press, 2002.

[21] S. Doshvarpassand, C. Wu, X. Wang, An overview of corrosion defect characterization using active infrared thermography, Infrared physics & technology 96 (2019) 366–389.

[22] K.-Y. Choi, S. Kim, Morphological analysis and classification of types of surface corrosion damage by digital image processing, Corrosion Science 47 (2005) 1–15.

[23] M. Lovejoy, Magnetic particle inspection: a practical guide, Springer Science & Business Media, 1993.

[24] A. Zaki, H. K. Chai, D. G. Aggelis, N. Alver, Non-destructive evaluation for corrosion monitoring in concrete: A review and capability of acoustic emission technique, Sensors 15 (2015) 19069–19101.

[25] S. H. Aharinejad, A. Lametschwandtner, Microvascular corrosion casting in scanning electron microscopy: techniques and applications, Springer Science & Business Media, 2012.

[26] C. Debeunne, D. Vivet, A review of visual-lidar fusion based simultaneous localization and mapping, Sensors 20 (2020) 2068.

[27] Z. Zhang, A flexible new technique for camera calibration, IEEE Transactions on pattern analysis and machine intelligence 22 (2000) 1330–1334.

[28] J. Rehder, J. Nikolic, T. Schneider, T. Hinzmann, R. Siegwart, Extending kalibr: Calibrating the extrinsics of multiple imus and of individual axes, in: 2016 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2016, pp. 4304–4311.

[29] F. M. Mirzaei, D. G. Kottas, S. I. Roumeliotis, 3d lidar–camera intrinsic and extrinsic calibration: Identifiability and analytical least-squares-based initialization, The International Journal of Robotics Research 31 (2012) 452–467.

[30] S. Arshad, G.-W. Kim, Role of deep learning in loop closure detection for visual and lidar slam: A survey, Sensors 21 (2021) 1243.

[31] K. Koide, J. Miura, E. Menegatti, A portable three-dimensional lidar-based system for long-term and wide-area people behavior measurement, International Journal of Advanced Robotic Systems 16 (2019) 1729881419841532.

[32] S. Thrun, M. Montemerlo, The graph slam algorithm with applications to large-scale mapping of urban structures, The International Journal of Robotics Research 25 (2006) 403–429.

[33] W. Xu, F. Zhang, Fast-lio: A fast, robust lidar-inertial odometry package by tightly-coupled iterated kalman filter, IEEE Robotics and Automation Letters 6 (2021) 3317–3324. doi:10.1109/LRA.2021.3064227.

[34] W. Xu, Y. Cai, D. He, J. Lin, F. Zhang, Fast-lio2: Fast direct lidar-inertial odometry, IEEE Transactions on Robotics 38 (2022) 2053–2073.

[35] N. Akai, T. Hirayama, H. Murase, 3d monte carlo localization with efficient distance field representation for automated driving in dynamic environments, in: 2020 IEEE intelligent vehicles symposium (IV), IEEE, 2020, pp. 1859–1866.

[36] F. J. Perez-Grau, F. Caballero, A. Viguria, A. Ollero, Multi-sensor three-dimensional monte carlo localization for long-term aerial robot navigation, International Journal of Advanced Robotic Systems 14 (2017). URL: https://doi.org/10.1177/1729881417732757. doi:10.1177/1729881417732757.

[37] I. Vizzo, T. Guadagnino, B. Mersch, L. Wiesmann, J. Behley, C. Stachniss, KISS-ICP: In Defense of Point-to-Point ICP – Simple, Accurate, and Robust Registration If Done the Right Way, IEEE Robotics and Automation Letters (RA-L) 8 (2023) 1029–1036. doi:10.1109/LRA.2023.3236571.

[38] N. Akai, Efficient solution to 3D-LiDAR-based Monte Carlo localization with fusion of measurement model optimization via importance sampling, 2023. URL: https://arxiv.org/abs/2303.00216.

[39] R. P. de Figueiredo, S. Bøgh, Vision-based corrosion identification using data-driven semantic segmentation techniques, in: 2023 IEEE International Conference on Imaging Systems and Techniques (IST), 2023, pp. 1–6. doi:10.1109/IST59124.2023.10355698.

[40] H. Seo, M. Badiei Khuzani, V. Vasudevan, C. Huang, H. Ren, R. Xiao, X. Jia, L. Xing, Machine learning techniques for biomedical image segmentation: An overview of technical aspects and introduction to state-of-art applications, Medical Physics 47 (2020) e148–e167. doi:10.1002/mp.13649.

[41] D. Tøttrup, S. L. Skovgaard, J. l. F. Sejersen, R. Pimentel de Figueiredo, A real-time method for time-to-collision estimation from aerial images, Journal of Imaging 8 (2022). URL: https://www.mdpi.com/2313-433X/8/3/62. doi:10.3390/jimaging8030062.

[42] J. le Fevre Sejersen, R. Pimentel De Figueiredo, E. Kayacan, Safe vessel navigation visually aided by autonomous unmanned aerial vehicles in congested harbors and waterways, in: 2021 IEEE 17th International Conference on Automation Science and Engineering (CASE), 2021, pp. 1901–1907. doi:10.1109/CASE49439.2021.9551637.

[43] R. P. de Figueiredo, J. le Fevre Sejersen, J. G. Hansen, M. Brandão, E. Kayacan, Real-time volumetric-semantic exploration and mapping: An uncertainty-aware approach, in: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2021, pp. 9064–9070. doi:10.1109/IROS51168.2021.9635986.

[44] R. De Figueiredo, J. Le Fevre Sejersen, J. Hansen, M. Brandão, Integrated design-sense-plan architecture for autonomous geometric-semantic mapping with uavs, Front. Robot. AI 9 (2022). doi:https://doi.org/10.3389/frobt.2022.911974, funding Information: This work was supported by UKRI/EPSRC THuMP (EP/R033722/1) and the Smart Industry Program (European Regional Development Fund and Region Midtjylland, grant no.: RFM-17-0020). Publisher Copyright: Copyright © 2022 Pimentel de Figueiredo, Le Fevre Sejersen, Grimm Hansen and Brandão.

[45] M. Gruosso, N. Capece, U. Erra, Human segmentation in surveillance video with deep learning, Multimedia Tools and Applications 80 (2021) 1175–1199.

[46] T.-Y. Lin, M. Maire, S. J. Belongie, J. Hays, P. Perona, D. Ramanan,

P. Dollár, C. L. Zitnick, Microsoft coco: Common objects in context, ArXiv abs/1405.0312 (2014).

[47] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, B. Schiele, The cityscapes dataset for semantic urban scene understanding, in: Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[48] S. Zheng, J. Lu, H. Zhao, X. Zhu, Z. Luo, Y. Wang, Y. Fu, J. Feng, T. Xiang, P. H. Torr, et al., Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 6881–6890.

[49] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask r-cnn, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 2961–2969.

[50] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, Microsoft coco: Common objects in context, in: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13, Springer, 2014, pp. 740–755.

[51] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W. M. Wells, A. F. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer International Publishing, Cham, 2015, pp. 234–241.

[52] P. Furgale, J. Rehder, R. Siegwart, Unified temporal and spatial calibration for multi-sensor systems, in: 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, 2013, pp. 1280–1286.

[53] S. J. Wright, Numerical optimization, 2006.

[54] J. Beltran, C. Guindel, A. D. L. Escalera, F. Garcia, Automatic extrinsic calibration method for lidar and camera sensor setups, IEEE Transactions on Intelligent Transportation Systems 23 (2022) 17677–17689. doi:10.1109/TITS.2022.3155228.

[55] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, M. J. Marín-Jiménez, Automatic generation and detection of highly reliable fiducial markers under occlusion, Pattern Recognition 47 (2014) 2280–2292. doi:10.1016/j.patcog.2014.01.005.

[56] J. Levinson, S. Thrun, Automatic online calibration of cameras and lasers, Robotics: Science and Systems Foundation, 2013. URL: http://www.roboticsproceedings.org/rss09/p29.pdf. doi:10.15607/RSS.2013.IX.029.

[57] I. Sobel, History and definition of the sobel operator, Retrieved from the World Wide Web 1505 (2014).

[58] S. M. Isa, S. A. Shukor, N. Rahim, I. Maarof, Z. Yahya, A. Zakaria, A. Abdullah, R. Wong, Point cloud data segmentation using ransac and localization, in: IOP Conference Series: Materials Science and Engineering, volume 705, IOP Publishing, 2019, p. 012004.

[59] A. Nguyen, B. Le, 3d point cloud segmentation: A survey, in: 2013 6th IEEE conference on robotics, automation and mechatronics (RAM), IEEE, 2013, pp. 225–230.

[60] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3431–3440. doi:10.1109/CVPR.2015.7298965.

[61] E. A. Wan, R. Van Der Merwe, The unscented kalman filter, Kalman filtering and neural networks (2001) 221–280.

[62] M. Magnusson, A. Lilienthal, T. Duckett, Scan registration for autonomous mining vehicles using 3d-ndt, Journal of Field Robotics 24 (2007) 803–827.

[63] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, W. Burgard, G2o: A general framework for graph optimization, in: 2011 IEEE International Conference on Robotics and Automation, 2011, pp. 3607–3613. doi:10.1109/ICRA.2011.5979949.

[64] P. Besl, N. D. McKay, A method for registration of 3-d shapes, IEEE Transactions on Pattern Analysis and Machine Intelligence 14 (1992) 239–256. doi:10.1109/34.121791.

[65] J. S. Furtado, H. H. Liu, G. Lai, H. Lacheray, J. Desouza-Coelho, Comparative analysis of optitrack motion capture systems, in: Advances in Motion Sensing and Control for Robotic Applications: Selected Papers from the Symposium on Mechatronics, Robotics, and Control (SMRC'18)-CSME International Congress 2018, May 27-30, 2018 Toronto, Canada, Springer, 2019, pp. 15–31.

[66] W. Kabsch, A solution for the best rotation to relate two sets of vectors, Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography 32 (1976) 922–923.

[67] J. M. Bedkowski, T. Röhling, Online 3d lidar monte carlo localization with gpu acceleration, Industrial Robot: An International Journal 44 (2017) 442–456.

[68] Segments.ai, 2D & 3D data labeling, 2022. `https://segments.ai/`, Last accessed on 2022-11-30.

[69] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, Communications of the ACM 60 (2017) 84–90.

[70] B. Koonce, B. Koonce, Resnet 34, Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization (2021) 51–61.

[71] E. Bianchi, M. Hebdon, Development of extendable open-source structural inspection datasets, Journal of Computing in Civil Engineering 36 (2022) 04022039. doi:10.1061/(ASCE)CP.1943-5487.0001045.