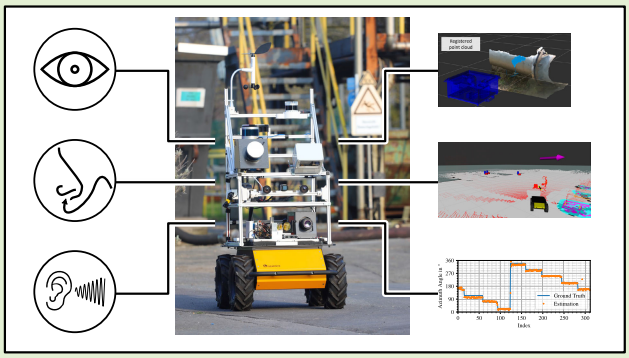
Evaluation of a Smart Mobile Robotic System for Industrial Plant Inspection and Supervision

Georg K. J. Fischer, Max Bergau, D. Adriana Gómez-Rosal, Andreas Wachaja, Johannes Graeter, Matthias Odenweller, Uwe Piechottka, Fabian Höflinger, Nikhil Gosala, *Graduate Student Member, IEEE*, Niklas Wetzel, Daniel Büscher, Abhinav Valada, *Member, IEEE*, and Wolfram Burgard, *Fellow, IEEE*

***Abstract*—Automated and autonomous industrial inspection is a longstanding research field, driven by the necessity to enhance safety and efficiency within industrial settings. In addressing this need, we introduce an autonomously navigating robotic system designed for comprehensive plant inspection. This innovative system comprises a robotic platform equipped with a diverse array of sensors integrated to facilitate the detection of various process and infrastructure parameters. These sensors encompass optical (LiDAR, stereo, and ultraviolet (UV)/infrared (IR)/RGB cameras), olfactory (electronic nose), and acoustic (microphone array) capabilities, enabling the identification of factors such as methane leaks, flow rates, and infrastructural anomalies. The proposed system underwent individual evaluation at a wastewater treatment site within a chemical plant, providing a practical and challenging environment for testing. The evaluation process encompassed key aspects such as object detection, 3-D localization, and path planning, achieving an average precision (AP) of around 0.7662 in the 2-D object detection. In addition, specific evaluations were conducted for optical methane leak detection and localization, demonstrating the system's capability to detect leaks as small as 40 mL min^{-1}. Furthermore, acoustic assessments focusing on pump equipment and gas leak localization yielded an absolute localization error of around 50 cm.**



***Index Terms*—Acoustic localization, acoustic signal processing, anomaly detection, autonomous robots, change detection, chemical plant supervision, distributed AI system, gas detection, industry 4.0, object detection.**

Manuscript received 12 February 2024; revised 3 April 2024; accepted 7 April 2024. Date of publication 1 May 2024; date of current version 14 June 2024. This work was supported in part by the German Federal Ministry of Education and Research (BMBF) under Grant FKZ 16ME0023K ("ISA4.0") and in part by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) under Grant FKZ 03EE3066D. The associate editor coordinating the review of this article and approving it for publication was Prof. Xiaofeng Yuan. *(Georg K. J. Fischer, Max Bergau, and D. Adriana Gómez-Rosal contributed equally to this work.) (Corresponding author: Georg K. J. Fischer.)*

Georg K. J. Fischer and Fabian Höflinger are with the Fraunhofer Institute for Highspeed Dynamics, Ernst-Mach-Institute (EMI), 79104 Freiburg, Germany (e-mail: georg.fischer@emi.fraunhofer.de).

Max Bergau is with the Sensors Automation Laboratory, Endress+Hauser Digital Solutions GmbH, 79110 Freiburg, Germany.

D. Adriana Gómez-Rosal, Nikhil Gosala, Niklas Wetzel, Daniel Büscher, and Abhinav Valada are with the Department of Computer Science, University of Freiburg, 79085 Freiburg, Germany.

Andreas Wachaja and Johannes Graeter are with dotscene GmbH, 79106 Freiburg, Germany.

Matthias Odenweller and Uwe Piechottka are with Evonik Operations GmbH, 45127 Essen, Germany.

Wolfram Burgard is with the Department of Engineering, University of Technology Nuremberg, 90443 Nuremberg, Germany.

This article has supplementary downloadable material available at https://doi.org/10.1109/JSEN.2024.3390622, provided by the authors.

Digital Object Identifier 10.1109/JSEN.2024.3390622

I. Introduction

CHEMICAL production plants must adhere to stringent safety standards, and the responsibility of routine plant inspection typically falls on trained professionals. These inspections involve recording essential process parameters and identifying anomalies during safety rounds. This may include assessing local gauge values, visually inspecting for abnormalities, conducting olfactory and acoustic inspections to detect gas or fluid leaks, and identifying changes in the environment such as broken pipes or blocked passages. In environments within Ex-Areas or those that are harsh or remote, deploying human inspectors can pose significant dangers and incur financial burdens. Alternatively, the installation of fixed sensor equipment, while effective, can be costly and add to the maintenance workload. Moreover, the deployment of high-precision sensors across an entire plant is often economically impractical, especially when looking at older brownfield plants. Recognizing these challenges, there is a growing interest in mobile inspection robots. Recent advancements in robotic technology, exemplified by robots such as Boston Dynamics Spot and others, have made it possible to access more challenging areas of industrial plants [1]. While

these robots primarily serve as equipment carriers, numerous questions remain unanswered. These include determining the optimal sensors for deployment, devising effective data processing methods, and developing efficient navigation strategies for these robots within the plant environment. As mobile inspection robots gain prominence, addressing these questions becomes imperative for their successful integration into industrial safety and inspection processes.

Sensing the state of industrial processes entails a diverse array of sensors. For instance, detecting and localizing gas leakages in chemical plants involve a range of sensors from metal oxide sensors to laser-based detection systems [2]. This diversity extends across industries, addressing changes in the environment such as obstacles obstructing critical transport routes, malfunctioning equipment (e.g., pipes, motors, and valves), and parameters such as foam build-up, which require optical assessment. Replacing human inspection professionals presents a challenging task, as each use case typically requires bespoke development. This may involve the definition of specific physical parameters to observe, selecting an appropriate sensor, establishing a tailored signal, and data processing chain. However, different use cases can be independently implemented, with the robot serving the role of safely transporting measurement equipment to and from sites. Advancements in mapping [3], [4], localization [5], and navigation [6] in mobile robotics are essential for achieving human-level navigation autonomously, encompassing tasks such as object detection and obstacle avoidance. Although the opportunities for autonomous supervising robots in industrial plants are growing, few works of robots equipped with different sensing capabilities for such tasks have been reported or are still at a development stage, and furthermore, fewer works have been validated in a fully working industrial plant. This work aims to contribute to this field. In this article, we introduce a system that outlines the evolutionary trajectory in this domain. We delve into the development of the robot's navigation capabilities and present a subset of developed use cases focused on gathering essential process information.

This article serves as an extension of the previously presented conference paper [7], augmenting the results with tangible data and providing an extensive literature review of existing works. In addition, this contribution includes a more comprehensive exploration of the working principles of our inspection system and the associated sensors in a dedicated section.

II. Related Work

Robots have established their presence in the manufacturing sector, particularly in roles involving goods transportation within factories and warehouses, as seen with automated guided vehicles (AGVs) [8], [9]. However, the application of mobile robotics in plant inspection presents a more complex challenge. On the one hand, specific sensors need to be developed for particular use cases, while, on the other hand, navigation and locomotion capabilities must reach a level that ensures safe and secure operation in hazardous areas and effective collaboration with humans and other vehicles.

1) Robotic System: The concept of industrial inspection robots is not new; nevertheless, mobile robots for industrial inspection are an active research topic. Furthermore, their mobility within dynamic environments is a more recent development [10], [11], [12], [13], [14], [15]. Early works addressed inspection of industrial pipelines [16] and nuclear power plants [17], [18]. Later, these proposals evolved to include a broader range of industrial applications, such as inspection of waste deposits and electric power cables [19]. Soon after, these experiments ventured into the civil realm with inspection of building facades, concrete constructions [20], structures such as ships, wind turbines, and aircraft [21], and even examination of underwater facilities [22]. Furthermore, a recent development includes multiple proposals of systems for surveying nuclear power plants for radiation levels [23], [24].

Recent work in plant inspection [25] reviews chemical sensing applications for mobile robots, addressing challenges such as the quest for improved chemical sensors and the development of more intelligent behavior.

In the context of diverse industries, a study [26] investigates various approaches to inspection in the oil and gas sector. This review encompasses diverse designs for inspections, including those tailored for vertical structures, pipelines, and underwater environments. The introduction of robots into inspection processes significantly enhances safety by eliminating the need to deploy humans in risky areas. Yet, expertise in technology deployment is crucial, as highlighted in another study [27] that delves into various use cases for robots in hazardous sites, emphasizing the need for specialized knowledge in deploying this technology.

Focusing on the exploration and mapping of gas sources, a specific study [2] is centered on gas detection principles and reviews existing works proposing mobile inspection vehicles. Together, these works showcase the evolving landscape of mobile robotics in-plant inspection, emphasizing both the technological advancements and the need for domain-specific expertise in deploying these systems effectively and safely.

In the study [28], a multirobot cooperative system is proposed, aiming to fuse data captured by multiple robots and process them through downsampling, compression, and dynamic bandwidth management. This approach is geared toward achieving a low-latency digital twin. Given the practical challenge of sensor crosstalk, a multirobot system could potentially offer a solution to this issue. In addition, from an economic standpoint, there is potential to enhance efficiency by deploying expensive yet capable robots in challenging areas while employing cheaper and simpler robots for less demanding environments.

However, autonomously conducted inspections of industrial settings, such as processing plants, present ongoing research challenges, such as regarding autonomous navigation, retrieval of measurements in disordered environments, or its consequent interpretation. Hence, to faithfully relieve human inspectors from these challenging tasks, the proposals need to enhance their intelligence and reliability. In this regard, few initiatives aim to design an inspection robot with all their required functions as a centralized, and yet flexible, system. This is the gap that our system seeks to address.

2) Microphone Array: Sound source localization (SSL) is a prevalent subject in robotics, with a comprehensive review of common methods provided in [29]. Its applications span human–machine interaction (HMI), speaker or asset localization, and event detection such as identifying leakages in the surrounding environment [30], [31], [32], [33]. While gas or air leak localization is crucial for industrial inspection and has been previously addressed, existing approaches often rely on fixed installed acoustic transducers and static time-domain features for locating points of leakage in pipes [34]. The work in [35] applies a four-element linear array to find two leakage sources in the emission range of 63–187 kHz. The work in [33] utilizes ultrasonic emissions from leaks in pressurized pipes, employing a peak search on a beam-formed spectrum. By leveraging multiple poses of a robot, potential leaks are localized, necessitating a microphone array with 32 elements. In contrast, our work introduces an algorithm capable of achieving accurate sound source localization with a significantly reduced number of microphones, providing a more resource-efficient solution for industrial applications.

3) Gas Cameras: Modern optical gas imaging (OGI) cameras typically utilize a mid- or long-wave-infrared (IR) camera in combination with a bandpass filter to detect and visualize gas leaks [36]. The gas of interest absorbs parts of the omnipresent background IR radiation, leading to a contrast in the IR camera image. This principle allows for the detection of a wide range of gases, including methane, ammonia, and sulfur hexafluoride. In this work, we present an active gas camera that does not rely on background illumination. Instead, it employs its own active illumination in the form of a laser, which allows for more sensitivity, gas selectivity, and accurate concentration estimation. It comes to the cost of a shorter detection range and the need for a background [37]. Such active approaches are yet rare due to costs and complexity. A laboratory setup has been published by Strahl et al. [38] and a much simplified version of a camera by Nutt et al. [39]. Commercially, the company QLM Technologies is working on an advanced active scanning approach [40].

The main contributions of this article are given as follows.

1) An autonomous mobile robotic system equipped with a diverse and integrated set of sensors, including olfactory, optical, and acoustic capabilities. This advanced system is designed not only for gas detection but also for infrastructure monitoring.
2) A robot navigation framework, enabling flexible and independent mobility within industrial settings.
3) In-depth examination of the performance of the integrated sensors under diverse conditions. Notably, the study includes a detailed analysis of sensor functionality within the wastewater treatment facility of a chemical plant situated in Marl, Germany.
4) To the best of the authors' knowledge, an active gas camera has been mounted on a mobile robotic platform for the first time. Since the active approach requires reflected light from a background, the robot has the potential to adapt its filming position, thereby finding ideal illumination locations.
5) Apart from hyperspectral imaging, which is typically significantly more expensive, to the best of our knowledge, for the first time, methane leaks as small as 40 mL min^{-1} leaks have been filmed in industrial conditions in real time.

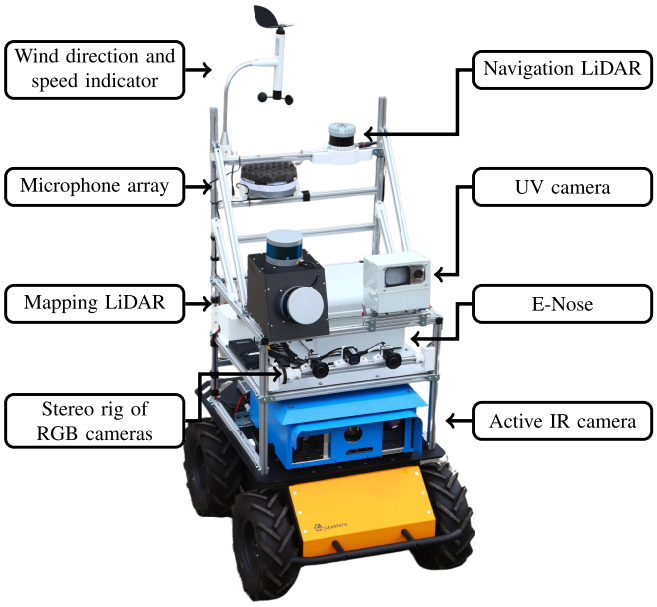


Fig. 1. Mobile robotic platform with multimodal sensors.

III. System Architecture

Our robotic system, as shown in Fig. 1, consists of a mobile robot platform integrated with diverse multimodal sensors, as summarized in Table I. The robot navigates the industrial plant autonomously with its onboard processing capabilities. However, tasks that require more resources, such as object or anomaly detection, are executed on a remote server. To this end, the data are transmitted over the wireless network employing the robot operating system (ROS) [41]. Table I gives an overview of the equipped sensors and devices.

Two approaches of simultaneous localization and mapping (SLAM) are employed: a 2-D and a 3-D SLAM. The 2-D SLAM is used for localization and navigation in the industrial area. This leverages the nature of the mobile ground robot and turns the navigation into a simpler, more reliable, and faster task. On the other hand, the 3-D SLAM is primarily used to generate a dense 3-D map representation of the environment. This fine representation allows for posterior tasks such as finding geometrical differences (while navigating) and embedding relevant information in the 3-D map.

The core software component responsible for centralizing the interactions, control, and data access is deployed on the robot's onboard computer and is managed by the ROS middleware. It provides modules for package management, hardware abstraction, and communication in a networked fashion. This enabled the connectivity between the robot and desktop computers to each single board computer (SBC) contained in the sensing devices.

Fig. 2 shows the system architecture and communication channels, comprised of the computer cluster on the mobile robot platform and the stationary components, such as the industrial remote computer and the remote joystick control. Most of the sensing devices count with an SBC, which is responsible for the connectivity to other onboard computers

TABLE I
OVERVIEW OF EQUIPPED SENSORS AND DEVICES

Device	Specs.	Use Cases
Electronic Nose	• 3 × non-selective MOX sensors	• Methane, CO2, flammable gas detection
Active IR Camera	• Methane rate $\geq 40\,\mathrm{mL\,min^{-1}}$ • Wind speed $\leq 2\,\mathrm{m\,s^{-1}}$ • Distance to Leak $\leq 3\,\mathrm{m}$	• Methane Leak Localization • Concentration Length Measurements • Flowrate Quantification
UV Camera	• 365 nm UV-LEDs • OpenMV Cam H7R2 CMOS camera	• UV-excited fluorescence • Oil film detection
Microphone Array	• UCA • 5 × TDK ICS-40720 MEMS-microphones • ∅ 6.8 cm	• Acoustic anomaly detection • Acoustic machine condition monitoring • Leakage detection
Mapping LiDAR	• 2 × Velodyne VLP16 laser scanner • IMU	• 3D mapping • Geometric changes and anomaly detection
Passive Cameras	• 2× RGB Blackfly S GigE FLIR cameras • FLIR Boson Long wave infrared (LWIR) thermal camera	• Dynamic object detection (pedestrians, obstacles, etc.)
Navigation LiDAR	• OS128 Ouster LiDAR. • 90 deg vertical field of view, 50 m range	• 2D mapping • Localization and navigation

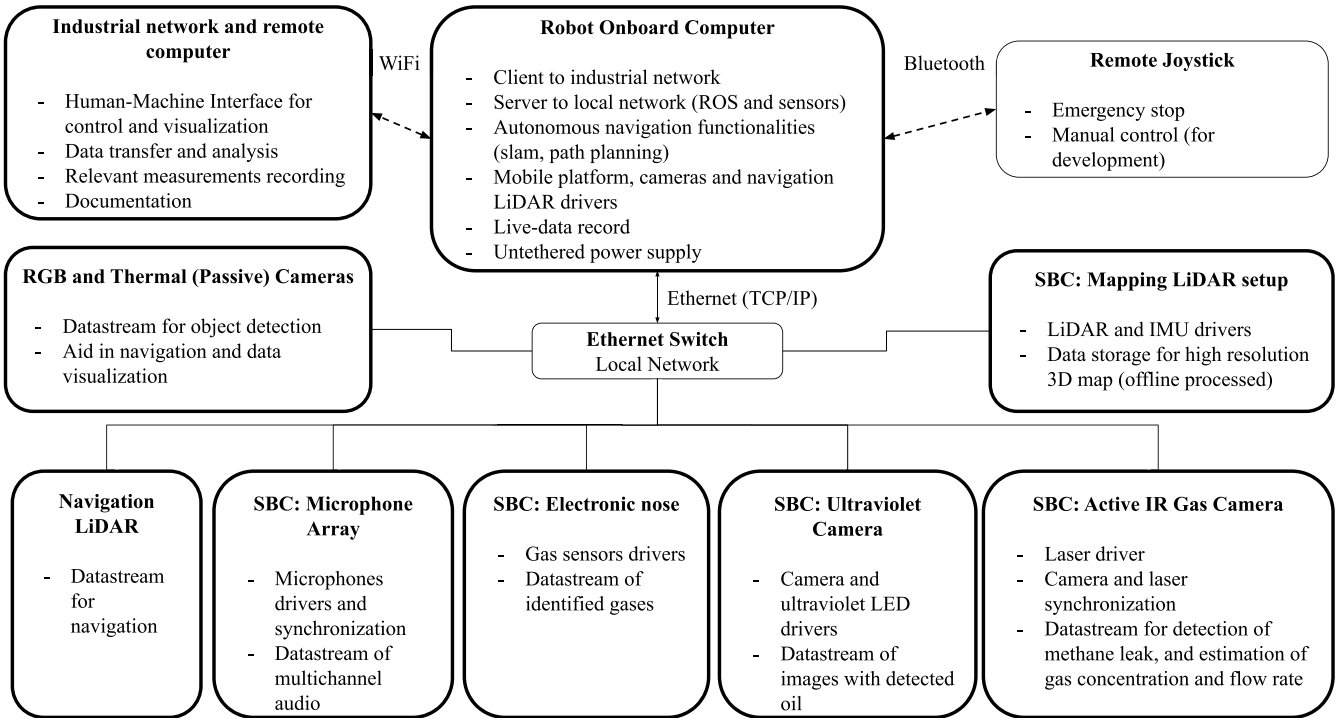


Fig. 2. System architecture with communication between the components. Solid or dotted lines between the boxes represent, respectively, Ethernet or wireless connectivity.

via a local Ethernet network, among other data configuration and processing functions.

The system is controlled live from a computer through the industrial WLAN, and for development purposes, the robot platform is manually controlled with a remote joystick. The onboard sensing devices stream the data via the ROS middleware; where predefined data types are used for internal communication.

A. Electronic Nose

For the task of gas anomaly detection, we developed a *electronic nose* (E-Nose). Its working principle is based on the integration of different gas sensors. We employ three nonselective metal oxide semiconductor (MOX) sensors and, in combination with optical (nondispersive IR and laser scattering) technologies, we measure air contaminants, methane, CO2, and flammable gases. The second category of gas

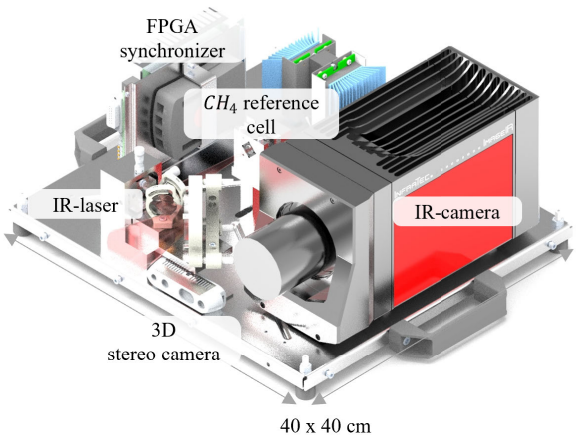


Fig. 3. Setup of the active gas camera. A detailed description of its working principle can be found in [37].

sensors, based on electrochemical principles is specific and fast in their response. These data allowed the training of a supervised learning network able to identify gas signatures. In addition, relevant information, such as humidity, wind speed, and direction, was logged.

B. Active IR Gas Camera

We developed an active IR gas camera (IRcam) for the sensitive detection of methane leaks and the estimation of gas concentration and flow rate. The light detection is carried out by an ImageIR 8300 camera from Infratec, while its field of view is illuminated by a 25 mW tunable ~3260 nm single-mode distributed feedback (DFB)-interband cascade laser from nanoplus.

Three consecutive images are recorded, while the laser's wavelength is tuned over a methane absorption line. We process the image batches as described in [37] using the Tunable Diode Laser Absorption Spectroscopy (TDLAS) to obtain concentration length information. Thereby, the synchronization between laser tuning and camera trigger is achieved with a redpitaya STEMlab 125-10 SBC with a field programmable gate array (FPGA) on board. To ensure wavelength stability across different temperatures and to handle the aging effects of the laser diode, a reference gas cell is filled with methane. Less than 1% of the laser light is coupled out of the diverging laser beam and fed through the cell, as shown in Fig. 3. We feed the gas-image stream at 5 Hz to a 3-D convolutional neural network (CNN) trained to estimate the flow rate of the gas leak in real time. Due to the nature of the TDLAS-based gas detection, we gain more precise concentration length information than state-of-the-art gas cameras, which leads to more accurate flow rate predictions [42], [43].

C. UV Camera

We developed a remote fluorescence detection system for the recognition of oil films on surfaces and potential oil rests. This system consists of a 365 nm ultraviolet (UV) light-emitting diode (LED), ODS75 Smart Vision Lights, and a CMOS camera, OpenMV Cam H7R2. Pairs of images are taken by the camera: one image with ambient light and one with additional UV illumination. The difference between these two frames yields an image containing UV-excited fluorescence to reveal information about oil films due to their aromatic molecular structures [44]. While this sensor is able to detect oil films up to a distance of 1 m in indoor lighting conditions, the sensitivity decreases outdoors, in particular in sunny conditions.

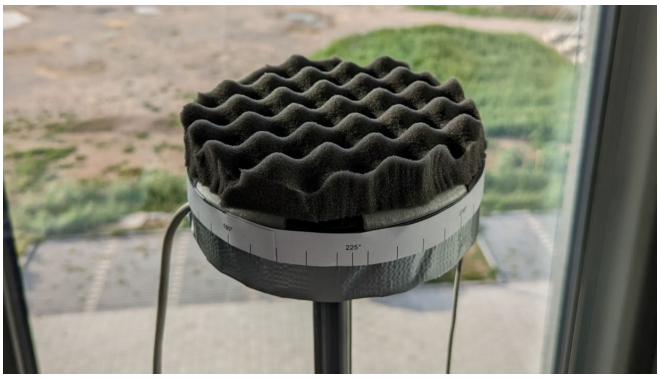


Fig. 4. UCA microphone array with a windscreen for outdoor usage. Details of the employed algorithms can be found in [45].

D. Microphone Array

To provide the robot with hearing capabilities, we developed a five-microphone uniform circular array, as shown in Fig. 4. We employ TDK ICS-40720 microelectromechanical systems (MEMS) microphones with a sampling rate of 96 kHz, and an intermicrophone spacing of ≈8 cm. The setup is capable of localizing sound sources with a maximum frequency of around 2.1 kHz using a specifically designed method for direction-of-arrival (DoA) estimation.

The algorithms under consideration have been thoroughly examined in a laboratory setting, as outlined in [45]. These algorithms, originating from classical subspace DoA estimation techniques such as multiple signal classification (MUSIC), have been expanded to accommodate the UCA configuration. Recognizing that subspace methods are traditionally formulated for narrowband estimation, whereas wideband signals are expected (as detailed in a previous publication), these algorithms have been further adapted to support an arbitrary frequency range, extending up to the spatial Nyquist frequency. The resulting algorithms comprise a two-step process, as depicted in Fig. 5. Initially, a computationally inexpensive coarse frequency-domain frequency-invariant beamformer (FDFIB) guess step [46] is implemented, followed by a more resource-intensive fine-search step within the coherent subspace (CSS) MUSIC spectrum [47]. This hybrid approach ensures an optimal balance between computational efficiency and accuracy. Notably, the algorithms exhibit the capability to resolve multiple sources, facilitated by a selection criterion applied to the singular values of the spectral covariance matrix. However, it is important to note that the algorithm's performance is constrained by the number of available microphones in the array.

We detect sound anomalies by recognizing sound samples as normal or abnormal through a supervised learning model. The network is trained on the MIMII [48] dataset and can be extended to include sound samples from the plant where the robot is operated. The anomalies, constituted by unknown

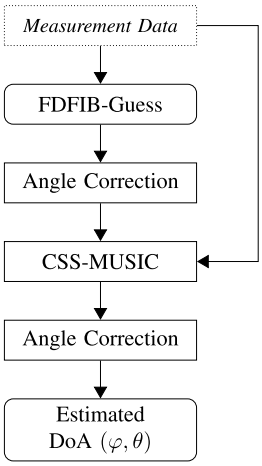


Fig. 5. Sequential steps of the DoA estimation algorithm, starting with a coarse but computationally efficient guess and progressing to a more detailed peak search in the wideband MUSIC spectrum for refined accuracy.

sound samples, are detected using a deep neural network with an encoder–decoder architecture. To this end, the network learns features of the sound samples by encoding the sample into a latent vector embedding and then decoding the vector to recreate the input sound sample. For unknown sounds, the input cannot be reconstructed, which indicates an anomaly.

E. LiDAR Mapping System

We developed a task-specific LiDAR setup to build accurate 3-D maps of the environment of the robot. We, furthermore, map the deep sewage channels at the plant. The unit comprises two Velodyne VLP16 laser scanners mounted on a cuboid structure. While the top laser is leveled with the ground to capture the distant surroundings, the front laser scanner is steeply inclined, allowing it to capture the ground and possible channels in front of the robot. We employ an inertial measurement unit (IMU) to track the short-term motion of the system, increasing mapping accuracy and robustness. A dense 3-D map of the environment is obtained by fusing the continuously captured data with a 3-D SLAM approach when the robot is in motion. The resulting map of the environment is represented as a point cloud in 3-D space. With this generic representation, various applications are realized, such as embedding other sensor information (e.g., temperature), geometric alignment for inventory purposes, changes or anomalies detection, and, as future work, semantic segmentation with deep learning techniques.

F. Passive Cameras

The optical perception system includes three forward-facing cameras composed of a stereo rig of two RGB cameras, Blackfly S GigE FLIR, with a resolution of 1544 $\times$ 2064, and a thermal camera, and FLIR Boson long wave IR (LWIR), with 512 $\times$ 640 pixels and a spectral range of 7.5–13.5 μm. These three cameras are synchronized with an external trigger at a frame rate of 30 frames/s.

The live object detection module employs the images from the RGB cameras to support the navigation functions. It detects dynamic agents in the scene, such as cars and pedestrians, and is based on CenterNet [49], a network for object detection. CenterNet is a fully supervised model that first predicts the object centers and their corresponding center offsets and then combines these values to estimate bounding boxes around dynamic objects.

G. Mobile Robot Navigation System

We employ the Clearpath Husky A200 as the mobile robotic platform since it is a rugged, all-terrain robot equipped with an onboard computer that is designed to operate in indoor and outdoor environments. The robot achieves a safe maximum speed of 1 m/s in forward movement with the sensor arrangement weighing about 35 kg. It also supplies electrical power to most of the sensors, enabling an untethered performance in a self-contained fashion. The onboard computer gets connected to the wireless industrial network, from which a remote computer accesses and controls the system through the human–machine interface over Robot Operating System (ROS).

To enable autonomous navigation, we use an OS128 Ouster LiDAR. It has a vertical field of view of 90°, 50 m range, 128 times 1024 channels at a 20 Hz rate, and a built-in IMU. The navigation functionality consists of the tasks of localization, mapping, and path planning.

We solved the 2-D SLAM task with an efficient Rao-Blackwellized particle filter (RBPF) that creates grid maps [50], while the localization in a 2-D grid map builds on adaptive Monte Carlo localization (AMCL) [51] using an RBPF. To this end, the 3-D point clouds from the Ouster LiDAR were projected into 2-D range scans.

The path planning task involves two hierarchized planners. While we solve the global planning using the A^* [52] approach, we employ the timed-elastic-bands (TEB) [53] approach for local planning.

The choice of TEB for our system over the traditional dynamic window approach (DWA) [54] arises from observed drawbacks with the latter. First, DWA presented suboptimal performance due to its inability to execute reversal motions, a consequence of constant control actions along the prediction horizon. In addition, on-site rotations required an extra tuning step which was empirically determined. In contrast, TEB demonstrated significant improvement in the controller performance as a result of its formulation, which prioritizes time-optimal solutions. TEB exceled in cases of navigation around obstacles, which required temporary deviations from the path. Although experimental versions of the planner incorporate dynamic obstacle avoidance, its performance relies heavily on obstacle tracking and the accuracy of its state estimation. To support these aspects, we complement the planner with a reinforcement learning (RL) algorithm.

Such an algorithm uses the dynamic obstacles detected in the RGB images and projects them in the bird's-eye view (BEV) using the Ouster LiDAR data. We combine features extracted from this local semantic map via a CNN with the robot's low-dimensional internal state, such as the navigation goal and recent controls.

IV. Experiments

The experiments were conducted at the wastewater treatment plant within the chemical facility of Marl, Germany. This industrial setting offered a ground-level route encompassing

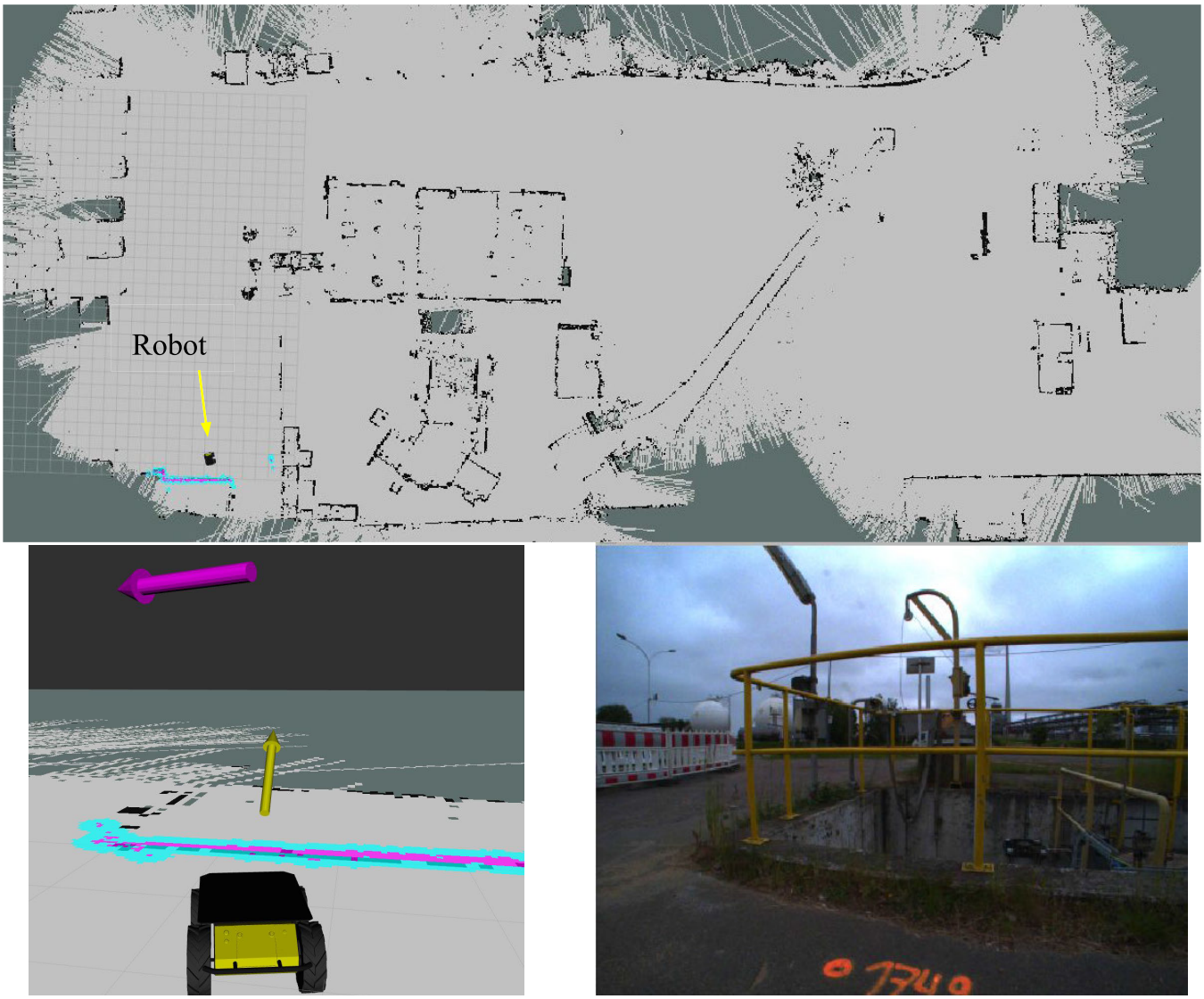


Fig. 6. Visualization of the robot in the map and markers for wind and sound direction. The three images depict the same instance of the robot navigating the test site in real life. Top: Full 2-D occupancy grid map of the test area, shown in grey and black from a top-down perspective. The position of the real robot is shown on the map with its 3-D model. Bottom left: Perspective rear view of the 3-D model robot in the visualizer, in the same position as the upper figure. Its navigation on the map is enabled by the cost map (in cyan and magenta), which is aligned with the map. Simultaneously, the purple arrow indicates the wind direction (data obtained from the e-nose), and the yellow arrow indicates the audio direction (data from the microphone array). Bottom right: Image from the RGB camera on the right side of the robot. The robot is facing a fence, which is recognized in the map and consequently creates the costmap. This supports localization in the map.

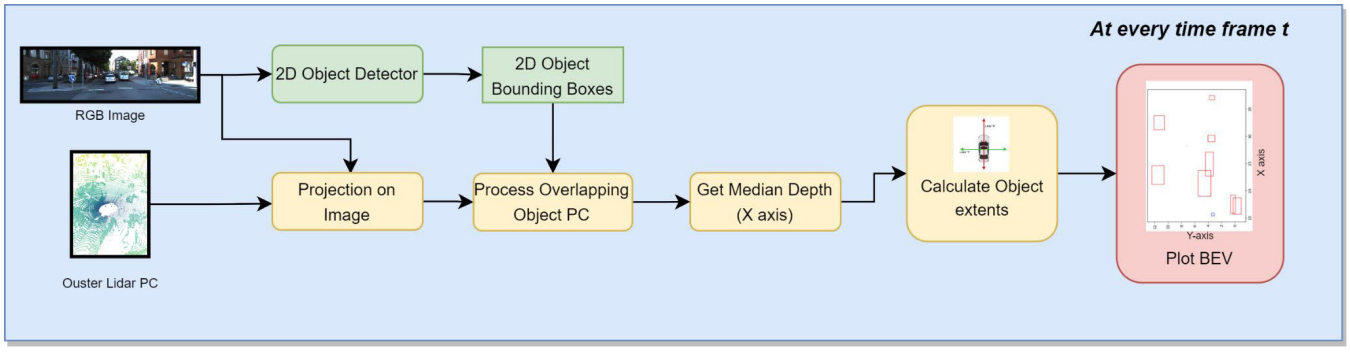


Fig. 7. Pipeline of object detection and its 3-D position estimation for robot navigation in a dynamic environment.

diverse points of interest, including tasks involving hot pipes, noisy pumps, and sewage canals with occasional variations in gas levels. The discussion in this section is divided into three main parts. First, the focus is on autonomous navigation capabilities. The presentation includes a detailed discussion of the experiments conducted to evaluate the robot's ability to navigate autonomously in this complex industrial environment. Moving on to the second part, the section delves into gas

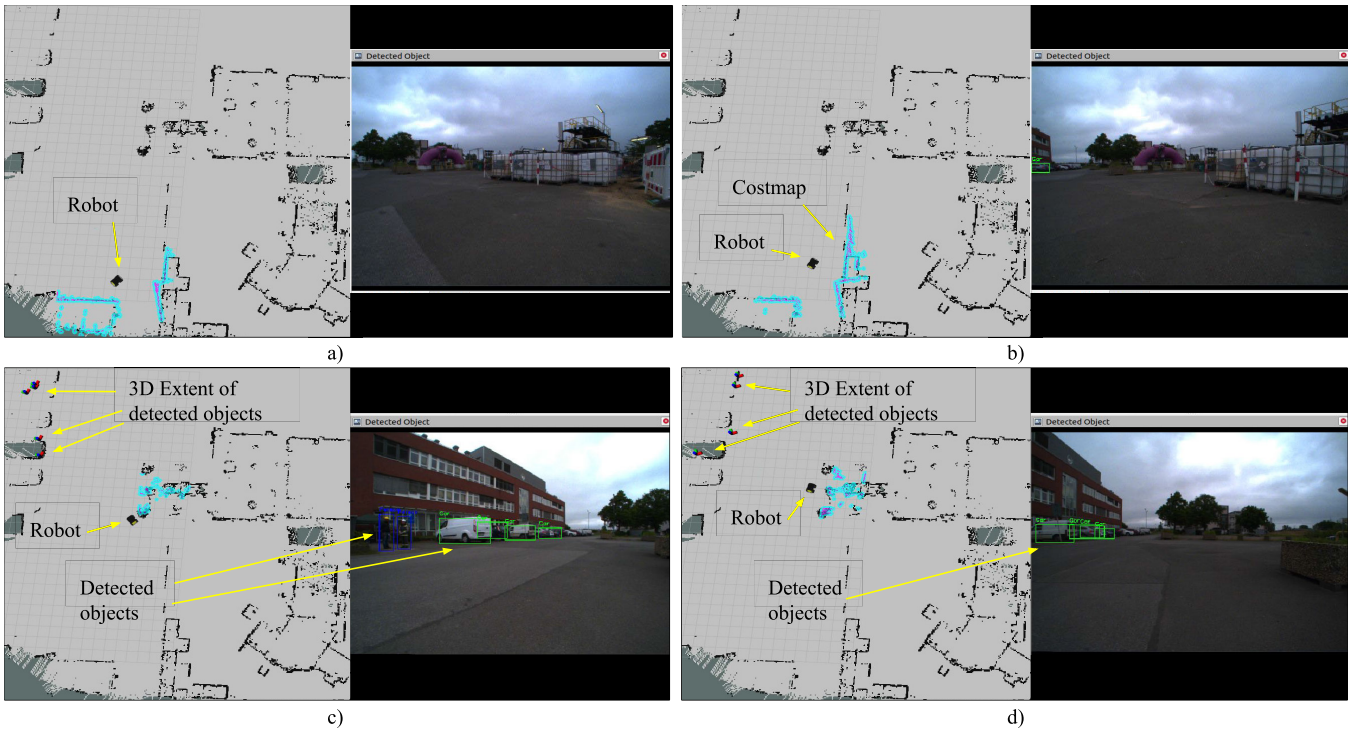


Fig. 8. Visualization from a top-view perspective of the robot navigating the test area and detecting objects. The order of the images corresponds to successive time points following the alphabetical order, with (a) being the initial time point, followed by (b)–(d). Each image shows the real-world localization of the robot, with the robot model superimposed on the 2-D map, its localization (costmap matching the map), and the image taken by the right RGB camera on the real robot. Top left: First position of the robot in a north-easterly direction. The accompanying RGB image shows some white industrial containers, a white fence, and a purple pipe in the background. While the proximity of the containers and the fence contributes to the localization of the robot by the costmap, they remain unidentified by the object-detection module. Top right: Second position of the robot. As the robot has moved to the north-east, the costmap (colored in cyan) accompanies the movement and by matching the 2-D map stabilizes the localization. The RGB image shows a closer look at the pink pipe, together with the first car detected in a small green box. Bottom left: Third position of the robot, where it has moved in a north-westerly direction, losing the purple pipe from its field of view. The robot is now facing a parking lot where people are also present. The detected objects are classified as people (in blue boxes) or cars (in green boxes) in the RGB image and their 3-D extents are projected onto 2-D map with small markers. Bottom right: Final position of the robot, after a slight north-east movement from the previous position. The cars are still in the robot's field of view, so their position is still shown on the 2-D map. The fact that the costmap matches the 2-D map means that the robot is still correctly localized.

sensing experiments. Special attention is given to the outcomes obtained from the IR camera, providing a comprehensive analysis of the robot's performance in tasks related to gas detection within the plant. In the final part of this section, the focus shifts to acoustic experiments. The initial set of experiments involves the localization of pump equipment using acoustic signals, followed by experiments on the detection and localization of acoustic leaks. The discussion concludes with the introduction of a more general approach to anomaly detection using acoustic data.

A. Navigation Experiments

Robot information, related to position and localization (Odometry), was computed through probabilistic approaches using the navigating LiDAR data. This point cloud was used in the tasks of mapping, localization, and global/local path generation through the Dijkstra algorithm and the DWA, respectively. The 2-D mapping capabilities were compared across resolutions from 0.01 to 0.2 m, which had a direct effect on the map size and update speed.

The robot was tasked with navigation missions and a visualization interface from a remote station, allowed to know relevant live information from the robot and the sensors. This is displayed in Fig. 6 with a representation of the real robot on a map alongside indicators sensor data. In the figure, the 2-D occupancy grid map of the complete test site is displayed. When the "real" robot is navigating the site, its position is signed on the map by a 3-D model of the robotic platform. As the real robot changes its position in the test site, the 3-D model updates its localization through the costmap computed with the particle filter and the "live" LiDAR data. In the same figure, the three images correspond to the same moment as the robot navigates the real-life test site. From the lower images, a closer look at the perspective rear view of the 3-D model robot in the visualizer shows the correspondence of the costmap with the 2-D map, allowing a correct localization and consequently reliable navigation. Finally, an inspection of the "live" data transmitted by the RGB camera lets us know that the "real" robot is in front of a fence, acting as a landmark for the 2-D map and enhancing the localization. In addition, diverse RL agents were trained using the soft actor-critic [55] algorithm with varying hyperparameters to compare the ability to navigate around dynamic objects in simulations and real-world scenarios predicatively.

To achieve more robust navigation in environments with dynamic obstacles, an object detection module for obstacle avoidance was developed. This module can recognize car and people categories, as well as determine their 2-D and 3-D extents. This leads to dynamic obstacle recognition and avoidance while navigating. Such a task is divided into two

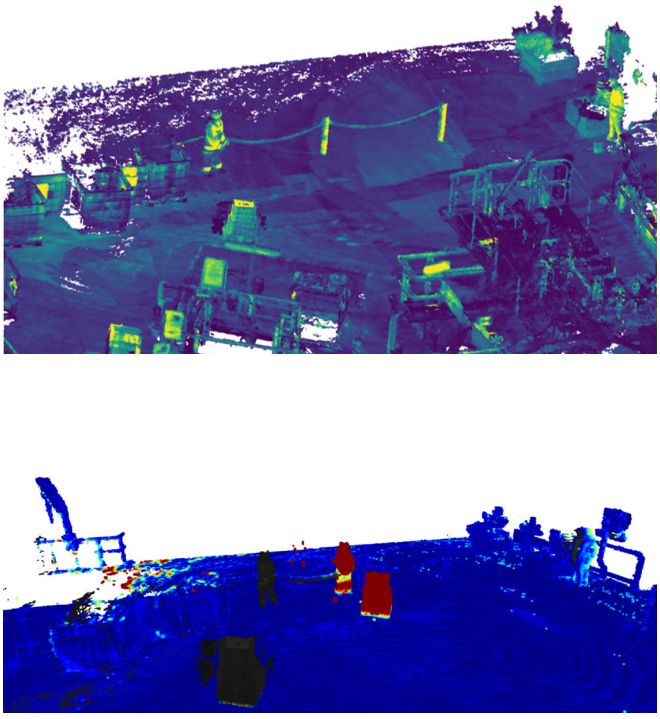


Fig. 9. Reference point cloud built by the Mapping LiDAR Setup. Colorization is based on surface reflectivity for IR light (top). Results of the change detection algorithm. In this experiment, objects and a person were relocated from one location (gray) to a new location (red). Colors represent the distance to the reference mesh (bottom).

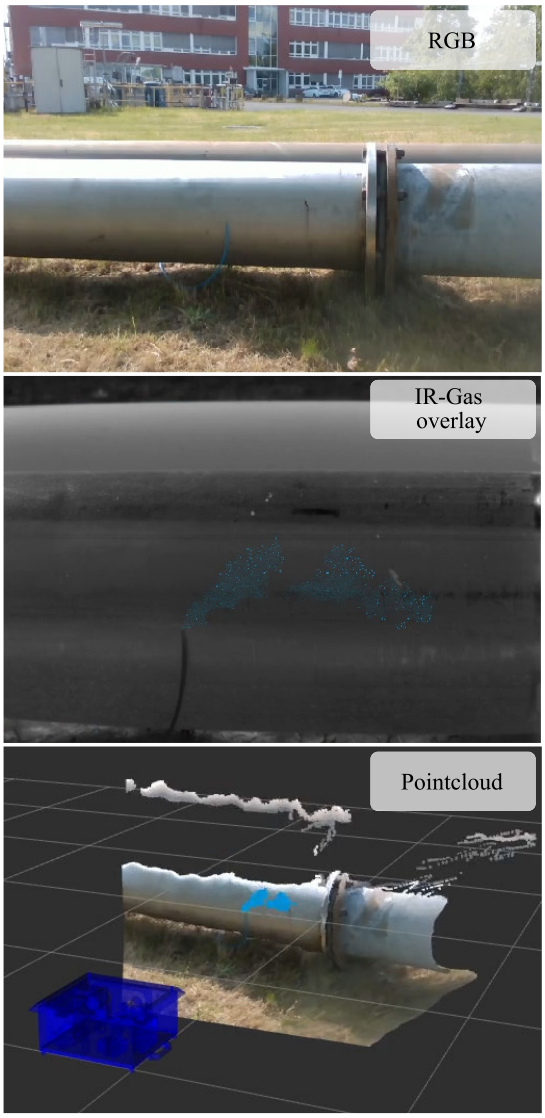


Fig. 10. Different image streams of the active gas camera are shown filming an artificial methane leakage on an industrial plant. The registered point cloud (bottom) is possible through an RGB-Depth-IR calibration. A video of this scene, from which the images are taken, is published together with this work.

subtasks: 1) recognition of specific objects in the scene using an object detection pipeline and 2) projection of the 2-D bounding boxes of the detected objects into the 3-D world.

1) 2-D Object Detection: The first subtask leverages the state-of-the-art object detection framework CenterNet, which is a single-stage object detection framework that accepts an RGB image as input and outputs a center point heatmap while simultaneously regressing the height, width, and offsets for every object in the image. These separate outputs are later fused with a postprocessing step to generate the final 2-D bounding boxes. To train CenterNet for the object recognition task, numerous labeled images were required in principle. To avoid the time-consuming task of manual image annotation for the object detection module training, we leveraged preexisting object detection ground truths from a different domain and then adapted the model to work on the chemical plant dataset. To this end, we pretrained the model using labels from the autonomous driving dataset KITTI [56] comprising 3712 train and 3769 validation samples and then fine-tuned it on 167 manually annotated images from our chemical plant and validating the model on 17 images.

The evaluation of 20 manually annotated images resulted in Intersection over Union (IoU) values and average precision (AP) scores at the confidence threshold of 0.3–0.6855. These values were 0.300 and 0.668 for the Car class and 0.300 and 0.703 for the Pedestrian class, respectively. The accuracy of locating only the objects in the image without considering their width and height predictions is referred to as the AP of Centerpoint matches and obtained a value of 0.7662. Due to the limited train and test datasets available at the time, this is an optimal result for the scope of this task application. The object detection module estimates the dynamic object location in the scene but is not the last step for obstacle avoidance. Since the bounding box covers a major part of the object, this detection is enough given that the LiDAR point projection helps in determining the actual object location. Furthermore, the size of the bounding box gets reduced after the object detection so that no background is later used, which improves the accuracy of the final object estimation for navigation.

2) Estimation of 3-D Position: The second subtask concerns the estimation of the 3-D locations of the detected objects and leverages the depth information obtained from the onboard LiDAR sensor along with the extrinsic calibration between the LiDAR and camera sensors. The LiDAR point cloud is projected onto the RGB image using the cameras' intrinsic parameters coupled with the extrinsic transformation between the two sensors. For every detected object, the LiDAR points outside the 2-D extent of the bounding box are filtered out.

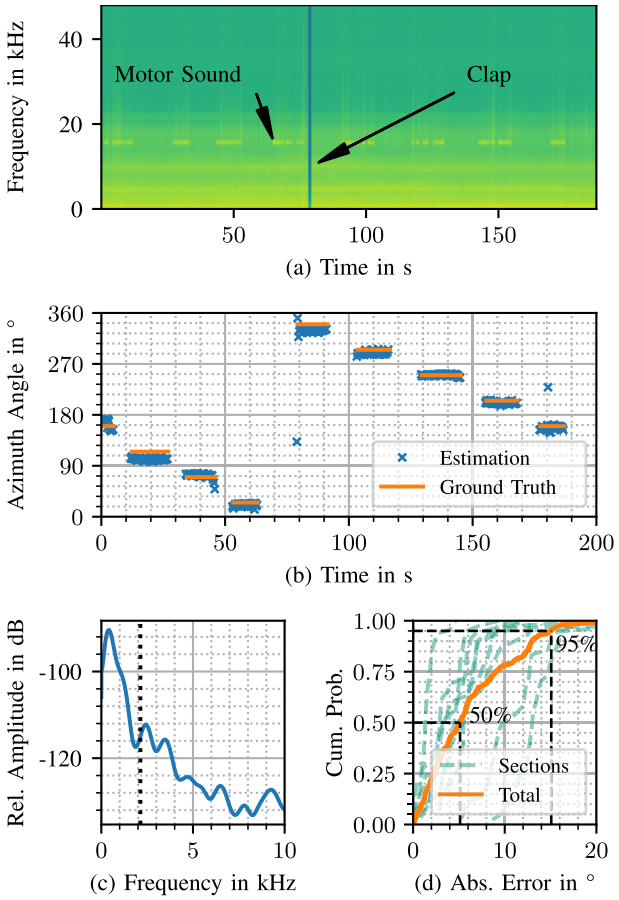


Fig. 11. Static experiment. (a) Features the spectrogram of the recorded signal, highlighting the discernible motor noise just below 20 kHz during robot movement. A distinctive clap performed by the experimenter is denoted in the plot when the robot faces 180° away from the working pump. (b) Evolution of the estimated DoA throughout the experiment run. (c) Frequency power distribution of the sounds emitted by the working pump. In this instance, a significant portion of the frequency content falls below the array's maximum spatial frequency. (d) Cumulative error distribution of the DoA error. The DoA estimation performance demonstrates a median circular error of approximately 5°, with a 95% confidence interval of around 15°.

Afterward, the bounding boxes are shrunk by a factor of 0.7 to alleviate potential noise introduced by incorrect intrinsic or extrinsic calibration matrices, thus improving the robustness of the model. The filtered LiDAR point cloud represents all the points belonging to the object being considered, which is then used to compute the 3-D extent of the object. These 3-D extents are then flattened along the height dimension to determine the extent of the object in the BEV space.

Fig. 7 illustrates the pipeline of the object detection modules and its position estimation. This information is utilized to plan safe collision-free paths around movable obstacles.

Fig. 8 shows sequential captures of the robot navigating the test site in real life in alphabetical order. It is observed that as the robot navigates close to fixed objects at the time the map was created, the robot recognizes the features (such as fences, pipes, and buildings) and the generated costmaps enable the localization. Accordingly, these elements remain unidentified by the object detection module. When a car or a human enters the field of view of the RGB cameras, the system identifies them and obtains their 3-D extent. These data are consequently displayed in the 2-D map with red and blue markers and considered by the path planner to generate a collision-free path around these objects.

B. 3-D Mapping and Change Detection

Detecting and tracking geometric changes in industrial environments are crucial for ensuring safe plant operation. Moreover, it can be the keystone for predictive maintenance. To evaluate the capability of the module for 3-D mapping and change detection, multiple data recordings were obtained with the robot in the chemical plant Marl. The foundation is a highly detailed 3-D point cloud captured by the Mapping LiDAR Setup, as described in Section III-E. This reference model is converted into a continuous surface representation, a triangular mesh. Test point clouds recorded in other runs are then compared to this mesh by computing the nearest distances of all points to the reference mesh. In this way, different clusters of points can be identified that represent geometric change, as illustrated in Fig. 9. The proposed method allows for intuitive anomaly detection and assessment.

C. Gas Detection Experiments

The active gas camera was used to detect artificial gas leaks that we installed on an industrial site of a chemical park in Germany. Fig. 10 shows a $\sim$40 mL min^{-1} methane leak that we place in front of a pipeline. The flow rate was set using a mechanical flowmeter from *Brooks instruments*, which had 40 mL min^{-1} as the lower limit. The leak was filmed from 2 m distance. The point cloud is achieved by a geometric calibration of the stereo camera to the IR camera. This allows projecting the source of the gas leak to the depth frame and projecting it at the correct distance in the point cloud. The gas overlay on the IR image is scaled such that stronger occupancy relates to a higher methane concentration length. The scene is also available as a video[1] published together with this article. The red-masked objects are detected as moving objects and are not considered to be gas information. A description of how these algorithms work can be found in [37].

Methane leaks of $\sim$40 mL min^{-1} could be visualized, while this limit strongly depends on the wind conditions. It should be noted that the active approach is challenged by the wind as it uses subsequent images for the concentration calculation, in which the wind disturbs the static scene. An improvement of the camera during this project resulted in stable methane detection up to wind speed of $\sim$2 m s^{-1} by minimizing the relevant times in between the recorded subsequent frames. Apart from hyperspectral imaging that is typically significantly more expensive, to the best of our knowledge, this is the first reported gas camera to visualize such small leaks under real-world conditions in real-time.

D. Acoustic Experiments

The acoustic experiments are structured into three distinct parts. Initially, a static phase is implemented to assess the Direction-of-Arrival (DoA) estimation accuracy using the sound emitted by pumps. This phase serves to evaluate the precision of the DoA estimator under controlled conditions.

[1]https://dx.doi.org/10.21227/4ms5-rs57

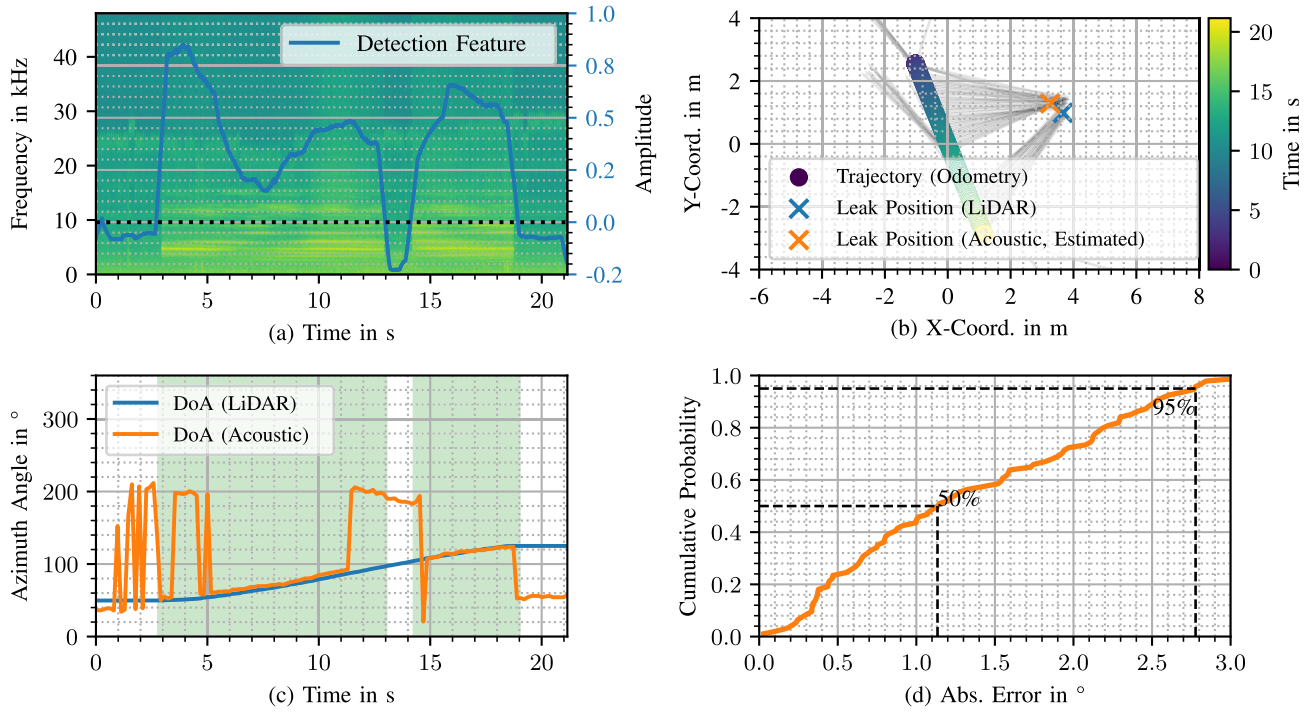


Fig. 12. Dynamic experiment. (a) Spectrogram of the recording signal, incorporating the played-back sound of a leak. Overlayed on the spectrogram is the crafted detection feature designed for the leak signal. (b) Trajectory of the robot during the experiment, with the measured leak positions obtained from both LiDAR and acoustic estimation. The single DoA measurements are represented by thin black lines. The position of the leak can be performed by acoustic means to an error of around 0.5 m to the LiDAR measurement. (c) Comparison between the estimated acoustic DoA values and the LiDAR-measured DoA values. Larger deviations happen when another sound source is interfering with the measurement. Shading highlights regions with a positive detection feature. (d) Cumulative error distribution of the absolute circular error. The system demonstrates the capability to estimate with a median accuracy of 1.13° (or 2.8° at the 95th percentile).

Following the static phase, a dynamic experiment is executed, involving the robot being driven into the proximity of the pumps. The objective of this phase is to evaluate the source detection capabilities in a more dynamic setting, considering the robot's movement in relation to the pumps. In the final part of the experiments, the focus shifts to anomaly detection. Various experiments are conducted to assess the system's ability to detect anomalies in the acoustic environment. This phase aims to evaluate the robustness and effectiveness of the anomaly detection mechanisms implemented in the system.

1) Static Experiments (Pump Equipment Localization): In this experiment, the robot is positioned in front of an operational pump. Subsequently, the robot undergoes rotational movement by a specific degree in each step, as illustrated in Fig. 11. The analysis reveals a median circular error of approximately 5°, accompanied by a 95% confidence interval of around 15°. The evaluation of the cumulative distribution function (cdf) for individual sections indicates no significant deviation. Therefore, it can be inferred that the DoA estimation performance remains isotropic across the azimuthal range.

2) Dynamic Experiment (Leak Localization): A sound speaker is positioned near a pipe, emitting a recorded sound mimicking a pipe leak. The robot follows a linear trajectory [refer to Fig. 12(b)] during the experiment, and due to the recording's brevity, a brief interruption occurs. To create a detection feature, a simple inner product detection method is employed based on the frequency distribution of the leak signal. Using the LiDAR sensor, the position of the speaker is established as a reference. Since a substantial portion of the leak signal surpasses the array's maximum spatial frequency, conventional DoA estimation methods discussed earlier are not applicable. Instead, the well-known steered response power phase transform (SRP-PHAT) algorithm [57] is employed for DoA estimation in this experiment. With the LiDAR reference, a DoA estimation accuracy of approximately 1.13° (or 2.8° at the 95th percentile) is achieved, considering only the estimated values falling within regions where the detection feature exceeds 0. Combining the odometry-based position information from the robot with the estimated DoA values, the leak's position is determined. This localization is executed by intersecting various pairs of generated lines, and the median of these intersections yields the leak's position. This estimated position is within 0.5 m of the LiDAR measurement.

3) Anomaly Detection: Two strategies were tested to detect sound anomalies, namely, reconstruction error, and latent space distance. In the reconstruction error strategy, the network tries to recreate the input sound signal using the intermediate latent vector. Since the network is solely trained on normal samples, it can recreate (decode) normal samples but fails to recreate an anomalous sound sample. The reconstruction error between the input and output samples is compared against a threshold value and used to differentiate between a normal and an anomalous sound sample. The latter strategy relies on the fact that latent vectors of similar objects are clustered near one another in the latent space; hence, the latent vector of an anomalous sample gets placed far from the latent vectors of the normal samples. An anomaly is detected when the distance

between the latent vector of the current sample and the mean of all latent vectors from training is above a predefined threshold.

The encoder–decoder model was initially trained on the MIMII [48] dataset, which comprises numerous sequences of normal and anomalous sound samples. From them, 4800 samples were used for training and 447 for validation and 1524 for testing. The reconstruction error-based strategy achieved an accuracy of 69.62% on the test set, while the latent distance-based approach achieved an accuracy of 82.28% on the test set. In addition, the model was further trained to include sound samples from the Marl chemical plant, adding splits of 32 train, six validation, and seven test samples. With this extended dataset, the model achieved an accuracy of 84% using the reconstruction error-based approach and 100% using the latent distance-based approach. In the field, anomalies were simulated by playing "broken" sounds near the pumps.

V. Conclusion and Outlook

This article introduced a robotic system designed for the comprehensive inspection of an industrial chemical plant. The system encompasses various key features, including navigation capabilities, and 2-D object detection with a demonstrated general AP of 0.7662, covering objects such as cars and pedestrians. Although we acknowledge that the dataset may seem limited, the collection of additional data by sampling in industrial settings, or by implementing data augmentation techniques is left as a future endeavor. Such plans may improve the robustness of the methods. The AP of 2-D object detection can be improved by annotating more data or using more state-of-the-art object detection networks. More recently, it can also use foundation models such as DINOv2 [58] to get rich object features and then fine-tune the model on a small subset of data. Through extrinsic calibration of multiple optical sensors, the system is proven to estimate the 3-D position of detected objects, which, in turn, is employed for planning collision-free paths. In addition, this article highlights the sensitivity and the real-world capability of the active gas camera. It visualizes an artificial methane leak of only $40\,\mathrm{mL\,min^{-1}}$ in real time and successfully estimates its 3-D position by combining depth information of the stereo camera up to 2 m distance. The final aspect presented is the elucidation of the working principles behind the acoustic pump equipment and leak localization. This section demonstrates the system's capability to localize a leak with an absolute error of around 0.5 m, while the robot is in motion.

Routine inspection rounds are mandatory in many companies in process industries. The frequency and duration of these field rounds depend on the potential hazard of the plant, and they are conducted 24/7. The reduced availability of qualified candidates for such jobs, the desire to increase their appeal, and the need to collect reliable and objective sensor data fuel the development and testing of robots that can take over these tasks in diverse companies with industrial processes. Whereas navigation with legged robots in such plants seems already feasible, further development of adequate sensors and algorithms for the evaluation and interpretation of their signals is still necessary. Furthermore, the ability of humans to depart from routine rounds when detecting an incident in peripheral vision is an open issue that needs to be transferred to such robots.

References

[1] E. Guizzo, "By leaps and bounds: An exclusive look at how Boston dynamics is redefining robot agility," *IEEE Spectr.*, vol. 56, no. 12, pp. 34–39, Dec. 2019.

[2] A. Francis, S. Li, C. Griffiths, and J. Sienz, "Gas source localization and mapping with mobile robots: A review," *J. Field Robot.*, vol. 39, no. 8, pp. 1341–1373, Dec. 2022.

[3] N. Vödisch, D. Cattaneo, W. Burgard, and A. Valada, "Continual SLAM: Beyond lifelong simultaneous localization and mapping through continual learning," in *Proc. Int. Symp. Robot. Res.* Geneva, Switzerland: Springer, 2022, pp. 19–35.

[4] J. Arce, N. Vödisch, D. Cattaneo, W. Burgard, and A. Valada, "PADLoC: LiDAR-based deep loop closure detection and registration using panoptic attention," *IEEE Robot. Autom. Lett.*, vol. 8, no. 3, pp. 1319–1326, Mar. 2023.

[5] D. Sorrenti, D. Cattaneo, and V. Abhinav, "CMRNet++: Map and camera agnostic monocular visual localization in LiDAR maps," in *Proc. ICRA Workshop Emerg. Learn. Algorithmic Methods Data Assoc. Robot.*, 2020, pp. 1–5.

[6] A. Younes, D. Honerkamp, T. Welschehold, and A. Valada, "Catch me if you hear me: Audio-visual navigation in complex unmapped environments with moving sounds," *IEEE Robot. Autom. Lett.*, vol. 8, no. 2, pp. 928–935, Feb. 2023.

[7] D. A. Gómez-Rosal et al., "A smart robotic system for industrial plant supervision," in *Proc. IEEE SENSORS*, Oct. 2023, pp. 1–4.

[8] A. M. Zanchettin, N. M. Ceriani, P. Rocco, H. Ding, and B. Matthias, "Safety in human-robot collaborative manufacturing environments: Metrics and control," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 2, pp. 882–893, Apr. 2016.

[9] E. A. Oyekanlu et al., "A review of recent advances in automated guided vehicle technologies: Integration challenges and research areas for 5G-based smart manufacturing applications," *IEEE Access*, vol. 8, pp. 202312–202353, 2020.

[10] M. Edwards, "Robots in industry: An overview," *Appl. Ergonom.*, vol. 15, no. 1, pp. 45–53, Mar. 1984. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003687084901212

[11] M. Shneier and R. Bostelman, "Literature review of mobile robots for manufacturing," Nat. Inst. Standards Technol., Gaithersburg, MD, USA, Tech. Rep. ISTIR 8022, 2015.

[12] O. Duran, K. Althoefer, and L. D. Seneviratne, "State of the art in sensor technologies for sewer inspection," *IEEE Sensors J.*, vol. 2, no. 2, pp. 73–81, Apr. 2002.

[13] M. Karkoub, O. Bouhali, and A. Sheharyar, "Gas pipeline inspection using autonomous robots with omni-directional cameras," *IEEE Sensors J.*, vol. 21, no. 14, pp. 15544–15553, Jul. 2021.

[14] D. Zhang, J. Cao, G. Dobie, and C. MacLeod, "A framework of using customized LiDAR to localize robot for nuclear reactor inspections," *IEEE Sensors J.*, vol. 22, no. 6, pp. 5352–5359, Mar. 2022.

[15] G. Dobie, R. Summan, S. G. Pierce, W. Galbraith, and G. Hayward, "A noncontact ultrasonic platform for structural inspection," *IEEE Sensors J.*, vol. 11, no. 10, pp. 2458–2468, Oct. 2011.

[16] P. Ambati, K. M. S. Raj, and A. Joshuva, "A review on pipeline inspection robot," in *Proc. 3RD Int. Conf. FRONTIERS AUTOMOBILE Mech. Eng. (FAM)*, 2020, Art. no. 060002, doi: 10.1063/5.0033998.

[17] S. Yamamoto, "Development of inspection robot for nuclear power plant," in *Proc. IEEE Int. Conf. Robot. Automat.*, vol. 2, Jan. 1992, pp. 1559–1566.

[18] X. G. Fu, G. Z. Yan, B. Yan, and H. Liu, "A new robot system for auto-inspection of intersected welds of pipes used in nuclear power stations," *Int. J. Adv. Manuf. Technol.*, vol. 28, nos. 5–6, pp. 596–601, Mar. 2006.

[19] B. Jiang, A. Sample, R. Wistort, and A. Mamishev, "Autonomous robotic monitoring of underground cable systems," in *Proc. 12th Int. Conf. Adv. Robot.*, 2005, pp. 673–679.

[20] I. Muhammad, K. Ying, M. Nithish, J. Xin, Z. Xinge, and C. C. Cheah, "Robot-assisted object detection for construction automation: Data and information-driven approach," *IEEE/ASME Trans. Mechatronics*, vol. 26, no. 6, pp. 2845–2856, Dec. 2021.

[21] R. Almadhoun, T. Taha, L. Seneviratne, J. Dias, and G. Cai, "A survey on inspecting structures using robotic systems," *Int. J. Adv. Robotic Syst.*, vol. 13, no. 6, Dec. 2016, Art. no. 172988141666366.

[22] X. Wang, T. Luo, and Y. Lei, "Inspection robot design for underwater structure in complex flow environment," in *Proc. Int. Conf. Service Robot. (ICoSR)*, Jun. 2022, pp. 193–197.

[23] B. Nouri Rahmat Abadi et al., "CARMA II: A ground vehicle for autonomous surveying of alpha, beta and gamma radiation," *Frontiers Robot. AI*, vol. 10, Mar. 2023, Art. no. 1137750.

[24] D. Mitchell et al., "Lessons learned: Symbiotic autonomous robot ecosystem for nuclear environments," *IET Cyber-Syst. Robot.*, vol. 5, no. 4, Dec. 2023, Art. no. e12103.

[25] H. Ishida, Y. Wada, and H. Matsukura, "Chemical sensing in robotic applications: A review," *IEEE Sensors J.*, vol. 12, no. 11, pp. 3163–3173, Nov. 2012.

[26] L. Yu et al., "Inspection robots in oil and gas industry: A review of current solutions and future trends," in *Proc. 25th Int. Conf. Autom. Comput. (ICAC)*, Sep. 2019, pp. 1–6.

[27] K. A. Kas and G. K. Johnson, "Using unmanned aerial vehicles and robotics in hazardous locations safely," *Process Saf. Prog.*, vol. 39, no. 1, Mar. 2020, Art. no. e12066.

[28] H. Kivrak, M. Z. Karakusak, S. Watson, and B. Lennox, "Cyber–physical system architecture of autonomous robot ecosystem for industrial asset monitoring," *Comput. Commun.*, vol. 218, pp. 72–84, Mar. 2024.

[29] C. Rascon and I. Meza, "Localization of sound sources in robotics: A review," *Robot. Auto. Syst.*, vol. 96, pp. 184–210, Oct. 2017.

[30] K. Nakamura, K. Nakadai, and G. Ince, "Real-time super-resolution sound source localization for robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2012, pp. 694–699.

[31] G. Fischer et al., "Multimodal indoor localization: Fusion possibilities of ultrasonic and Bluetooth low-energy data," *IEEE Sensors J.*, vol. 22, no. 6, pp. 5857–5868, Mar. 2022.

[32] J.-M. Valin, F. Michaud, J. Rouat, and D. Letourneau, "Robust sound source localization using a microphone array on a mobile robot," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2003, pp. 1228–1233.

[33] A. Schenck, W. Daems, and J. Steckel, "AirleakSLAM: Detection of pressurized air leaks using passive ultrasonic sensors," in *Proc. IEEE SENSORS*, Oct. 2019, pp. 1–4.

[34] F. Wang, W. Lin, Z. Liu, S. Wu, and X. Qiu, "Pipeline leak detection by using time-domain statistical features," *IEEE Sensors J.*, vol. 17, no. 19, pp. 6431–6442, Oct. 2017.

[35] Y. Yan, Y. Shen, X. Cui, and Y. Hu, "Localization of multiple leak sources using acoustic emission sensors based on MUSIC algorithm and wavelet packet analysis," *IEEE Sensors J.*, vol. 18, no. 23, pp. 9812–9820, Dec. 2018.

[36] A. P. Ravikumar, J. Wang, and A. R. Brandt, "Are optical gas imaging technologies effective for methane leak detection?" *Environ. Sci. Technol.*, vol. 51, no. 1, pp. 718–724, 2017.

[37] M. Bergau, T. Strahl, B. Scherer, and J. Wöllenstein, "Real-time active-gas imaging of small gas leaks," *J. Sensors Sensor Syst.*, vol. 12, no. 1, pp. 61–68, Feb. 2023. [Online]. Available: https://jsss.copernicus.org/articles/12/61/2023/

[38] T. Strahl, J. Herbst, A. Lambrecht, E. Maier, J. Steinebrunner, and J. Wöllenstein, "Methane leak detection by tunable laser spectroscopy and mid-infrared imaging," *Appl. Opt.*, vol. 60, no. 15, May 2021. [Online]. Available: https://www.osapublishing.org/ao/abstract.cfm?doi=10.1364/AO.419942

[39] K. J. Nutt, N. Hempler, G. T. Maker, G. P. A. Malcolm, M. J. Padgett, and G. M. Gibson, "Developing a portable gas imaging camera using highly tunable active-illumination and computer vision," *Opt. Exp.*, vol. 28, no. 13, p. 18566, Jun. 2020. [Online]. Available: https://www.osapublishing.org/abstract.cfm?URI=oe-28-13-18566

[40] J. Titchener et al., "Single photon LiDAR gas imagers for practical and widespread continuous methane monitoring," *Appl. Energy*, vol. 306, Jan. 2022, Art. no. 118086. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0306261921013714

[41] M. Quigley et al., "ROS: An open-source robot operating system," in *Proc. ICRA Workshop Open Source Software*, vol. 3, 2009, p. 5.

[42] M. Bergau, T. Strahl, K. Ludlum, B. Scherer, and J. Wöllenstein, "Flow rate quantification of small methane leaks using laser spectroscopy and deep learning," *Process Saf. Environ. Protection*, vol. 182, pp. 752–759, Nov. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957582023010637

[43] J. Wang, J. Ji, A. P. Ravikumar, S. Savarese, and A. R. Brandt, "VideoGasNet: Deep learning for natural gas methane leak classification using an infrared camera," *Energy*, vol. 238, Jan. 2022, Art. no. 121516. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0360544221017643

[44] L. Stasiuk and L. Snowdon, "Fluorescence micro-spectrometry of synthetic and natural hydrocarbon fluid inclusions: Crude oil chemistry, density and application to petroleum migration," *Appl. Geochem.*, vol. 12, no. 3, pp. 229–241, 1997.

[45] G. Fischer et al., "Localization of acoustic gas leakage sources with a circular microphone array," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, 2021, pp. 1–7.

[46] T. Do-Hong, F. Demmel, and P. Russer, "A method for wide-band direction-of-arrival estimation using frequency-domain frequency-invariant beamformers," in *IEEE Antennas Propag. Soc. Int. Symp. Dig. Held Conjunct, USNC/CNC/URSI North Amer. Radio Sci. Meeting*, Jun. 2003, pp. 244–247.

[47] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 4, pp. 823–831, Aug. 1985.

[48] H. Purohit et al., "MIMII dataset: Sound dataset for malfunctioning industrial machine investigation and inspection," 2019, *arXiv:1909.09347*.

[49] K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang, and Q. Tian, "CenterNet: Keypoint triplets for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2019, pp. 6569–6578.

[50] G. Grisetti, C. Stachniss, and W. Burgard, "Improving grid-based SLAM with rao-blackwellized particle filters by adaptive proposals and selective resampling," in *Proc. IEEE Int. Conf. Robot. Autom.*, Apr. 2005, pp. 2432–2437.

[51] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics* (Intelligent Robotics and Autonomous Agents). Cambridge, MA, USA: MIT Press, 2005.

[52] P. E. Hart, N. J. Nilsson, and B. Raphael, "A formal basis for the heuristic determination of minimum cost paths," *IEEE Trans. Syst. Sci. Cybern.*, vol. SCS-4, no. 2, pp. 100–107, Jul. 1968, doi: 10.1109/TSSC.1968.300136.

[53] B. Magyar, N. Tsiogkas, J. Deray, S. Pfeiffer, and D. Lane, "Timed-elastic bands for manipulation motion planning," *IEEE Robot. Autom. Lett.*, vol. 4, no. 4, pp. 3513–3520, Oct. 2019.

[54] D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance," *IEEE Robot. Autom. Mag.*, vol. 4, no. 1, pp. 23–33, Mar. 1997.

[55] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 1861–1870.

[56] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, 2013.

[57] J. H. DiBiase, "A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays," Ph.D. thesis, Brown Univ., Providence, RI, USA, 2000. [Online]. Available: http://www.glat.info/ma/av16.3/2000-DiBiaseThesis.pdf

[58] M. Oquab et al., "DINOv2: Learning robust visual features without supervision," 2024, *arXiv:2304.07193*.

Georg K. J. Fischer received the B.Sc. degree in electrical engineering from TU Braunschweig, Braunschweig, Germany, in 2017, and the M.Sc. degree in electrical engineering and information technology from the Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, in 2019.

He joined the Fraunhofer Ernst-Mach-Institute (EMI), Freiburg, Germany, in 2020, as a Research Assistant in the field of signal processing.

Max Bergau received the B.Sc. and M.Sc. degrees in physics from the University of Freiburg, Freiburg, Germany, in 2013 and 2015, respectively. He is pursuing the Ph.D. degree in computer vision and optical gas imaging (OGI) with the University of Freiburg, and Endress+Hauser, Freiburg.

He worked at SensoPart, Gottenheim, Germany, for five years with vision sensors for robotics.

D. Adriana Gómez-Rosal received the B.Sc. degree in mechatronics from the National Polytechnic Institute, Mexico City, Mexico, in 2011, where she specialized in robotics and control, the M.Sc. degree in computer science from the Center for Research and Advanced Studies of the National Polytechnic Institute, in 2016, and the M.Sc. degree in space robotics and automation from Aalto University, Espoo, Finland, and Luleå Tekniska Universitet, Luleå, Sweden, in 2018.

After gaining years of industrial experience in companies such as Intel Guadalajara Design Center, Zapopan, Mexico, and ABB, Auburn Hills, MI, USA, working on real-world robotics applications, she joined the University of Freiburg, Freiburg, Germany, in 2021, where she currently works as a Doctoral Research Assistant at the intersection of mobile robotics, machine learning, and perception.

Andreas Wachaja received the Dipl.-Ing. degree in mechanical engineering from the Karlsruhe Institute of Technology, Karlsruhe, Germany, in 2013.

He was a Researcher in the field of mobile robotics, human–machine interaction, and LiDAR mapping at Albert-Ludwigs-University Freiburg, Freiburg, Germany, from 2013 to 2019. Since 2019, he has been a Co-Founder and a Lead Engineer at dotscene GmbH, Freiburg, a company that focuses on the development of mobile laser scanning technology.

Johannes Graeter received the diploma in mechanical engineering from the Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, and the INSA de Lyon, Villeurbanne, France, in 2013, and the Ph.D. degree in mechanical engineering from the Institute of Measurement and Control, KIT, in 2019, in the topic of Visual- and LiDAR odometry for autonomous driving.

After two years at Daimler AG, Stuttgart, Germany, working in the field of behavior prediction, he joined dotscene GmbH, Freiburg, Germany, in 2021, where he is the Lead SLAM Software Engineer.

Matthias Odenweller received the diploma and Ph.D. degrees in physics from Goethe-Universität Frankfurt am Main, Frankfurt, Germany, in 2005 and 2010, respectively.

He joined the Department of Process Technology, Evonik, Essen, Germany, in 2011, with a main focus on automation topics, such as process analytical technology (PAT), and emerging new technologies, such as robotics for plant inspection.

Uwe Piechottka received the master's degrees in electrical engineering and control from the University of Duisburg, Duisburg, Germany, and the University of Washington, Seattle, USA, in 1984 and 1985, respectively, and the Ph.D. degree from the University of Duisburg, in 1989, with focus on nonlinear systems and robotics.

Afterwards he joined Degussa AG which is now Evonik Industries AG. He held different positions in maintenance, engineering, and process technology. Currently he is Head of the group Automation & Process Analytical Technology in the department Process Technology and Engineering.

Fabian Höflinger received the bachelor's degree in automation engineering from the University of Applied Sciences Ravensburg-Weingarten, Weingarten, Germany, in 2005, the master's degree in automation and energy systems from the Mannheim University of Applied Sciences, Mannheim, Germany, in 2007, and the Ph.D. degree in microsystems engineering from the University of Freiburg, Freiburg, Germany, in 2014, with a focus on localization systems.

He was with Junghans Feinwerktechnik, Dunningen/Seedorf, Germany, where he developed components for telemetric systems. From 2007 to 2010, he was a Development Engineer. Since 2010, he has been with the Laboratory for Electrical Instrumentation and Embedded Systems, Department of Microsystems Engineering, University of Freiburg, where he has also been a Group Leader since 2014 and has a research group in the field of indoor localization. Since 2019, he has also been working at the Fraunhofer Ernst-Mach-Institute (EMI), Freiburg.

Nikhil Gosala (Graduate Student Member, IEEE) received the B.E. degree in computer science from the Birla Institute of Technology and Science—Pilani, Pilani, India, in 2017, and the M.Sc. degree in computer science from ETH Zürich, Zürich, Switzerland, in 2020.

He is currently working as a Doctoral Researcher at the Robot Learning Laboratory, University of Freiburg, Freiburg, Germany, specializing in robot perception and autonomy.

Niklas Wetzel received the diploma in mathematics from the University of Freiburg, Freiburg, Germany, in 2017.

He joined the Autonomous Intelligent Systems Laboratory, Department of Computer Science, University of Freiburg, as a Research Assistant, in 2018.

Daniel Büscher received the Ph.D. degree in particle physics from the University of Freiburg, Freiburg, Germany, in 2012.

He is currently a Postdoctoral Researcher with the Autonomous Intelligent Systems Group, University of Freiburg, headed by Wolfram Burgard. His research focuses on autonomous robot navigation and perception using deep learning.

Abhinav Valada (Member, IEEE) received the M.S. degree in robotics from Carnegie Mellon University, Pittsburgh, PA, USA, in 2013, and the Ph.D. degree in computer science from the University of Freiburg, Freiburg, Germany, in 2019.

He is a Full Professor and the Director of the Robot Learning Laboratory, University of Freiburg. He is a member of the Department of Computer Science, a Principal Investigator at the BrainLinks-BrainTools Center, and a Founding Faculty of the European Laboratory for Learning and Intelligent Systems (ELLIS) unit, Freiburg. His research lies at the intersection of robotics, machine learning, and computer vision with a focus on tackling fundamental robot perception, state estimation, and planning problems using learning approaches in order to enable robots to reliably operate in complex and diverse domains.

Dr. Valada is a Scholar of the ELLIS Society, a DFG Emmy Noether Fellow, and the Co-Chair of the IEEE RAS TC on Robot Learning.

Wolfram Burgard (Fellow, IEEE) completed the dissertation at the Institute of Computer Science, University of Bonn, Bonn, Germany, in 1991.

He has been a Founding Chair of the Department of Engineering and a Professor of Artificial Intelligence and Robotics with the University of Technology Nuremberg, Nuremberg, Germany, since 2022. Prior to this assignment, he was a Professor of Autonomous Intelligent Systems with the University of Freiburg, Freiburg, Germany. He has published over 350 papers and articles at conferences and in journals on *Robotics and Artificial Intelligence*. In 2005, he coauthored two books.

Prof. Burgard is a member of the National Academy of Sciences Leopoldina and an Honorary Member of the European Coordination Committee for Artificial Intelligence, the Institute of Electrical and Electronics Engineers, and the American Association for Artificial Intelligence. He received the Gottfried Wilhelm Leibniz-Prize in 2009, the Advanced Grant of the European Research Council in 2010, and the Technical Field Award of the Institute of Electrical and Electronics Engineers in 2021. He also coordinated the Cluster of Excellence BrainLinks-BrainTools from 2012 to 2019, which was funded by the German Research Foundation. He served as the President of the IEEE Robotic and Automation Society from 2018 to 2019.