
EKSPLOATACJA I NIEZAWODNOŚĆ
MAINTENANCE AND RELIABILITY
Polskie Naukowe Techniczne Towarzystwo Eksploatacyjne
Warszawa
Polish Maintenance Society
Warsaw

**Eksploatacja i Niezawodnosc – Maintenance and Reliability**
**Volume 28 (2026), Issue 3**
journal homepage: http://www.ein.org.pl

Article citation info:
Koszyk J, Hyla B, Pieczonka L, Ambrozinski L, Multi-sensor robotic mapping with thermal and acoustic imaging for non-contact technical diagnostics, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2026: 28(3) http://doi.org/10.17531/ein/217575

# Multi-sensor robotic mapping with thermal and acoustic imaging for non-contact technical diagnostics

Indexed by: Web of Science Group

**Joanna Koszyk[a], Bartosz Hyla[a,*], Lukasz Pieczonka[a], Lukasz Ambrozinski[a]**

[a] Department of Robotics and Mechatronics, AGH University of Krakow, Poland

## Highlights

- The work presents automated mapping and inspection of industrial environment.
- Presents thermal and acoustic modalities for cooperative fault detection.
- Utilizes point cloud for robot navigation and diagnostic data representation.
- Multi-sensor data fusion to localize and identify anomalies in the machinery.

## Abstract

Machine diagnostics is essential in maintaining high operational efficiency and minimizing machinery downtime. One way to achieve this goal is by implementing automated diagnostic systems. This paper presents a multi-sensor mapping system developed on a quadruped robotic platform. The robot combines diagnostic data from thermal and acoustic sensors with LiDAR data to create a 3D map of the working environment. Anomalous readings of temperature and sound pressure levels are identified and indicated on a 3D map with distinct color values through data fusion. Validation tests performed in a laboratory environment confirmed the feasibility of this approach by correctly identifying simulated cases of malfunctioning machinery. The system robustness was evaluated through repeatability tests. The proposed multi-modal mapping solution enables reliable monitoring of the technical state of machinery in complex environments.

**Keywords**
technical diagnostics, data fusion, quadruped robot, multi-modal mapping, robot perception

## 1. Introduction

Machine diagnostics is essential in preventing severe machine faults and reducing downtimes [17]. Early detection of malfunctioning machinery helps to improve safety and effectiveness of the production process [12,22]. Currently, on-demand nondestructive evaluation (NDE) involve trained human personnel and are usually performed during planned machinery downtimes [1]. Such a solution increases operational efficiency by avoiding unplanned downtimes caused by unpredictable machine failures. However, scheduled maintenance periods are still required to allow human personnel to perform the necessary inspections in a safe working environment, which negatively affects overall operational efficiency. To resolve this issue, a robot might be applied to perform the inspection in an automated manner during the regular operation of the machinery. Employing a mobile robotic platform to perform inspection tasks should enhance the diagnostic process by making it more time-efficient [5] and objective due to its reduced dependency on human factors [13]. Moreover, inspections carried out by humans are often limited to one modality at a time, while a robot can simultaneously carry and control multiple sensors and process the measurements at the same time. Certainly, human inspectors

---

(*) Corresponding author.
E-mail addresses: J. Koszyk, (ORCID: 0009-0007-3609-9670) jkoszyk@agh.edu.pl, B. Hyla (ORCID: 0009-0003-7153-6486) bhyla@agh.edu.pl, L. Pieczonka (ORCID:0000-0003-3623-3984) lukasz.pieczonka@agh.edu.pl, L. Ambrozinski (ORCID:0000-0002-3311-2821) ambrozin@agh.edu.pl

also rely on their own senses, which might help discover abnormal events in the field. However, the performance of hearing and eyesight is limited, person-specific, and did not evolve for anomaly detection in industrial environments. Acoustic symptoms of machinery failure often extend beyond the audible range, and thermal symptoms are too subtle to be detected by humans from afar.

Developing an autonomous robotic system to perform diagnostics, however, is a complicated task. Robot navigation in a complex environment is itself a complex matter that requires solving the obstacle avoidance [18], SLAM (simultaneous localization and mapping) [15] and path planning [19,24]. In addition to sensing and navigating through the environment, the robot needs to collect and process diagnostic data, which further increases the overall complexity.

A mobile robot suitable for operating in industrial environments can be customized for nondestructive inspections. Acquiring multi-sensory diagnostic data while navigating through an industrial environment delivers an abundance of information that allows for the identification of anomalous states of the machinery. It is essential to process and interpret this data so that the anomalies can be detected and localized with reference to the global frame of reference. This requires efficient geometrical mapping of the working environment in 3D and fusing this information with diagnostic data.

In this article, we propose a diagnostic method designed for automated mapping and inspection of industrial environment. The approach integrates multi-sensor data fusion to identify and localize anomalous states of machinery during autonomous missions in complex settings. The main novelty of this paper lies in presenting diagnostic data as a point cloud representation, enabling spatially accurate visualization of anomalies within the 3D environment. The proposed technique is validated in a laboratory environment simulating real industrial conditions, demonstrating its capability to provide comprehensive anomaly indications through multi-modal sensing.

## 2. Related Work

Autonomous diagnostics plays a crucial role in enhancing efficiency, safety, and reduce costs [7]. Although the solution to the problem strongly relies on the field and application, several studies have been performed in the area of robotic inspections. Different types of robots have been developed and adapted to meet the diverse operational demands of diagnostics, including ground-based robots [14,27], aerial drones [16,20,29], climbing robots [25], and underwater vehicles [21]. Each platform type presents distinct advantages depending on the environment, inspection modality, and surface accessibility. Recent studies emphasize the growing relevance of quadruped robots in the field of NDE [8,26]. Recent implementations demonstrate that these robots can reliably carry thermal [4], visual and other sensors, supporting multi-modal NDE inspections without significant compromise in mobility or battery life.

Robotic multi-modal perception allows to use the advantages of multiple sources and compensate for single sensor limitations by enhancing measurements through sensor fusion. In literature, combining heterogenous sensors on robots has been utilized mainly for simultaneous localization and mapping (SLAM) to improve accuracy and robustness of pose estimation [6,9,11].

The use of sensor fusion can provide reliable pose estimation on various mobile platforms, including legged robots. For example Belter and Nowicki [2] proposed a fusion of robot odometry and RGB-D cameras to enhance pose estimation of walking robots. Through a combination of visual SLAM and legged odometry, optimisation-based motion estimation was used as a feedback, making the algorithm robust to rapid movements. Another approach, Multi-Modal SLAM (MIMOSA) [10] was proposed to enhance robustness of pose estimation in GPS-denied environments depending on odometry, point cloud data and IMU readings. The data was fused through first estimating the position from each sensor separately and then performing factor graph optimization.

More recent works have explored the use of unconventional sensors like thermal cameras for pose estimation. The work of Bultmann et. al. [3] proposed a combination of LiDAR, RGB camera and thermal data. The authors prepared a CNN-based neural network to acquire semantic information. The solution prepared in the study allowed to mitigate domain adaptation issue.

Ongoing investigations are focused on sensor fusion to acquire a higher level of perception to the extent of panoptic segmentation. In [28], instance recognition is obtained through fusion of depth, geometric information and semantic labels.

Sensor fusion in the field of mobile robotics is usually considered for improvement of localization and mapping. Despite different data characteristics, these measurements are often used to acquire the same information, which is pose estimation. In our work, we explore the possibility of combining autonomously acquired diagnostic data from sources that contain non-overlapping information. Acoustic and thermal data can complement each other’s measurements by providing independent assessment of level of threat severity. Creating a fused diagnostic system aims to construct a comprehensive view on the occurrence of anomalies in the environment.

## 3. Experimental setup

### 3.1. Inspection system

The following section describes the hardware and software components of the diagnostic system mounted on the quadruped robot shown in Fig. 1. The inspection payload consisted of a thermal camera FLIR A70, an acoustic camera Sorama L642v+, a LiDAR sensor Ouster OS-1 with 32 channels, and an on-board computer NVIDIA Jetson Orin Nano. The system was running on Ubuntu 20.04 Focal and ROS Noetic. The FLIR A70 thermal camera had a field of view of hFOV × vFOV = 51° × 39° and spatial resolution of width × height = 640 × 480 pixels. The Sorama L642v+ acoustic camera was set to record frequency bandwidth from 5000 to 10000 Hz, and its field of view was hFOV x vFOV = 90° x 51°. The thermal camera captures radiometric images, enabling temperature measurement in each pixel with a 2 Kelvin accuracy, and stores it in the form of a matrix. An acoustic camera delivers a visual image of the environment with a color scale overlay of the sound pressure level values, representing the sound sources. LiDAR sensor allows for the capture of a spatial representation of the environment in the form of a point cloud. In this study, X, Y, and Z coordinates of points are collected. LiDAR does not capture color values.

The inspection payload was mounted on a quadruped mobile robot. The mobile platform, *Spot*, from *Boston Dynamics*, has five stereo cameras that are employed for obstacle avoidance and are used to handle rough terrain or stairs. The advantage of using a legged robot instead of a wheeled platform is its capability to move in complicated environments with uneven ground.

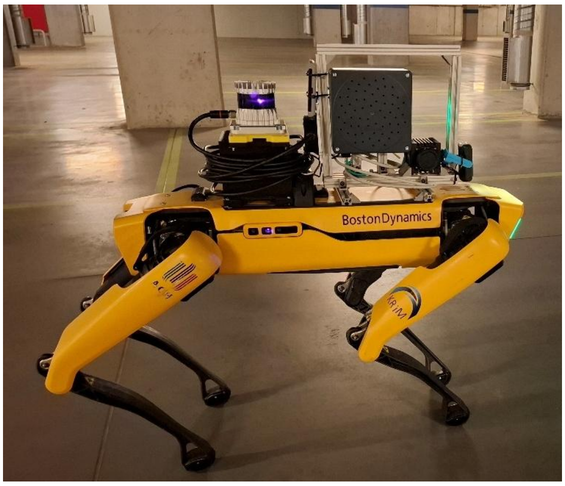

Figure 1. Mobile platform equipped with diagnostic system.

### 3.2. Test Field Configuration

The mobile platform can perform autonomous missions in large, complex, multilevel environments. In this work, we limited the working area to a single-level environment of an industrial robots laboratory. To access the room, the robot had to traverse a corridor and walk down the stairs. To mimic the industrial environment, robotic manipulators and auxiliary equipment present in the laboratory were actively operating during the test. A view of the scene is shown in Fig. 2.

Figure 2. Test environment illustrating a quadruped robot equipped with an inspection system and a malfunction-simulation board.

To enable simulation of 4 pneumatic and 8 electrical malfunctions, a test board shown in Fig. 3 was prepared. The setup is controlled with a *Raspberry Pi* microcomputer, which can activate faulty modules to intentionally generate acoustic and thermal responses on demand. These signals are subsequently interpreted by the robot's diagnostic sensors as indicators of malfunction.

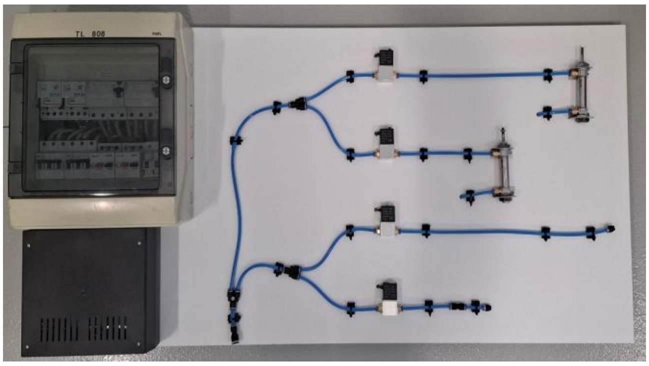

Figure 3. Pneumatic and electrical test board used to simulate malfunctions in selected nodes of the system.

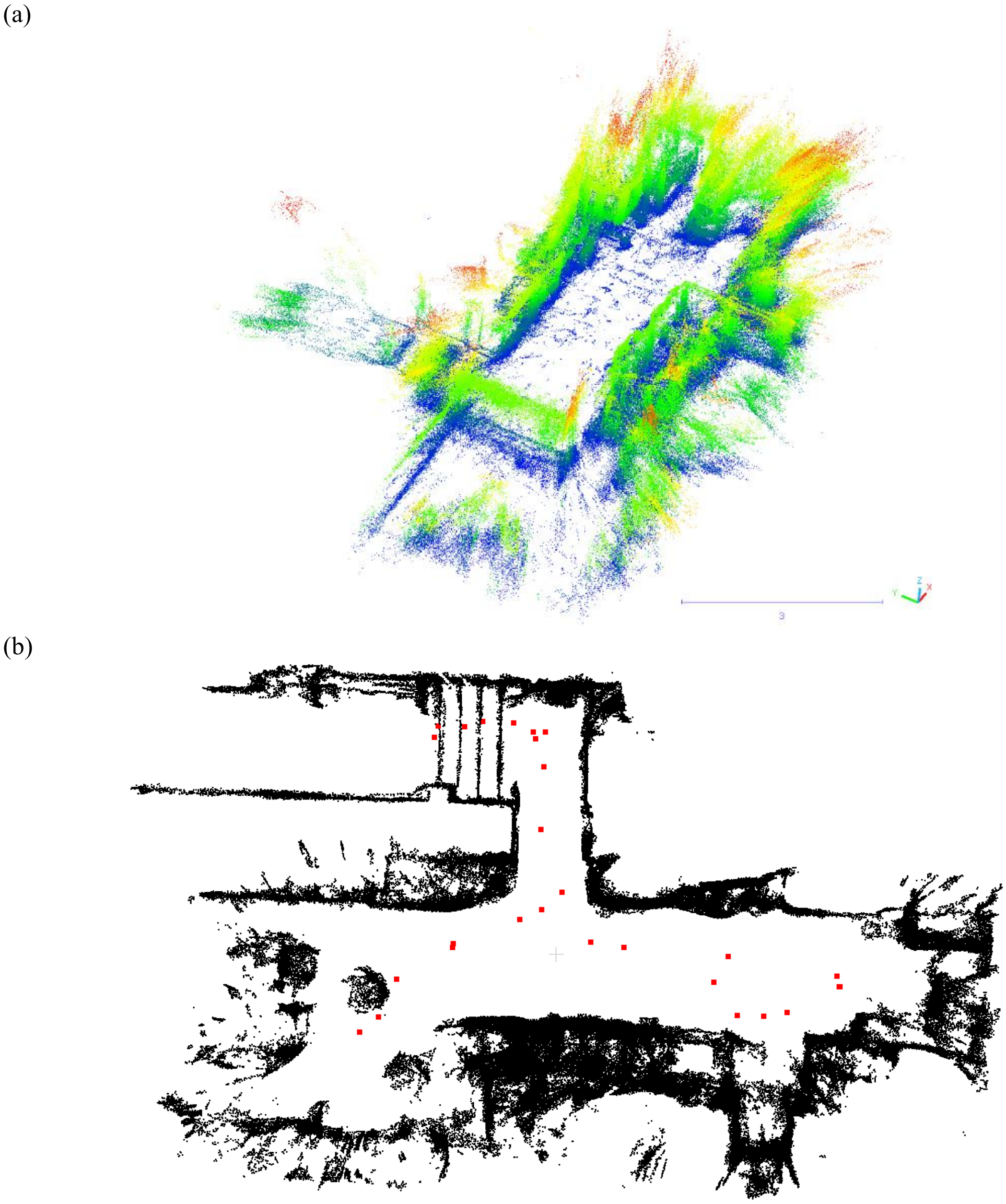

Figure. 4 Map of the scene obtained from the robot's cameras (a) point cloud (b) 2D map with waypoints indicated as red squares.

### 3.3. Autonomous inspection

The mobile platform features Autowalk® capability. In this mode, the robot moves autonomously without the operator’s supervision. To realize this task, visual reference markers called fiducials must be placed around the environment, and the operator has to plan the robot’s path and actions in advance. When performing a mission, the robot repeats prerecorded moves, undertaking necessary modifications in response to the changes to the environment. During the recording, the environment representation is gathered for further self-reliant navigation. The map in the form of a point cloud is obtained with the robot’s depth cameras and mostly contains characteristic visual features such as edges of the objects in the surroundings. The map of the test field, acquired using the robot’s built-in cameras, is shown in Fig. 4a. This map is obtained through a separate SLAM process, as a built-in robot function. The map is used for navigation in Autowalk®, but is not further utilized in our diagnostic mapping process.

Besides acquiring the map, the robot stores information on the route, the waypoints it is supposed to cross, fiducials, and localizations where it should reach the previously determined pose to steer the diagnostic sensors towards the board. The waypoints are visualized on 2D map for better clarity in Fig. 4b. The 2D map is created by projection of all points from 3D map obtained from robot’s cameras (Fig. 4a) to the XY plane. The waypoints are recorded during Autowalk® mission planning and the robot moves from a waypoint to waypoint during operation. The depth camera map and waypoints are used for navigation solely because the method is not sufficiently accurate for diagnostic mapping. The data from all sensors mounted on the robot was gathered in Autowalk® mode while performing a mission without the operator’s supervision. The mission was repeated 5 times in order to assess the repeatability of the inspection process in the view of external disturbances.

## 4. Multi-sensor mapping of the environment

This section is focused on a diagnostic mapping solution that combines a LiDAR sensor with thermal and acoustic detections. The multi-sensor mapping method, proposed in the study, is performed by first projecting thermal and acoustic images onto point cloud, which involves obtaining coordinate transformation of sensors, described in section 4.1. The next stage, described in section 4.2, requires pose estimation of consecutive point clouds. Multi-sensor diagnostics (section 4.3) are done through anomaly-based coloring. Finally, multi-sensor map of the area is obtained through combining thermal and acoustic maps, which is described in section 4.4.

### 4.1. Coordinate Transformation for Diagnostic Imaging

The first step of our proposed framework is thermal and acoustic image projection to corresponding point cloud data. Data frames are matched based on timestamp difference. Given the time acquisition difference lower than 0.05 s, the data is considered to be gathered at the same time. Further step involves image projection to a point cloud within the same time stamp taking into account corresponding orientation of the sensing devices. Exemplary sensor alignment on the robot is visualized in Fig. 5.

Thermal cameras measure thermal radiation emitted within their field of view. Temperature distribution is given across the image.

To obtain a combination of thermal data and LiDAR scans a point cloud in 3D space is defined as $P = \{p_i \in \boldsymbol{R}^3\}$. Each point $p_i$, represented through x, y, z coordinates, is projected onto the projection plane Π in accordance to equations 1 to 4.

$$\Pi(p_i) = \left(\frac{u\prime}{w}, \frac{v\prime}{w}\right), \tag{1}$$

where $u$ and $v$ represent the pixel format on x and y dimensions and $w$ is the distance from the projection plane. The values are expressed as:

$$\begin{bmatrix} u' \\ v' \\ w \end{bmatrix} = \boldsymbol{K} \cdot \boldsymbol{R} \cdot (p_i - t) \tag{2}$$

where $t$ is a translation from laser scanner to thermal camera given as [dx dy dz]. **R** is rotation matrix of the cameras regarding to the point cloud coordinate system (see Fig 5b) expressed as:

$$\boldsymbol{R}^{3x3} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix} \tag{3}$$

***K*** is a camera calibration matrix given as:

$$\boldsymbol{K}^{3x3} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{width}{2\tan\left(\frac{hFOV}{2}\right)} & 0 & \frac{width}{2} \\ 0 & \frac{height}{2\tan\left(\frac{vFOV}{2}\right)} & \frac{height}{2} \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

This step is followed by checking the constraints given by the image frame expressed as:

$$u \in <u_{min}, u_{max}>, v \in <v_{min}, v_{max}>, z > 0, \tag{5}$$

where $u_{min}, u_{max}, v_{min}, v_{max}$ are image pixel limits and $z$ above 0 filters points that lie behind the camera.

Consecutive step involves assigning the temperature value given the projected coordinates:

$$T(p_i') = Im(\Pi(p_i')) = Im(u, v). \tag{6}$$

Acoustic cameras acquire intensity of sound sources within the frame. The camera acquires Acoustic cameras capture the intensity of sound sources within their field of view. They provide both the spatial localization of each source and its corresponding sound intensity. The data acquired by an acoustic camera are represented as an image that depicts the spatial distribution of sound sources across a specified frequency range. The projection of an acoustic image is generated using a procedure analogous to that employed for thermal image projection. Note, however, that these two systems are not necessarily synchronized, moreover they can operate at two different intervals. Therefore, the thermal and acoustic point clouds may be created from different, independent LiDAR scans.

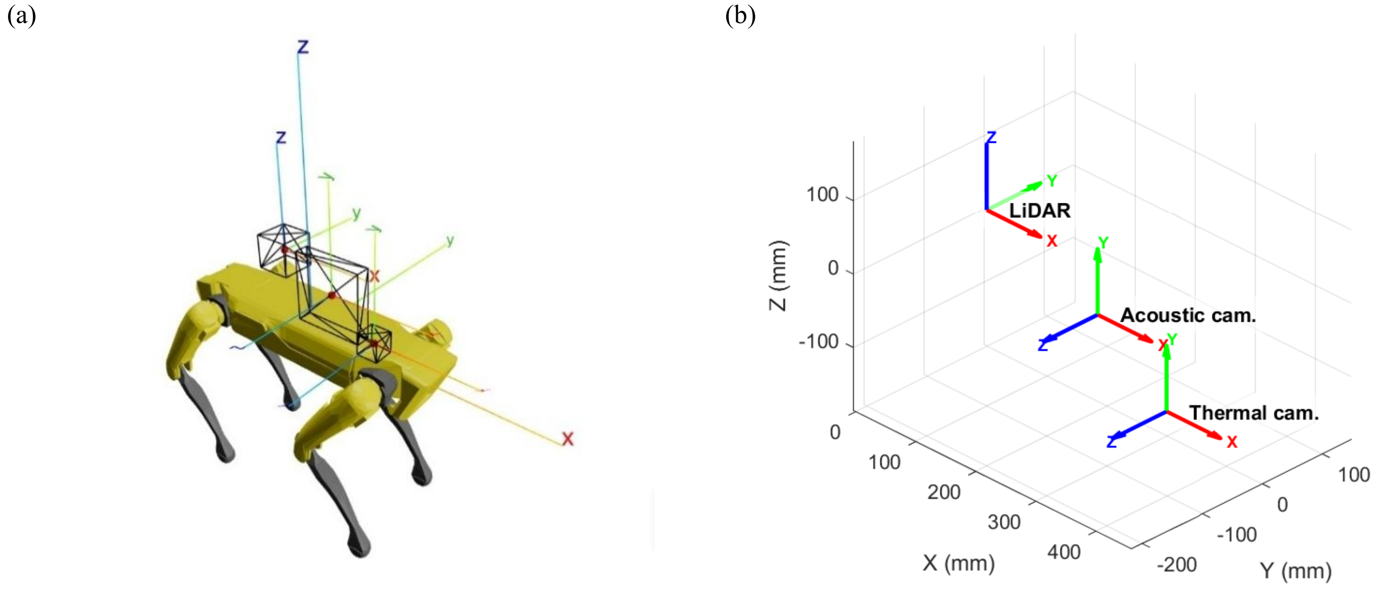

Figure 5. (a) Sensor alignment on the robot, (b) spatial alignment of the sensor coordinates.

### 4.2. Pose estimation

Pose estimation for each data snapshot is achieved through point cloud registration, which aligns spatial data across successive scans to determine the robot's position within the environment. At each measurement location, the local LiDAR point cloud is preprocessed by filtering points based on their Z-axis values, effectively removing ceiling and floor data to facilitate subsequent registration. In the consecutive step, the quaternion data is extracted from LiDAR's built-in IMU, and initial point cloud alignment is performed using the IMU data. In the final stage, ICP (Iterative Closest Point) [23] is performed to match consecutive point clouds. Based on the registration, a relative transformation is found and applied to update the robot's global pose. The process is repeated for each LiDAR scan.

### 4.3. Multi-sensor diagnostic

To distinguish individual sources of diagnostic indications on the resulting unified map, classification scheme organized into two distinct damage type is introduced. These indications are used next to apply coloring of the point cloud. Standard 24-bit encoding is utilized to represent three separate color channels: red, green, and blue, each corresponding to specific diagnostic modalities.

The thermal indications are coded in a red channel by linear mapping temperatures from 0˚C up to an alarm level set above 50˚C. The threshold value can be adjusted to the tested object in the range from 0˚C to 250˚C.

An acoustic camera, used in our study, returns monochromatic images highlighting areas with sound sources.

The raw acoustic indication is linearly mapped with an alarm level set above 50dB. The resulting acoustic indications were marked on a channel encoding blue color. The threshold value can be adjusted to the tested object in the range from 0dB to 120dB.

The green channel is left to represent common parts, i.e., areas with temperature and sound level exceeding the set threshold at the same time. These areas are interpreted as abnormal system operation. To determine the intersection of the areas where thermal and acoustic alarms are raised, it is necessary to project diagnostic information onto the point cloud.

### 4.4. Area mapping

The 3D model of the environment is created form a point cloud obtained from a single scan captured at the same localization as the corresponding thermal and acoustic images. The images are projected onto the point cloud collected at a given measurement location. This projection is performed by computing the image coordinates for each point in the scan. Each point in in the point cloud is subsequently annotated with diagnostic data extracted from the corresponding pixels in the thermal and acoustic images.

Once this step is completed global mapping is performed. Pose estimation is used to calculate the transformation matrices for the consecutive measurement areas yielding independent thermal and acoustic maps of the entire area. Multi-sensor maps are created by transforming the point clouds with diagnostic data to KDTrees of the voxel size of 0.01 m. The points in the acoustic point cloud localized inside the voxels in the thermal KDTree, that has a temperature above the alarm level, are marked green. Similarly, the points in the thermal point cloud that are localized inside the voxels of the acoustic KDTree, with a raised acoustic alarm, also marked green.

## 5. Experimental Results

In this section, we describe the prepared diagnostic mobile system as well as the measurements conducted in a test field. Moreover, the results obtained are presented and discussed. Section 5.2 presents exemplary results as well as potential false alarms generated by individual sensors. The mapping procedure, detailed in Section 5.3, emphasizes the integration of data into a unified representation. Section 5.4 discusses the evaluation of the system within the context of a repeatability experiment.

### 5.1. Scene mapping

Registered point clouds sequence counts 1100 scans of the environment. Consecutive scans transformations are computed with ICP method described in section 4.2 and scans are finally aligned into the single dense point cloud.

The map containing all scans aligned within a common reference frame is shown in Fig. 6.

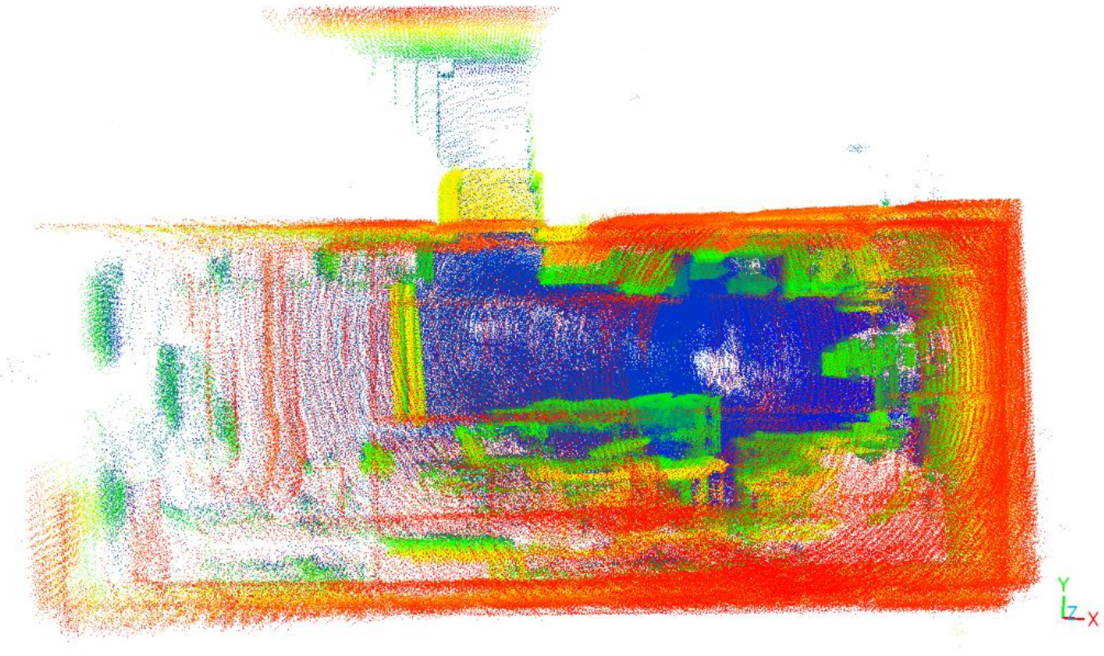

Figure 6. Aligned LiDAR scans of the test scene based on pose estimation. Colors indicate height ramp.

### 5.2. Sensor response

Examples of test-board operation can be seen in Fig. 7 and Fig. 9, where acoustic and thermal anomalies are visible.

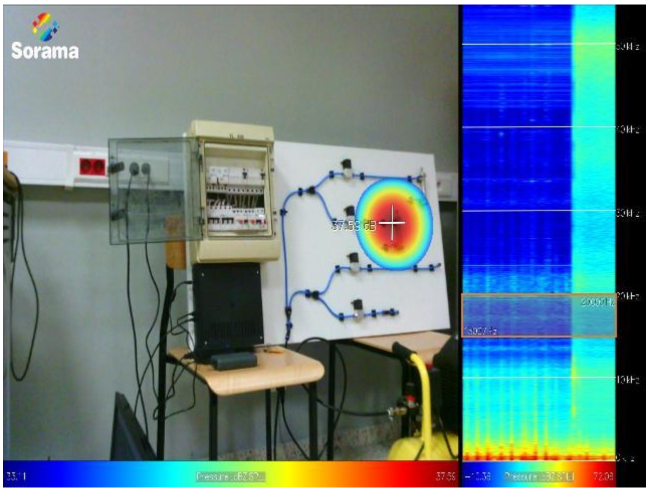

Figure 7. Acoustic camera indication of compressed air leak detection from the pneumatic tube.

Acoustic indicators are also visible in Fig. 8. They occur because the industrial manipulators are equipped with pneumatic suction cups to move objects in their operational fields. The flat walls in the laboratory are effective reflectors for acoustic waves, thus, few indications of the sound sources on the walls are also registered. These features are discarded from anomaly detections, as they origin form normal operation behavior.

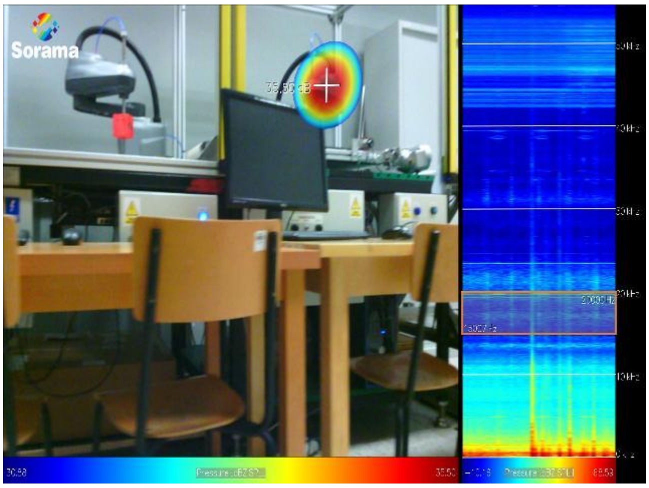

Figure 8. Acoustic camera readings from manipulators with pneumatic end effector.

Thermal camera provides information on the temperature distribution in its field of view. Fig. 9 presents the testing board with temperature exceeding 35˚C (in the center of the image) and a computer display operating in the temperature range of approximately 40˚C.

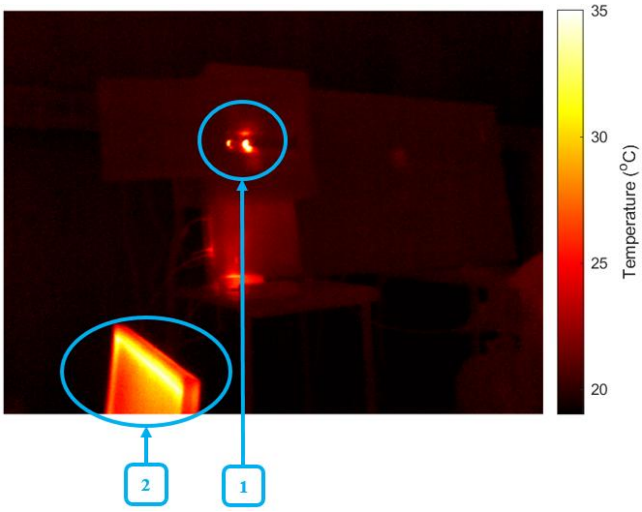

Figure 9. Thermal anomaly detection from the test case. (1) thermal malfunction of a relay switch simulated on testing board. (2) thermal signature of a correctly operating element.

In the collected data, none of the environment parts in the inspected environment indicated faults in both modalities simultaneously, however many elements exceeded one of the alarm thresholds. The testing board is prepared to be the only element that exceeds both thermal and acoustic alarm levels simultaneously. The combined information from the two fields is necessary to distinguish the true damage indications.

### 5.3. Multi-sensor mapping

The measurements obtained from the diagnostic payloads are fused to create the multi-sensor map. Thermal and acoustic indications are separately projected onto the map of the scene. The result of thermal mapping is shown in Fig. 10. The points with temperature above 35 ˚C are marked with red color. The rest of the point cloud without anomaly indication is colored black. As expected, the indications originate from the test stand.

Acoustic mapping marked more areas than thermal mapping, as shown in Fig. 11. Identified acoustic anomalies are marked with blue color. The field consisted of a few acoustic sources. Additionally, the reflections from walls were also detected by the camera and are present on the point cloud.

Thermal and acoustic maps were combined to create a multi-sensor representation of the environment. The result of data fusion is shown in Fig. 12. Red indicates thermal-only anomalies, blue indicates acoustic-only anomalies, and detections recorded by both sensors are colored green.

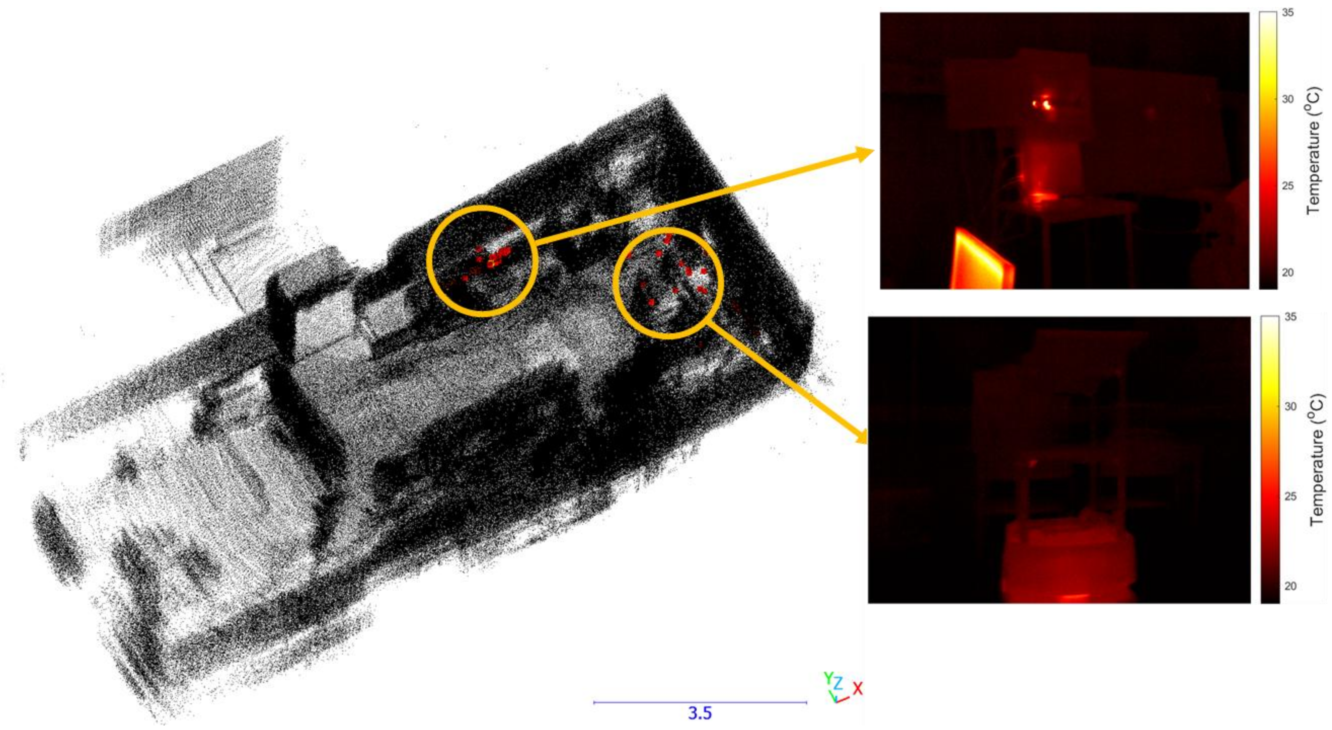

Figure 10. Thermal mapping. The acquired point cloud has black coloring in regions where no anomaly occurred. Points with thermal anomaly indication are marked red. Detected thermal anomalies shown next to the point cloud were testing board and laboratory equipment present in the environment.

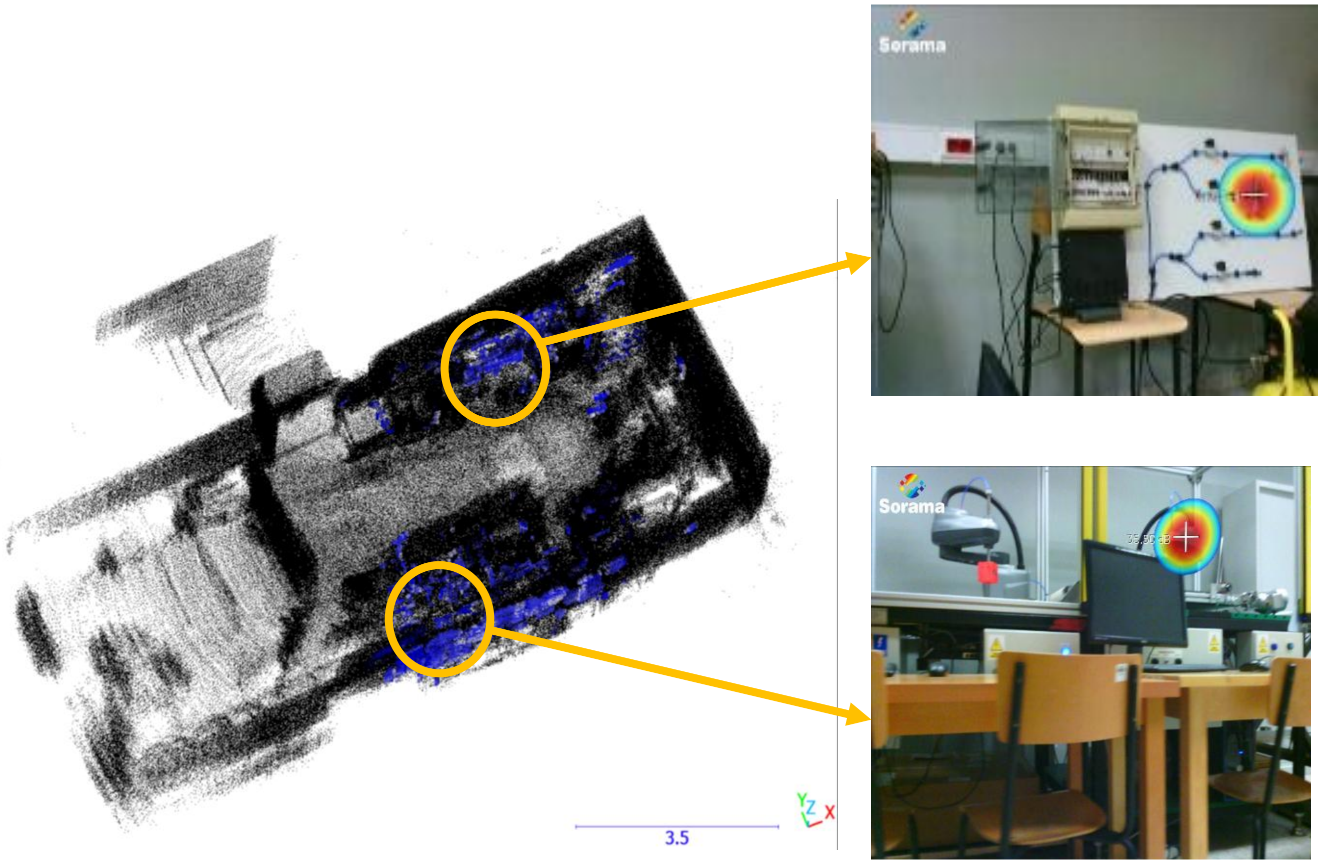

Figure 11. Acoustic mapping. The acquired point cloud has black coloring in regions where no anomaly occurred. Points with acoustic anomaly indication are marked blue. The main sources of acoustic signal are shown next to the point cloud.

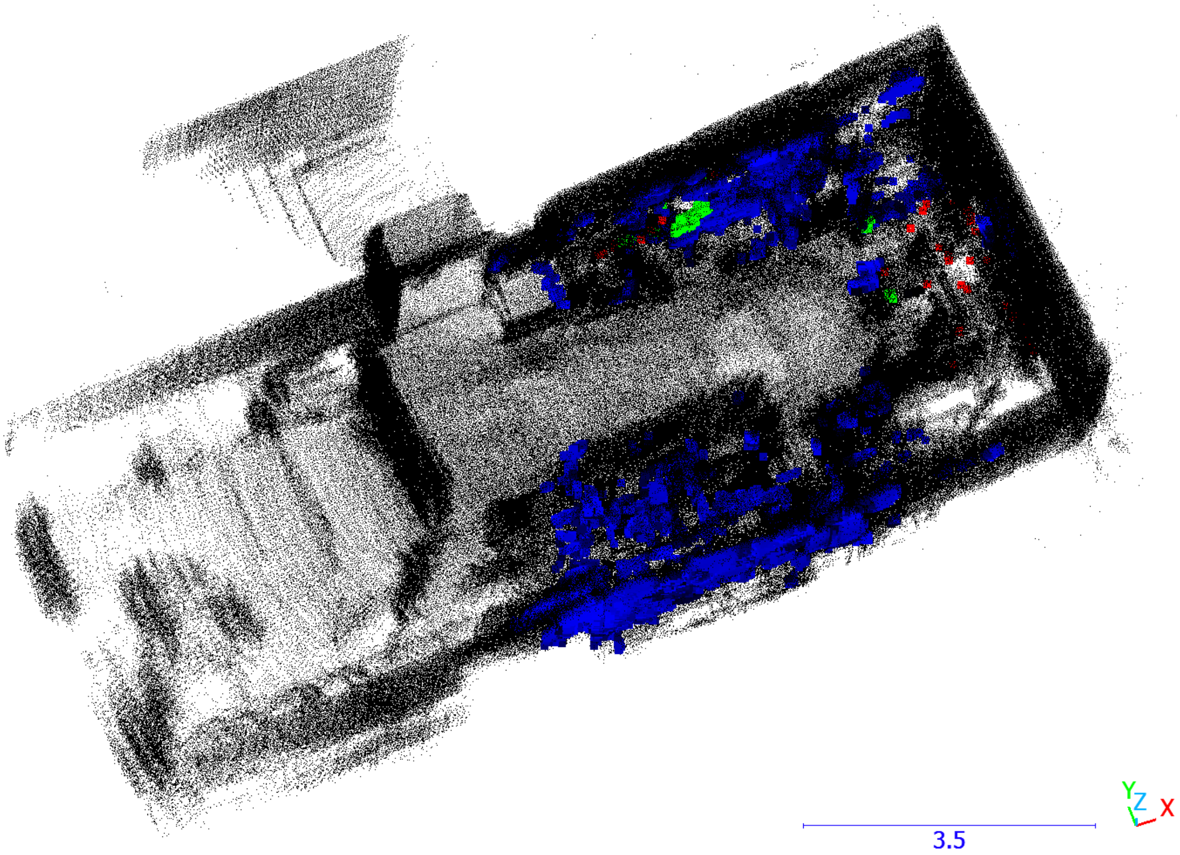

Figure 12. Multi-sensor map. The point cloud is a result of fusing thermal and acoustic map. Red color indicates the presence of thermal anomaly, blue acoustic and green both.

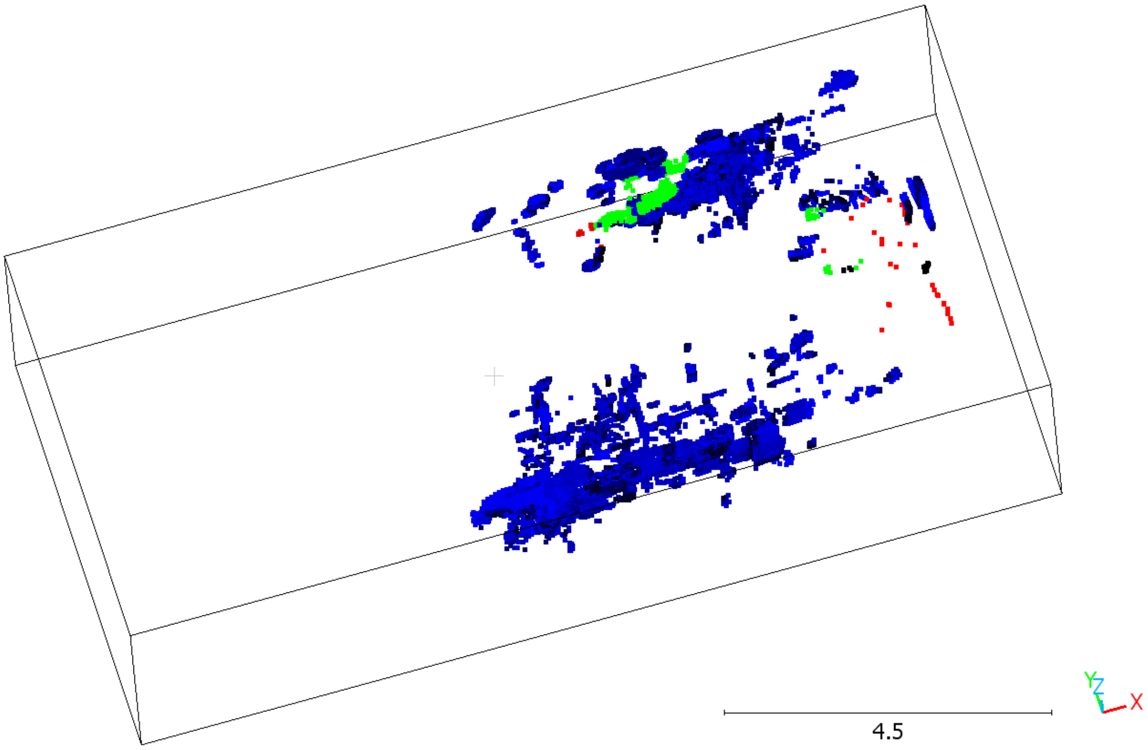

Figure 13. Point cloud showing the anomalies only. Red color indicates the presence of thermal anomaly, blue acoustic and green both. The bounding box shows the dimensions of the room.

For clarity, anomalies only are shown in Fig. 13. The diagnostic data indicated on the map can be utilized to determine areas that require further inspections.

### 5.4. Repeatability tests

The robustness of the system was assessed through repeatability tests. The experiment was conducted 5 times by making the robot repeat the same mission path and gather data during the mission execution. Number of points in point clouds for each repetition averaged to 1.32 ± 0.3 mil. Time for one repetition averaged to 71.2 ± 9 s.

As the diagnostic information stems from exceeded temperature or sound indications maximum values of these quantities were extracted from subsequent acquisition results and illustrated in Fig. 14a and b respectively. In the case of thermal results, the maximum values were detected on the testing board in proximity to the ground truth location. In the case of acoustic data, however, only 2 of 5 maximal indications were located close to the leaking actuator. This result can be explained by other sound sources, e.g. vacuum suction cups on robotic manipulators operating in the inspected room.

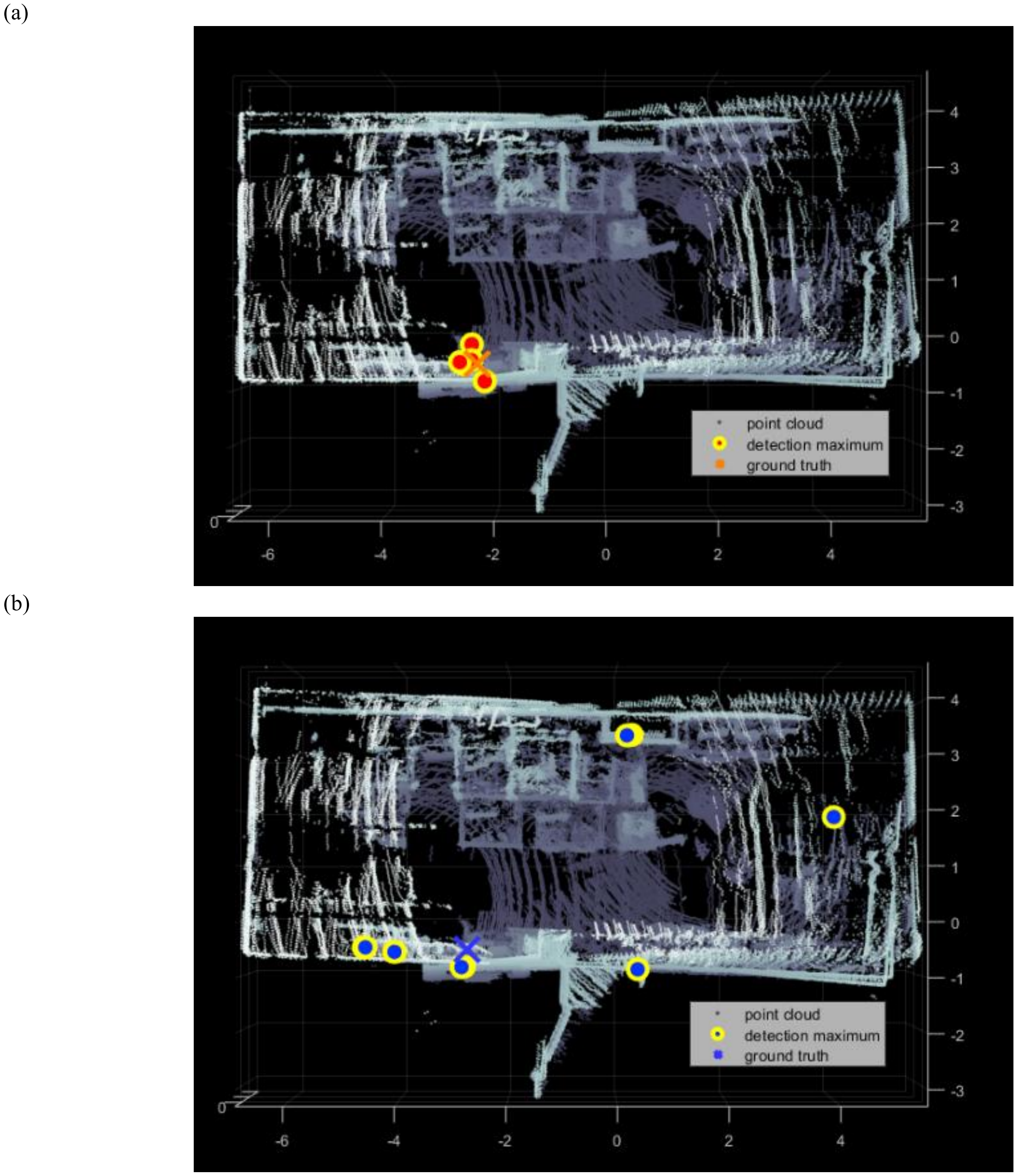

Figure 14. Distribution of diagnostic values obtained for repetition test. (a) Thermal data. (b) Acoustic data.

### 5.5. Reprojection error calculation

To evaluate the accuracy of the LiDAR-camera extrinsic calibration, a reprojection error analysis was conducted using a dedicated calibration target observed simultaneously by the LiDAR sensor and thermal and acoustic cameras. Due to the lack of distinctive point features that can be automatically extracted from the LiDAR point cloud, the reference points on the calibration target were manually annotated.

For each experimental configuration, three corresponding points were identified in the LiDAR point cloud that correspond to automatically detected points on a specifically prepared calibration board in the thermal images. The calibration target was positioned at multiple distances and orientations with respect to the sensor's placement in order to assess the stability of the calibration under varying geometric conditions as shown in Figure 15. The evaluated configurations included target distances of 1.0 m, 1.5 m, and 2.0 m, as well as target displacement to match azimuth of −15° and +15° from the center of the image at a distance of 2.0 m.

For each target configuration, the reprojection error was evaluated using three manually annotated point correspondences. The average root mean square error was computed separately for each configuration. The table 1 summarizes the reprojection error for all evaluated configurations.

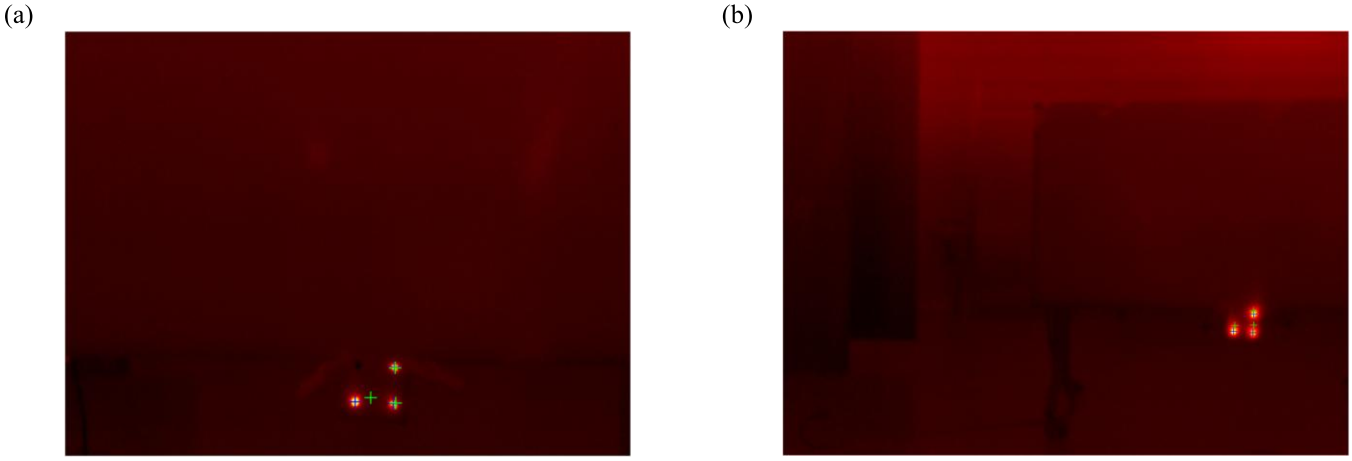

Figure 15. Thermal image from configuration (a) distance 1 meter, (b) distance 2 meters and azimuth 15° from the sensor. Blue plus sign denotes object centers detected on the image Green plus sign denotes projection of corresponding point cloud points onto an image plane. Calibration board consists of 3 thermal sources arranged in 3 square's corners spaced 60mm from the neighboring heater.

Table 1. Average RMSE of reprojection error for 5 tested thermal configurations.

| 1 m azimuth: 0° | 1.5 m azimuth: 0° | 2 m azimuth: 0° | 2 m azimuth: 15° | 2 m azimuth: -15° |
|---|---|---|---|---|
| 4.9105 px | 5.1575 px | 3.4097 px | 2.4027 px | 5.3821 px |

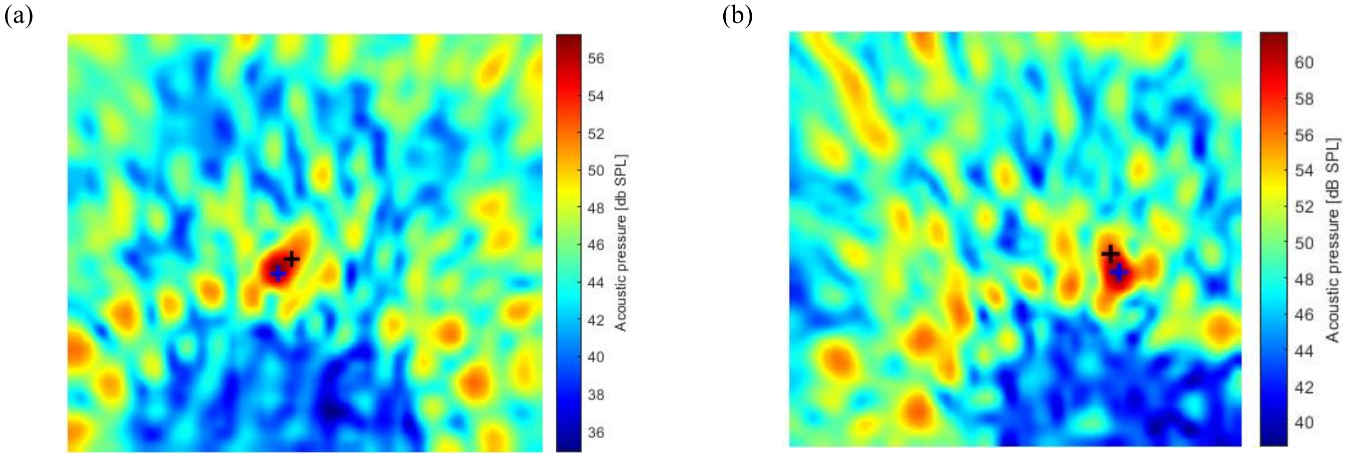

Figure 16. Acoustic pressure spatial distribution form configuration (a) distance 2m azimuth 0°, (b) distance 2m azimuth 15°.

The same reprojection error analysis procedure was repeated for acoustic camera with LiDAR setup using pneumatic test board presented in Fig 3 with single acoustic source. Testing scenarios were chosen the same as in the case of thermal camera and exemplary scenario of raw acoustic distribution image is presented in Figure 16.

For each tested scenario root mean square error was calculated between center of the detection and projection of the closes point from point cloud.

Above evaluation presented a reprojection error analysis for validating the LiDAR–camera extrinsic calibration. The error values above expectation in some experimental scenarios are caused by limited vertical resolution of the LiDAR sensor and manual annotation. The results demonstrate that the estimated transformation achieves consistent reprojection accuracy across multiple target distances and orientations, confirming its suitability for multimodal sensor fusion tasks.

The presented reprojection error analysis is subject to several limitations. First, the number of evaluated point correspondences is limited due to the manual annotation required for LiDAR point cloud labeling. Second, the reported reprojection error includes uncertainty introduced by the manual annotation process and by the spatial resolution of the LiDAR sensor.

Despite these limitations, the analysis provides a meaningful validation of the calibration accuracy under varying geometric conditions relevant to the intended application.

Table 2. RMSE of reprojection error for 5 tested acoustic configurations.

| 1 m azimuth: 0° | 1.5 m azimuth: 0° | 2 m azimuth: 0° | 2 m azimuth: 15° | 2 m azimuth: -15° |
|---|---|---|---|---|
| 4.9299 px | 7.1036 px | 21.3143 px | 16.4053px | 7.9080 px |

## 6. Discussion

While the repeatability analysis confirmed consistent thermal source localization, the acoustic anomaly was correctly flagged on the test table only 2 times out of 5 repetitions. This result suggests that the current acoustic diagnostic logic requires further extension to improve robustness and reduce false indications under realistic scene variability. At the same time, this limitation does not compromise the proposed multi-modal mapping pipeline reliably co-registers and visualizes heterogeneous thermal and acoustic cues in a common 3D representation, enabling consistent spatial interpretation and providing a foundation for future improvements of the decision-level diagnostic algorithms. To reduce false positive indications, further acoustic noise mitigation could be implemented e.g. spatial clustering such as DBSCAN. Clustering detected anomalies in 3D space could enhance the recognition of false positives, including reflections. Spatial clustering could allow to detect relocation of cluster centers, which could enable differentiation between reflections and true indications.

The algorithm proposed in the study does not address the issues with projection in complex scenarios where sight is obstructed by an object visible only by one of the sensors. For example, if parts of the environment are visible by thermal camera and invisible by LiDAR, it can result in omitting points on thermal map. To resolve this issue, further work regarding geometric verification is required. Using line-of-sight or acoustic ray-tracing could help distinguish true anomaly sources. However, these methods were not implemented in this work because they could significantly increase computational complexity. The presented projection algorithm allows for computationally efficient framework and can scale to large environments.

Another limitation of our work can be seen in acoustic mapping result. Acoustic map has a slightly inaccurate indication locations near pneumatic manipulators between acoustic heatmap and LiDAR point cloud due to presence of acoustic reflections in consecutive frames. This effect mitigates the precision of source localization across acoustic map. In future works, further measurements oriented in the area based on acoustic map should be performed to find the accurate location of anomaly source and not the region.

The method presented in the paper relies on the intersection of diagnostic indications, which can result in ignoring anomalies that are producing only one sensing modality. This approach was selected to reduce false alarms and ensure that found anomalies are highly reliable in noisy industrial environments. In the future, other sensor fusion methods, such as probabilistic fusion, could be implemented to handle non-overlapping indications better.

The method proposed in this paper uses fixed thresholds selected by the expert knowledge of the system to recognize anomalies. Instead of this approach, adaptive thresholding could be deployed. However, presented approach focuses on recognizing all anomalies above a significant minimum level. This approach allows to ensure consistent and interpretable indications within the map.

## 7. Conclusions and Future Work

In this work, we presented a multi-sensor mapping based on a combined LiDAR, thermal, and acoustic mapping of the environment. The proposed solution is suitable for complex 3D environments, and it enhances the cognitive abilities of a mobile platform for machinery diagnostics.

Spatial diagnostic representation is achieved by fusing three separate heterogeneous data sources. Combining different types of information into one consistent data format provides comprehensive information on severity level of anomalies and their locations within the environment. The multi-sensor mapping process enhances the usefulness of the data as composing a 3D map can contribute to increasing understanding of the environment.

Our approach allows creating a map of the environment with anomaly indications. Applying a threshold enables automatic differentiation of anomalies from parts of the environment. Repeatability analysis confirmed the credibility of obtained results. Acoustic measurements provided higher variation in standard deviation, which reflects the problem with noise generated in real industrial environments, which hinders the detection of acoustic anomalies. In future works, semantic segmentation could be additionally performed to recognize false positives.

The technique can be further improved to actively act on the identified anomalies and perform additional actions, such as in-depth analysis of the identified areas to determine in more detail the sources and the severity of the problem. The method can be also used to prevent failures of devices by sending the information on detection to the operator or by performing fixes with a manipulator that can be installed on top of the robot.

Our approach features partial autonomy. The robot moves independently across the environment by repeating a pre-recorded route and adjusting behavior to emerging obstacles including dynamic objects e.g. humans or other mobile robots. Despite the autonomous mission manner, the autonomy level could be increased in the future. Future works could improve the system by adjusting behavior to the environment or detections up to extending inspection to exploration of unknown terrains.

## Acknowledgment

The work has been financed from the research subsidy no. 16.16.130.942 by Department of Robotics and Mechatronics, AGH University of Krakow.

## References

1. Abdo S, Hovanec M, Korba P, Svab P. Utilization of NDT methods in aircraft maintenance. 2021 New Trends in Aviation Development (NTAD), Košice, Slovakia, IEEE: 2021: 9–12, https://doi.org/10.1109/NTAD54074.2021.9746505.
2. Belter D, Nowicki M R. Optimization-based legged odometry and sensor fusion for legged robot continuous localization. Robotics and Autonomous Systems 2019; 111: 110–124, https://doi.org/10.1016/j.robot.2018.10.013.
3. Bultmann S, Quenzel J, Behnke S. Real-time multi-modal semantic fusion on unmanned aerial vehicles with label propagation for cross-domain adaptation. Robotics and Autonomous Systems 2023; 159: 104286, https://doi.org/10.1016/j.robot.2022.104286.
4. Chagoya J, Patel S, Koduru C et al. Data Collection, Heat Map Generation for Crack Detection Using Robotic Dog Fused with FLIR Sensor. SoutheastCon 2024, Atlanta, GA, USA, IEEE: 2024: 824–829, https://doi.org/10.1109/SoutheastCon52093.2024.10500184.
5. Cheng L, Gao H, Sun W et al. An Integrated Method for Predictive State Assessment and Path Planning for Inspection Robots in Island-Based Unmanned Substations. Eksploatacja i Niezawodność – Maintenance and Reliability 2025. doi:10.17531/ein/203994, https://doi.org/10.17531/ein/203994.
6. Dobrev Y, Flores S, Vossiek M. Multi-modal sensor fusion for indoor mobile robot pose estimation. 2016 IEEE/ION Position, Location and Navigation Symposium (PLANS), Savannah, GA, IEEE: 2016: 553–556, https://doi.org/10.1109/PLANS.2016.7479745.

7. Fischer G K J, Bergau M, Gómez-Rosal D A et al. Evaluation of a Smart Mobile Robotic System for Industrial Plant Inspection and Supervision. IEEE Sensors Journal 2024; 24(12): 19684–19697, https://doi.org/10.1109/JSEN.2024.3390622.
8. Hamrani A, Rayhan M M, Mackenson T et al. Smart quadruped robotics: a systematic review of design, control, sensing and perception. Advanced Robotics 2025; 39(1): 3–29, https://doi.org/10.1080/01691864.2024.2411684.
9. Khattak S, Nguyen H, Mascarich F et al. Complementary Multi–Modal Sensor Fusion for Resilient Robot Pose Estimation in Subterranean Environments. 2020 International Conference on Unmanned Aircraft Systems (ICUAS), Athens, Greece, IEEE: 2020: 1024–1029, https://doi.org/10.1109/ICUAS48674.2020.9213865.
10. Khedekar N, Kulkarni M, Alexis K. MIMOSA: A Multi-Modal SLAM Framework for Resilient Autonomy against Sensor Degradation. 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Kyoto, Japan, IEEE: 2022: 7153–7159, https://doi.org/10.1109/IROS47612.2022.9981108.
11. Kubelka V, Reinstein M, Svoboda T. Improving multimodal data fusion for mobile robots by trajectory smoothing. Robotics and Autonomous Systems 2016; 84: 88–96, https://doi.org/10.1016/j.robot.2016.07.006.
12. Mączak J, Więcławski K, Szczurowski K. New approach of model based detection of early stages of fuel injector failures. Eksploatacja i Niezawodność – Maintenance and Reliability 2023. doi:10.17531/ein.2023.1.6, https://doi.org/10.17531/ein.2023.1.6.
13. Magalhaes P, Ferreira N. Inspection Application in an Industrial Environment with Collaborative Robots. Automation 2022; 3(2): 258–268, https://doi.org/10.3390/automation3020013.
14. Montes De Oca-Mora N J, Woo-Garcia R M, Juarez-Aguirre R et al. Mobile robot with failure inspection system for ferromagnetic structures using magnetic memory method. SN Applied Sciences 2021; 3(12): 853, https://doi.org/10.1007/s42452-021-04833-9.
15. Nam D V, Danh P T, Park C H, Kim G-W. Fusion consistency for industrial robot navigation: An integrated SLAM framework with multiple 2D LiDAR-visual-inertial sensors. Computers and Electrical Engineering 2024; 120: 109607, https://doi.org/10.1016/j.compeleceng.2024.109607.
16. Nooralishahi P, Ibarra-Castanedo C, Deane S et al. Drone-Based Non-Destructive Inspection of Industrial Sites: A Review and Case Studies. Drones 2021; 5(4): 106, https://doi.org/10.3390/drones5040106.
17. Nooralishahi P, Rezayiye R K, López F, Maldague X P V. PHM-IRNET: Self-training thermal segmentation approach for thermographic inspection of industrial components. NDT & E International 2023; 138: 102884, https://doi.org/10.1016/j.ndteint.2023.102884.
18. Ospina R, Itakura K. Obstacle detection and avoidance system based on layered costmaps for robot tractors. Smart Agricultural Technology 2025: 100973, https://doi.org/10.1016/j.atech.2025.100973.
19. Ozdemir K, Tuncer A. Navigation of autonomous mobile robots in dynamic unknown environments based on dueling double deep q networks. Engineering Applications of Artificial Intelligence 2025; 139: 109498, https://doi.org/10.1016/j.engappai.2024.109498.
20. Pfändler P, Bodie K, Crotta G et al. Non-destructive corrosion inspection of reinforced concrete structures using an autonomous flying robot. Automation in Construction 2024; 158: 105241, https://doi.org/10.1016/j.autcon.2023.105241.
21. Pi R, Cieślak P, Esteba J et al. Compliant Manipulation With Quasi-Rigid Docking for Underwater Structure Inspection. IEEE Access 2023; 11: 128957–128969, https://doi.org/10.1109/ACCESS.2023.3332486.
22. Sabry A H, Ungku Amirulddin U A B. A review on fault detection and diagnosis of industrial robots and multi-axis machines. Results in Engineering 2024; 23: 102397, https://doi.org/10.1016/j.rineng.2024.102397.
23. Segal A, Haehnel D, Thrun S. Generalized-ICP. Robotics: Science and Systems V, Robotics: Science and Systems Foundation: 2009. doi:10.15607/RSS.2009.V.021, https://doi.org/10.15607/RSS.2009.V.021.
24. Szczepanski R, Tarczewski T, Erwinski K. Energy Efficient Local Path Planning Algorithm Based on Predictive Artificial Potential Field. IEEE Access 2022; 10: 39729–39742, https://doi.org/10.1109/ACCESS.2022.3166632.
25. Toman R, Rogala T, Synaszko P, Katunin A. Robotized Mobile Platform for Non-Destructive Inspection of Aircraft Structures. Applied Sciences 2024; 14(22): 10148, https://doi.org/10.3390/app142210148.
26. Tsenis T, Tromaras A, Kappatos V. NDE Using Quadruped Robotic Platform for Renewable Energy Producing Infrastructures. 2024 IEEE 8th Energy Conference (ENERGYCON), Doha, Qatar, IEEE: 2024: 1–6, https://doi.org/10.1109/ENERGYCON58629.2024.10488808.
27. Wang B, Zhang L, Kim J. Fault Detection and Diagnosis of Three-Wheeled Omnidirectional Mobile Robot Based on Power Consumption Modeling. Mathematics 2024; 12(11): 1731, https://doi.org/10.3390/math12111731.

28. Zhang J, Yuan X, Wang C et al. RF-Fusion: Robust and fine-grained volumetric panoptic mapping system with multi-level data fusion for robots. Information Fusion 2025; 122: 103209, https://doi.org/10.1016/j.inffus.2025.103209.
29. Zhao J, Xiao Z, Bai B et al. Defect Recognition of Transmission Line Unmanned Aerial Vehicle Inspection Images Based on Cascade R-CNN Algorithm. Eksploatacja i Niezawodność – Maintenance and Reliability 2025, https://doi.org/10.17531/ein/207303.