IEEE Robotics & Automation Magazine (RAM) paper, presented at ICRA 2026, Vienna, Austria. Cite as RAM paper.

Real-Time Localization Scoring for Challenging Industrial Environments

Abdurrahman Yilmaz
Istanbul Technical University

Umut Dumandag
Bluepath Robotics

Aydin Cagatay Sari
Bluepath Robotics

Ismail Hakki Savci
Bluepath Robotics

Hakan Temeltas
Istanbul Technical University

***Abstract*—Autonomous Mobile Robots (AMRs) are revolutionizing industries by enhancing flexibility and efficiency, particularly in dynamic environments such as automotive manufacturing. These environments pose challenges due to their constantly changing layouts, unpredictable obstacles, and varying conditions, which impact the performance of localization systems. This paper presents a novel real-time localization scoring architecture to address these challenges by quantifying the confidence in a robot's positioning system. The proposed *Localization Score* improves map reconciliation, manages sensor interference, adapts navigation strategies, and enhances traffic coordination. Extensive experimental studies, including real-world deployment in an operational automotive production factory, demonstrate the robustness, accuracy, and adaptability of the developed *Localization Score* algorithm. The results showcase its potential to significantly enhance the operational efficiency and reliability of AMRs in industrial settings.**

1. INTRODUCTION

Recently, autonomous mobile robots (AMRs) have emerged as key players in various industrial and commercial settings, offering flexibility and efficiency in tasks ranging from manufacturing to logistics [1]. This surge in demand has coincided with a projected growth rate of 23% (compound annual growth rate), propelling the market to exceed 10 billion USD by 2028 [2]. However, deploying AMRs in dynamic environments (e.g., automotive manufacturing factories) poses significant challenges. These challenges result from the inherent complexity of real-world scenarios, where factors such as changing layouts, unpredictable obstacles, and varying environmental conditions can impact the performance of localization systems [3]. Addressing these challenges is crucial for ensuring the smooth operation of AMRs in dynamic environments, where failures in localization can lead to inefficiencies and delays.

Digital Object Identifier 10.1109/10.1109/MRA.2025.3584350

Localization robustness typically depends on three key factors: detection (sensor sensitivity), matching (alignment between map and measurements), and accuracy (localization error) [4]. Ensuring robustness in these aspects is crucial for maintaining reliable localization, especially in highly dynamic industrial settings.

The *Localization Score* is a metric that quantifies the confidence in a robot's positioning system. This score is pivotal in addressing these challenges by enabling lifelong localization [5], which helps maintain continuous and efficient operations without interruptions.

The *Localization Score* facilitates enhanced map reconciliation when environmental changes cause discrepancies between a robot's internal map and the actual surroundings, ensuring up-to-date navigation data [6]. It can also aid in managing sensor interference by triggering adjustments in sensor settings or switching to more robust sensors, thereby preserving the integrity of localization data. The score can guide adaptive navigation strategies to optimize pathfinding and mobility in scenarios involving physical obstacles or diverse terrains. Traffic coordination can be improved through dynamic adjustments based on the *Localization Score*, promoting efficient flow and interaction not only between robots but also between robots and human workers [7].

The *Localization Score*'s scalability and role in continuous improvement across multiple robots help refine localization algorithms and adapt to larger operational demands [8]. By enabling reliable and precise localization that adapts to various challenges in real-time, the *Localization Score* can significantly enhance the operational efficiency and robustness of automation systems in automotive manufacturing, thereby preventing knock-on production stoppages or disruptions and maximizing factory efficiency.

In this study, therefore, a real-time localization scoring architecture is presented. Numerous studies have explored localization confidence estimation methods, including map-measurement coherence metrics that enhance reliability alerts [9], search-based metrics designed for frequent environmental changes by maintaining a short-term map alongside a static map [10], and Bayesian classifiers-based robust reliability estimation function integrated with Monte Carlo Localization (MCL) [11]. While these methods contribute significantly to localization confidence estimation, they often lack the adaptability required for dynamic and complex settings. Based on the key factors affecting localization robustness, our proposed *Localization Score* dynamically evaluates detection, matching, and accuracy, offering a simple yet robust framework tailored for lifelong localization in industrial applications by dynamically assessing confidence in continuously changing environments.

The aims and key contributions of the study are

1) developing a simple but effective real-time localization scoring architecture that quantifies the confidence in a robot's positioning system, addressing dynamic environmental challenges;
2) ensuring lifelong localization reliability by detecting localization failures in real-time and triggering the necessary environment map updates based on observing localization failures and maintaining continuous system performance even in frequently changing layouts;
3) implementing the proposed methods in a real factory environment. Unlike most studies that rely solely on simulations or field tests in laboratory settings, this research includes practical implementation, validating the feasibility and effectiveness of the approach and demonstrating its potential for industrial applications.

The remainder of the study is organized as follows: Section 2 summarizes related works in this field and compares them with the presented real-time localization scoring system. Section 3 elaborates on the challenges encountered in automotive manufacturing and factory operations, emphasizing the importance of the localization scoring system in resolving them. Section 4 details the proposed localization reliability evaluation system. In the subsequent section, Section 5, we present the experimental setup and extensive field test results in various automotive production scenes, demonstrating the importance and performance of real-time localization scoring, followed by a comprehensive discussion regarding the utilization of lifelong localization. Finally, Section 6 concludes the article and discusses future work opportunities.

2. RELATED WORKS

The field of AMRs has seen significant advancements, particularly in localization systems designed to ensure robust and accurate positioning in dynamic environments [12]. Various methods have advanced AMR localization, particularly in improving robustness and reliability. For example, Garrote et al. [13] combined reinforcement learning with particle filters for map updates, introducing an overlap score based on laser measurements and map cells. However, small deviations in orientation critically impacted mapping accuracy. Complementing this work, Wang et al. proposed a switching approach to enhance localization reliability in complex environments by defining each sensor-based localization subsystem's confidence in its accuracy [14]. This study, on the other hand, focuses on developing a localization reliability assessment method rather than improving the localization reliability of the systems.

Boukerche et al. highlighted that localization system performance improves with reliability-driven data, comparing reliability-aware and unaware systems [15]. This study underscores the importance of integrating reliability metrics but does not propose specific mechanisms for dynamic environments.

Ferranti et al. introduced a lightweight confidence measure metric for visually estimated poses based on spatial distribution inliers and coverage scores [16]. Despite the innovation, visual localization systems are less prevalent than range-based systems for AMRs due to their susceptibility to environmental variations.

In statistical methods, Chen et al. proposed a measurement index of localization accuracy for Ultra-Wide Bandwidth (UWB) systems, demonstrating higher reliability with smaller confidence interval radii [17]. Anagnostopoulos et al. proposed the Dynamic Accuracy Estimation (DAE) method for RSSI-based fingerprinting localization methods [18]. Although promising, the performance of such methods falls short in challenging factory environments requiring precise robotic navigation and material handling. Similarly, Nguyen et al. used confidence scores to estimate the reliability of Wi-Fi fingerprinting localization [19]. However, UWB, RSSI, and Wi-Fi fingerprint systems are inconvenient for highly reflective industrial environments like automotive factories [20].

Jing et al. evaluated localization system performance through accuracy, availability, continuity, and integrity metrics. Their method, tested on GNSS and inertial systems, highlights different types of localization errors but is more suited for outdoor applications than dynamic indoor factory environments [21].

Moon et al. focused on a failure recovery scheme for reliable localization using an Observation Likelihood Model [22]. Their approach addressed the design of the observation likelihood model, failure detection, and recovery. By employing semi-global localization, they effectively managed wheel slippage-like localization failures. However, their evaluation was limited to static or quasi-static environments, which may not reflect the complexities of dynamic settings. Similarly, He et al. enhanced the Adaptive Monte Carlo Localization (AMCL) by integrating the Iterative Closest Point (ICP) for localization confidence estimation [23]. This method allows particle swarms to update and converge faster based on ICP registration.

Other studies, such as Nobili et al. [9], emphasized map-measurement coherence for reliability evaluation, while Hroob et al. [24] extended confidence estimation to 3D environments using CNNs. While useful for creating alerts, the map-observation consistency metrics alone may not fully ensure accurate localization or capture localization failures in the quasi-static, instantaneously static but changing over time (e.g., warehouses), or dynamic environments. While the map-observation coherence metric is a necessary component for evaluating localization reliability, it alone cannot fully capture the complexities of dynamic environments. In our localization scoring mechanism, we incorporate this metric alongside additional measures, such as particle weight distributions and accuracy evaluations, to provide a more comprehensive assessment of localization reliability.

Addressing the challenges posed by dynamic objects, Demir et al. developed methods using Normal Distributions Transform (NDT) and LiDAR scan matching confidence estimation [25]. They proposed metrics like fitness score and transformation probability, which can be affected by dynamic objects and local minima. Their approach also included mean elevation angle estimation to mitigate the effects of dynamic objects. These methods share similarities with map measurement overlap distance-based evaluation, but require further refinement for dynamic environments.

Villacres et al. presented an RL-based resampling method for PF-based localization to counteract kidnapping [26]. Their method, however, is not designed to handle local localization failures common in dynamic industrial settings like factories.

PF variants are one of the most common frameworks used for indoor localization of the AMRs [27], [28]. Therefore, Akai et al. [11] enhanced Monte Carlo Localization (MCL) by integrating reliability estimation with a class-conditional measurement model. Their method improved robustness to environmental changes and enabled failure detection, but required accurate initial estimates and incurred computational overhead. Our approach builds on these methods by incorporating particle weights and distributions into a real-time scoring mechanism.

In summary, while many studies have made significant contributions to AMR localization, most are designed for static or quasi-static environments and struggle to maintain robustness in highly dynamic and complex settings. Our proposed real-time localization scoring architecture bridges this gap by providing a simple, scalable, and adaptive solution for lifelong localization in ever-changing environments.

IEEE Robotics & Automation Magazine (RAM) paper, presented at ICRA 2026, Vienna, Austria. Cite as RAM paper.

3. CHALLENGES IN AUTOMOTIVE MANUFACTURING AND FACTORY OPERATIONS

Mobile robots in automotive factories face numerous operational challenges that can significantly impact their efficiency and reliability. These challenges stem from the inherent characteristics of industrial environments and the specific demands of automotive production processes. The proposed *Localization Score* addresses these challenges by offering a quantitative measure of the confidence level in the robot localization system.

Challenge 1: Dynamic Environmental Changes: Automotive factories are subject to frequent changes in layout and operation zones due to varying production requirements. This dynamic nature can lead to discrepancies between the robot's internal map and the actual environment, leading to navigation errors.

Solution 1: By continuously monitoring the *Localization Score*, the system can detect when the robot's map becomes misaligned with the environment due to dynamic changes. This prompt detection allows for immediate corrective actions, such as map updates or on-the-fly recalibration of the localization system.

Challenge 2: Interference and Sensor Noise: The dense deployment of machinery and electronic devices in automotive factories generates significant electromagnetic interference, which can disrupt the signal integrity of the sensors used in robot localization systems. Additionally, heavy welding and metalworking equipment produce optical and acoustic noise, which can confuse sensors like LiDAR or cameras.

Solution 2: In environments with high electromagnetic and sensory noise, a dropping *Localization Score* can trigger enhanced filtering techniques or the switch to more robust sensors that are less susceptible to interference, thus maintaining localization integrity.

Challenge 3: Physical Obstacles and Varied Terrain: Robots often have to navigate around temporary obstacles like moved machinery, stockpiles of parts, or other robots and human workers. Furthermore, different areas of the factory may have varying floor conditions—from smooth concrete to grated metal surfaces—which can affect mobility and sensor readings.

Solution 3: The *Localization Score* can guide adaptive navigation strategies. For instance, if the score drops upon entering an area with complex terrain, the system can switch to alternative navigation modes that rely more on robust sensors or different types of sensors better suited to those specific environmental conditions.

Challenge 4: High Traffic Areas: Some zones in automotive factories experience high traffic flow, with both robots and human workers sharing the same pathways. Managing such interactions without collisions or unnecessary stoppages requires highly reliable and precise localization to ensure smooth and continuous robot operations.

Solution 4: In high-traffic areas, maintaining a high *Localization Score* is critical for avoiding collisions and ensuring efficient traffic flow. The score can be used to dynamically adjust the robot's speed, path, and operational timing to synchronize with surrounding traffic, thereby enhancing overall workflow efficiency.

Challenge 5: Scale and Complexity of Operations: As factories expand and incorporate more automated systems, the scale and complexity of operations increase. This growth requires a scalable localization system that can maintain high levels of accuracy and reliability across a larger array of robots and a more extensive operational environment.

Solution 5: The *Localization Score* allows for the aggregation of localization data across multiple robots, facilitating a collective improvement in localization strategies. This aggregation helps in refining the localization algorithms and adjusting the system to scale up to larger operational demands without losing accuracy or efficiency.

By directly addressing each of these challenges, the *Localization Score* not only ensures that the robots maintain high operational efficiency but also contributes to the overall robustness and reliability of the automation systems within the automotive factory environment.

4. REAL-TIME LOCALIZATION SCORING ARCHITECTURE

Evaluation of the localization system output is as vital as the localization algorithm itself for real-world scenarios such as those in automotive production lines. Beyond conventional approaches, the developed real-time localization scoring architecture in this study considers two main outputs. The first one involves inferences drawn from the weights and distributions of particles used in PF-based localization approaches to determine the position. The latter, pivotal metric is the incorporation of map measurement consistency for evaluating the reliability of localization outputs.

Particle Filter (PF) Data Information

PFs are widely favored in localization due to their ability to handle non-linear distributions, manage uncertainty, and adapt to diverse motion behaviors, rendering them ideal for real-time tasks. When assessing localization reliability using particle filter-based methods, attention is typically directed towards the distribution and weights of particles. Additionally, particles can be categorized into clusters based on their distributions, enhancing the granularity of localization analysis.

Weight analysis for particles and clusters: Each particle and cluster in a PF framework is assigned a weight representing its contribution to the estimated position. The weights can be evaluated through two primary metrics:

- **Maximum Weight:** Indicates the most significant particle or cluster. Higher maximum weights suggest a more confident localization.
- **Logarithmic Weight:** Measures the spread of weights among particles or clusters, providing insights into uncertainty. Lower logarithmic weights imply higher uncertainty.

From here on, let the number of particles be N and the number of clusters be C, and the weights of n^{th} particle and c^{th} cluster are defined as $\omega_p^{[n]}$ and $\omega_{cl}^{[c]}$ respectively. According to these definitions, the following particle and cluster weight equations can be derived:

$$\omega_p^{\max} = \max(\omega_p^{[n]}) \quad ; \quad \omega_p^{\max_{sc}} = \omega_p^{\max} \times N \tag{1}$$

$$\omega_{cl}^{\max} = \max(\omega_{cl}^{[c]}) \quad ; \quad \omega_{cl}^{\max_{sc}} = \omega_{cl}^{\max} \times C \tag{2}$$

$$\omega_p^{\log} = -\sum_{n=1}^{N} \omega_p^{[n]} \times \log(\omega_p^{[n]}) \quad ;$$
$$\omega_p^{\log_{sc}} = \frac{\omega_p^{\log}}{\log(N)} \tag{3}$$

$$\omega_{cl}^{\log} = -\sum_{c=1}^{C} \omega_{cl}^{[c]} \times \log(\omega_{cl}^{[c]}) \quad ;$$
$$\omega_{cl}^{\log_{sc}} = \frac{\omega_{cl}^{\log}}{\log(C)} \tag{4}$$

where $\omega_p^{\max}$ is the maximum weight of particles, $\omega_{cl}^{\max}$ is the maximum weight of clusters, $\omega_p^{\log}$ is logarithmic weight of particles, $\omega_{cl}^{\log}$ is logarithmic weight of clusters, $\omega_p^{\max_{sc}}$ is scaled maximum weight of particles, $\omega_{cl}^{\max_{sc}}$ is scaled maximum weight of clusters, $\omega_p^{\log_{sc}}$ is scaled logarithmic weight of particles, and $\omega_{cl}^{\log_{sc}}$ is scaled logarithmic weight of clusters. Scaled versions of these metrics normalize their values for better comparability.

When the uncertainty in localization increases, both the scaled maximum weight of clusters and particles increase. Consequently, if the specified boundaries are exceeded, it can be concluded that the localization output is unreliable. Conversely, the logarithmically scaled weight parameters decrease as uncertainty increases. Therefore, when the actual output falls below the specified threshold, we may infer that the localization system output is unreliable. Therefore, the weight reliability component (w) for the PF is computed based on these scaled maximum/logarithmic particle/cluster metrics as follows,

$$w = s_w \left(\frac{\sigma_p^{\max}}{\omega_p^{\max_{sc}}} + \frac{\sigma_{cl}^{\max}}{\omega_{cl}^{\max_{sc}}} + \frac{\omega_p^{\log_{sc}}}{\sigma_p^{\log}} + \frac{\omega_{cl}^{\log_{sc}}}{\sigma_{cl}^{\log}} \right) \tag{5}$$

where $\sigma_p^{\max}$, $\sigma_{cl}^{\max}$, $\sigma_p^{\log}$, and $\sigma_{cl}^{\log}$ represent specified thresholds for scaled maximum/logarithmic particle/cluster weight values. On the other hand, s_w is a scale constant for the weight reliability component to normalize the output of the overall measure.

Distribution analysis for particles: The distribution of particles can be investigated through the covariance matrix, which provides information about the particles. In 2D flat environments, such as factories, the components of the covariance matrix representing translation in the x and y directions and orientation variability around the z-axis characterize the distribution of particles. The uncertainty in localization increases as the covariance matrix components grow larger, implying a more scattered particle distribution. Conversely, smaller values indicate particles tightly clustered around the estimated pose.

To provide a single reliability measure, our localization scoring system considers the combined magnitude (hypotenuse) of these covariance components to compute the particle distribution component (d), such that

$$d = \exp\left(-s_d \frac{\sqrt{c_{x_{tr}}^2 + c_{y_{tr}}^2 + c_{z_{rot}}^2}}{\sigma^{\text{cov}}} \right) \tag{6}$$

where $c_{x_{tr}}$, $c_{y_{tr}}$, and $c_{z_{rot}}$ denote covariance matrix components for translation in the x and y directions

IEEE Robotics & Automation Magazine (RAM) paper, presented at ICRA 2026, Vienna, Austria. Cite as RAM paper.

[cm] and rotation around the z-axis [°], respectively. The use of these units ensures that both error types are represented on a consistent scale. In industrial applications, sub-centimeter and sub-degree accuracy are often the targets, making these units relevant. Similar to the weight analysis sub-section, σ^{cov} represents the specified threshold for the magnitude of covariance matrix components. On the other hand, s_d is a constant variable for the distribution reliability component to transfer the range of the metric to the desired interval. This score decreases as uncertainty increases, ensuring it aligns with intuitive expectations of reliability.

Map-Measurement Consistency

Next, a localization reliability metric was developed based on the consistency between the environment map and AMR laser measurements, independent of the localization method utilized. This metric aims to examine how well the environment map aligns with laser measurements relative to the determined pose of AMR. Two parameters can be investigated for map-measurement consistency. One parameter indicates the percentage of overlap between the map and measurements, while the other measures the average distance between overlapping points. A low overlap rate may suggest the need to update the relevant portion of the map, whereas the distance between overlapping points provides insight into the consistency of the overlap computation. Therefore, in the reliability metric, the inference is drawn from this average distance data, where a lower distance corresponds to higher localization performance. In contrast, an increase in distance corresponds to a decrease in reliability. Consequently, the following expression can be employed as the map-measurement consistency metric (c) within localization reliability analysis.

$$c = \exp\left(-s_c \frac{\overline{d}}{\sigma^{\text{cons}}}\right) \tag{7}$$

where σ^{cons} represents the specified threshold for the map-measurement consistency component and s_c constant ensures the consistency reliability component within specified ranges. Moreover, $\overline{d}$ denotes the average distance between the overlapping points on the map and laser measurements. For that purpose, synthetic laser measurements[1] are generated for the estimated pose and compared with the current laser measurements of the AMR. To compute the overlapping points between the map and measurements, the overlap rate calculation approach proposed by Yilmaz et al. [29] is employed. Similar expression in (7) is also employed by Zhu et al. [30]. While Zhu et al.'s method focuses on classifying observations as static or semi-dynamic to adjust observation weights dynamically, our approach directly quantifies map-measurement consistency to assess overall localization reliability. Moreover, Zhu et al.'s two-map approach contrasts with our use of a single map, aimed at realigning the robot's perception with environmental changes. These distinctions underline complementary but fundamentally different methodologies.

[1] https://github.com/yilmazabdurrah/synthetic_laser_measurement_MATLAB

Computation of Overall Localization Score

Expressions have been derived for calculating the values of the w metric obtained from PF particle and cluster weights, the d metric calculated based on the distribution of particles, and the c metric derived from map-measurement consistency. Utilizing these metrics, the overall localization reliability score (r) can be straightforwardly computed as follows by the weighted sum of defined metrics.

$$r = \omega_w w + \omega_d d + \omega_c c \tag{8}$$

The use of ω_w, ω_d, and ω_c weights allows for the maintenance of the localization reliability metric within the range of 0 to 1, as well as the ability to adjust the influence of any metric on the *Localization Score* as desired, increasing or decreasing its impact.

By synergizing these components, our algorithm ensures adaptability and versatility to localization failures and flexibility and robustness for complex environments with dynamic features.

5. EXPERIMENTAL STUDIES

The experimental studies section examines the performance and reliability of the localization scoring algorithm using T2000 robots of Bluepath Robotics. With their advanced navigation and perception systems, these robots are tested in different environments to assess how well the algorithm works in real-world applications. Each scenario is designed to address specific challenges, from static controlled environments to highly dynamic automotive manufacturing sites, ensuring a thorough algorithm evaluation. The results show the robustness, accuracy, and adaptability of the developed localization score algorithm, offering

Figure 1. Bluepath Robotics T2000 used in Ford Otosan's Production Factory located at Gölcük-Kocaeli/Turkey. For more details please visit Bluepath Robotics.

valuable insights for its use in industrial settings.

Experimental Setup

In this study, T2000 robots of Bluepath Robotics, a spin-off company of Ford Otosan, are used (see Fig. 1). The T2000 is designed for heavy-duty industrial applications, supporting a maximum payload of 2000 kg. The robot has a maximum speed of 1.5 m/s and is equipped with a Sick Laser Scanner and a RealSense RGBD camera for perception. It features high docking precision, supports remote control with safe on/off functionality, and complies with VDA5050 standards. It uses a hybrid navigation system combining natural navigation based on a particle filter localization approach on grid maps and line tracking, achieving a position accuracy of ±1 cm. The robot has 3D object/obstacle detection, obstacle avoidance, high-precision trajectory tracking, dynamic map updating, and enhanced perception features. It is actively used in Ford Otosan's automotive factory, demonstrating its practical application and reliability in a demanding industrial environment.

Application Scenarios

Application scenarios are thoroughly selected and fine-tuned to address the particular requirements of automotive applications. The purpose of each scenario is to evaluate the localization score algorithm in conditions that closely mimic real-world automotive settings. By customizing these scenarios to automotive contexts, we ensure that the algorithm is assessed under conditions directly relevant to its intended use, providing a comprehensive evaluation of its resilience and robustness.

In Scenario 1, the localization score is tested and tuned in a controlled Pre-Landing Area without dynamic elements, ensuring any score variations are due to the algorithm's performance.

Scenario 2 introduces a moderately dynamic real automotive manufacturing site. Here, the map changes and the tuned localization score from Scenario 1 are tested to observe consistent score drops and identify genuine declines in localization accuracy.

Scenario 3 presents a highly dynamic industrial environment, examining the reliability and behavior of the localization score in extreme conditions based on the findings of Scenarios 1 and 2.

In Scenario 4, the localization score triggers an in-house map update algorithm when score drops are detected. This updated map, once deployed, improves the localization score, creating a robust feedback mechanism for lifelong localization.

Collectively, these scenarios ensure the localization score algorithm is tested across varying environments, from static to highly dynamic, establishing a reliable and adaptive system capable of excellent performance in real-world scenarios.

The parameters described in (5), (6), and (7) (see Sect. 4) are selected and universally applied across all scenarios as follows: $s_w \cdot \sigma_p^{\max} = 0.2625$, $s_w \cdot \sigma_{cl}^{\max} = 0.2501$, $s_w/\sigma_p^{\log} = 0.1563$, $s_w/\sigma_{cl}^{\log} = 0.2143$, $s_d/\sigma^{\text{cov}} = 6.9315$, and $s_c/\sigma^{\text{cons}} = 2.7726$.

These parameters were first fine-tuned to balance the effects of uncertainties at reasonable levels. On the other hand, the weights in (8) are set as $\omega_w = \omega_d = \omega_c = 1/3$ to equalize the impacts of the sub-localization scoring metrics. The overall localization reliability score (r) is interpreted by analyzing how its values vary concerning environmental conditions across different scenarios. The categorization of this reliability score will be elaborated on within the scenario descriptions.

Finally, to showcase the advantages of our localization scoring architecture, we conducted both qualitative and quantitative comparisons against three state-of-the-art reliability approaches. These are Akai's *ALS Score* [11], Costa's *Dynamic Localization Score* [31], and Valencia's dual timescale *NDT-MCL Score* with Weber's implementation [10], [32].

IEEE Robotics & Automation Magazine (RAM) paper, presented at ICRA 2026, Vienna, Austria. Cite as RAM paper.

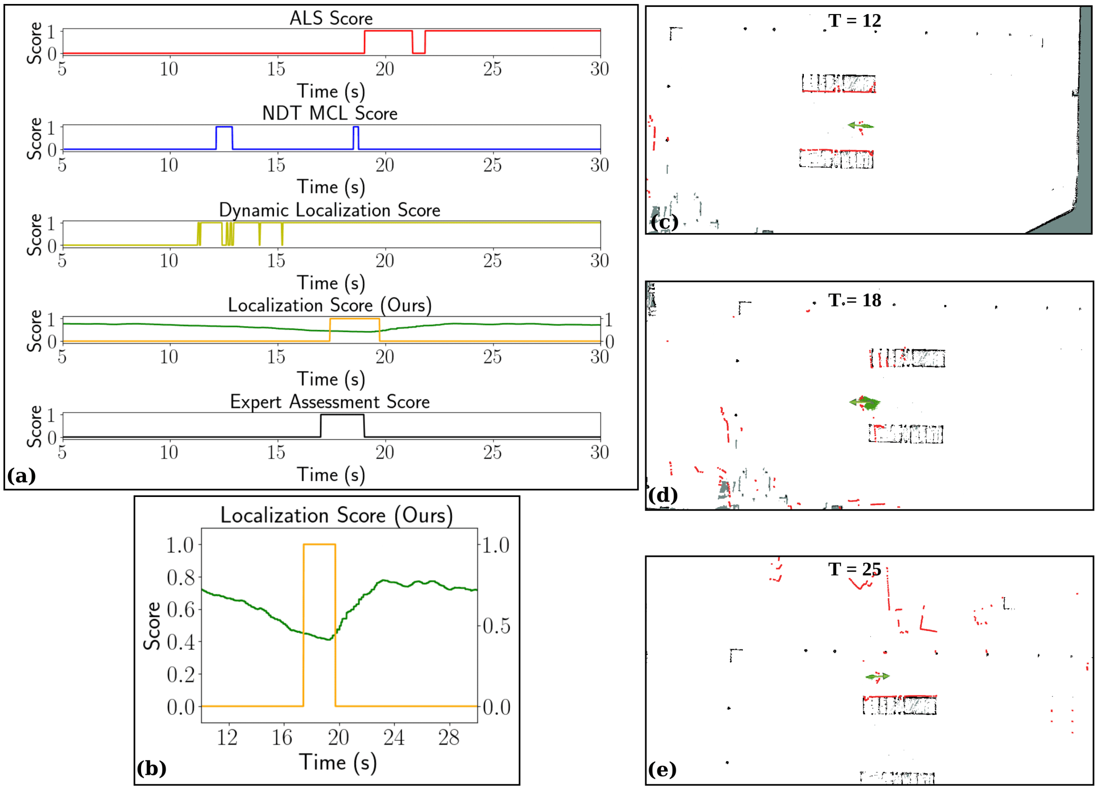

Figure 2. Localization scoring results in a controlled environment, specifically in the Pre-Landing area of the factory: (a) Localization scores vs. time for the compared algorithms (Value "0" indicates reliable localization, while "1" indicates unreliable localization), (b) close-up view of the section where our *Localization Score* system detects localization failure (the green line represents the raw output of our *Localization Score* system, and the yellowish line indicates the threshold-based 0-1 correspondence, where "0" corresponds to reliable localization and "1" corresponds to unreliable localization, based on the green line), (c-e) snapshots at different time instances, including occupancy grid map section of the environment, current laser measurements (red), and localization system estimates (green).

Scenario 1: Proof of Concept in a Controlled Static Environment

In this scenario, the primary focus is on evaluating the performance of the *Localization Score* algorithm in a static and controlled environment. The absence of dynamic elements ensures that the algorithm's efficacy in accurately determining the robot's position can be measured without interference from external variables.

The Traffic Management System (TMS) plays a crucial role in this setup. It issues specific orders to the robot, guiding it through a series of waypoints within the Pre-Landing Area. These waypoints form loops that the robot must navigate, allowing for repeated and consistent testing conditions.

This scenario aims to establish a baseline performance metric for the *Localization Score* algorithm by focusing on a static environment. The insights gained from this controlled testing will provide a foundation for subsequent evaluations in more dynamic and complex environments.

The observations on the Pre-Landing area have been indicated in Fig. 2. In this scenario, the AMR moved autonomously from one point to another while the reliability of the estimated pose was computed simultaneously. As in Fig. 2(b), our tests demonstrated that sharp decreases and/or lower values in the localization score corresponded to unreliable localization estimation; in other words, the AMR may have lost its pose at those times. For instance, while AMR poses are tracked properly at time instances shown in Fig. 2(c) and (e), the AMR lost its pose at the time instance highlighted in Fig. 2(d). To visualize the superior performance of the developed algorithm, we compared it with the state-of-the-art reliability approaches in

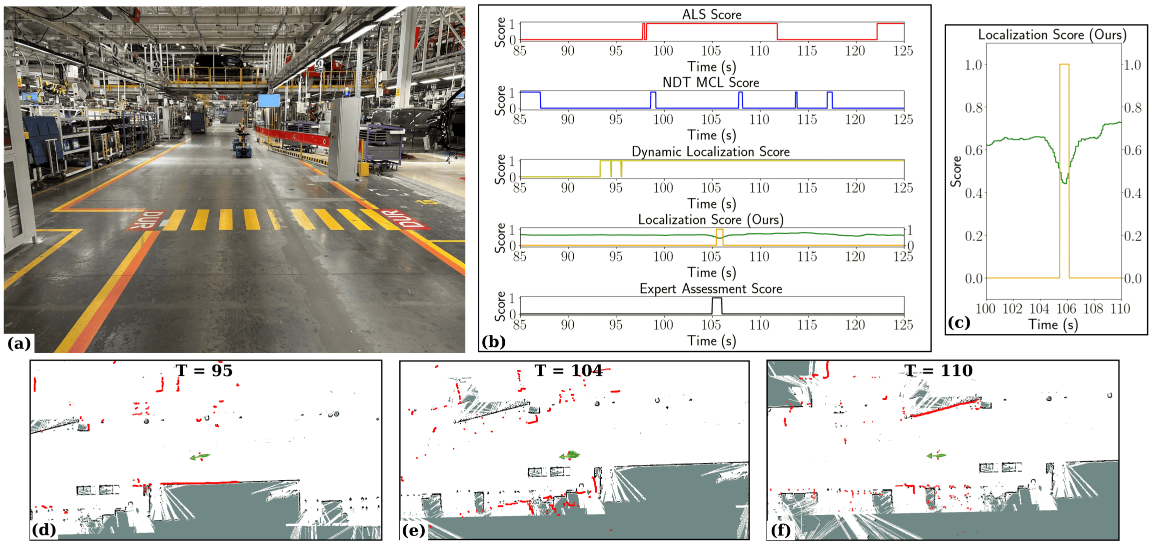

Figure 3. Localization scoring results in a real automotive production environment: (a) scene from real automotive production environment, (b) Localization scores vs. time for the compared algorithms (Value "0" indicates reliable localization, while "1" indicates unreliable localization), (c) close-up view of the section where our *Localization Score* system detects localization failure (the green line represents the raw output of our *Localization Score* system, and the yellowish line indicates the threshold-based 0-1 correspondence, where "0" corresponds to reliable localization and "1" corresponds to unreliable localization, based on the green line), (d-f) snapshots at different time instances, including occupancy grid map section of the environment, current laser measurements (red), and localization system estimates (green).

Fig. 2(a). Additionally, feedback from human operators, *Expert Assessment Score*, is also extracted as a ground truth (GT) localization reliability measure. In this figure and from now on, score "0" represents a "reliable" estimated pose, while score "1" denotes unreliable cases. The results indicate that the methods, except ours, often fail to properly estimate unreliable cases, highlighting the robustness and accuracy of our algorithm in such scenarios.

Observations from Scenario 1 indicated that a localization score of 0.74 and above corresponds to an almost perfect score. This threshold was refined in Scenario 1 and subsequently fed into Scenario 2 to determine what score represents unreliable cases.

Scenario 2: A Real Automotive Production Workshop

In this scenario, the *Localization Score* algorithm is tested in a real-world setting with moderate levels of dynamism, such as moving equipment and personnel. This setup presents a significant challenge due to mismatched LiDAR points on the map, which the robots must accurately navigate.

The robots are tasked with transporting automotive parts by pulling a dolly through a series of waypoints. These waypoints, provided by the TMS, form loops that the robots must follow. This setup ensures that the robots can be tested repeatedly under consistent yet moderately dynamic conditions.

By focusing on a real automotive manufacturing site with its inherent challenges, this scenario aims to evaluate the robustness and reliability of the *Localization Score* algorithm in a more complex and realistic environment. The insights gained from this testing will be critical for effectively refining the algorithm to handle real-world operational conditions.

Two scenarios from the real automotive production environment were demonstrated in Figs. 3 and 4. Similar to the Pre-Landing area case, the localization scoring strategy developed in this study predicts localization failures more robustly than the state-of-the-art method. As observed in Fig. 3(c) and Fig. 4(c), there is a sharp decline in the localization score during the period when AMR is lost. The developed method was able to detect the issue promptly in pose estimation.

In this scenario, it was observed by human operators that a localization score below 0.6 indicated a degradation in localization performance. This thresh-

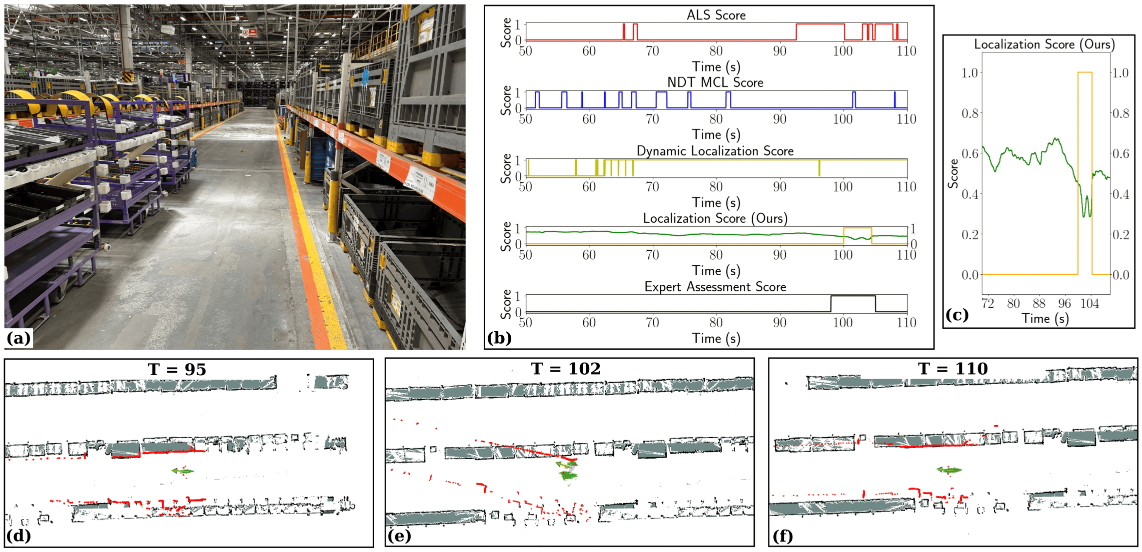

Figure 4. Localization scoring results in a real automotive production environment: (a) scene from real automotive production environment, (b) Localization scores vs. time for the compared algorithms (Value “0” indicates reliable localization, while “1” indicates unreliable localization), (c) close-up view of the section where our *Localization Score* system detects localization failure (the green line represents the raw output of our *Localization Score* system, and the yellowish line indicates the threshold-based 0-1 correspondence, where “0” corresponds to reliable localization and “1” corresponds to unreliable localization, based on the green line), (d-f) snapshots at different time instances, including occupancy grid map section of the environment, current laser measurements (red), and localization system estimates (green).

old was used as a benchmark and fed into Scenario 3. Scores between 0.6 and 0.74 were classified as “Good”.

Scenario 3: Highly-Dynamic Automotive Manufacturing Scene

In this scenario, the *Localization Score* algorithm is tested under highly dynamic conditions in a real-world automotive manufacturing site shown in Fig. 5(a). The presence of moving objects, particularly forklifts, poses significant challenges. These objects frequently pass by the robots, blocking considerable amounts of LiDAR points and increasing the number of mismatched points on the map.

The robots are tasked with transporting automotive parts by pulling a dolly through a series of waypoints provided by the TMS. The primary objective of this test is to evaluate the robustness and effectiveness of the *Localization Score* algorithm in a highly dynamic environment. By subjecting the algorithm to real-world operational challenges, such as frequent obstructions and mismatched LiDAR points, this scenario aims to validate its performance and reliability under extreme conditions.

As demonstrated by the *Localization Score* variation graphs in Fig. 5(b,c) and the corresponding scenes for different time instances in Figs. 5(d-f), localization score values often remained at low levels due to pose estimation difficulty in dynamic fields. It was observed that when the localization score dropped below 0.5, the robot was nearly lost. A score between 0.55 and 0.6 was classified as “Critical” and a score between 0.5 and 0.55 was categorized as “Marginal” while a score below 0.5 was allocated as “Lost”.

At t $\approx$ 90 (Fig. 5(d)), the AMR was still able to maintain localization despite noticeable mismatches between the LiDAR measurements and the map. Although environmental changes, such as moving obstacles, introduced inconsistencies, the system still classified localization as “Marginal” because the AMR was able to follow its expected path. At t $\approx$ 93 (Fig. 5(e)), a vehicle with a dolly passed directly in front of the AMR, significantly reducing the number of LiDAR matching points. This obstruction increased localization uncertainty, causing the localization score to drop below the 0.5 threshold, entering the “Lost” class. The system could no longer accurately determine the robot’s position due to the moving object

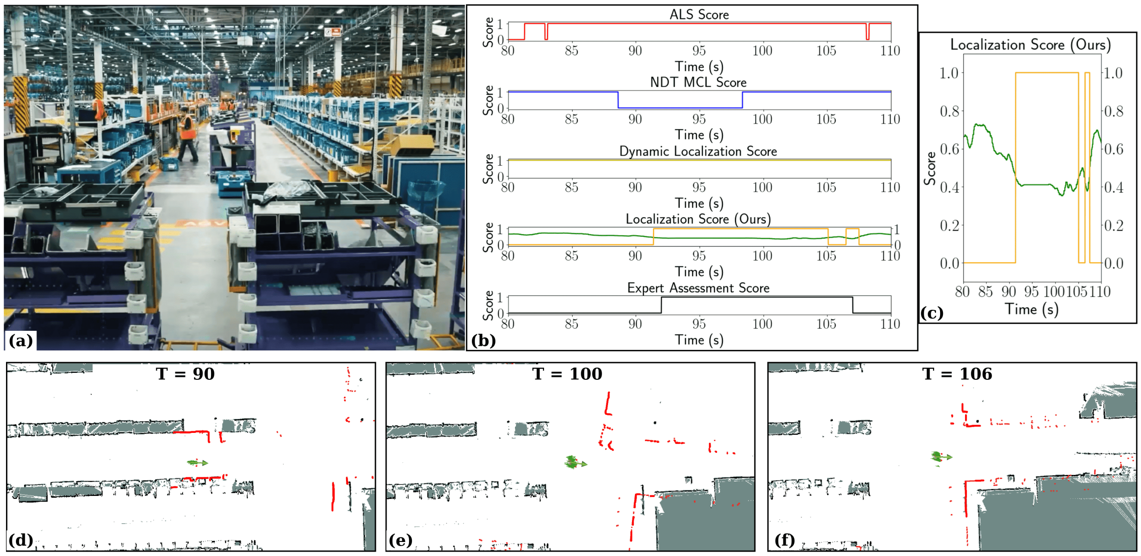

Figure 5. Localization scoring results in a highly-dynamic real automotive production environment: (a) scene from real automotive production environment, (b) Localization scores vs. time for the compared algorithms (Value "0" indicates reliable localization, while "1" indicates unreliable localization), (c) close-up view of the section where our *Localization Score* system detects localization failure (the green line represents the raw output of our *Localization Score* system, and the yellowish line indicates the threshold-based 0-1 correspondence, where "0" corresponds to reliable localization and "1" corresponds to unreliable localization, based on the green line), (d-f) snapshots at different time instances, including occupancy grid map section of the environment, current laser measurements (red), and localization system estimates (green).

blocking key reference points in the environment. At t $\approx$ 105 (Fig. 5(f)), after the forklift moved away, the localization system gradually regained accuracy as more LiDAR matching points became available. The localization score increased again, and the system transitioned back to the reliable state around t = 108.

The results clarify that the localization scoring algorithm developed effectively determines localization system reliability. For such cases, to improve the performance, the map parts can be updated.

Scenario 4: Updating Environment Map Based on Localization Scoring

The motivation behind this scenario lies in demonstrating that the proposed *Localization Score* is robust in detecting instances where the robot's localization begins to deteriorate. This capability ensures timely intervention by triggering the map update algorithm before navigation failures occur, thereby enabling reliable robot performance even in highly dynamic and ever-changing real-world conditions.

For this purpose, the *Localization Score* algorithm is continuously monitored, and specific locations where the localization score decreases are identified and recorded. These drops in the score trigger the use of an in-house developed map update algorithm. To trigger the map update algorithm, we selected a score of 0.55 as the lowest critical score. This approach ensures the robot does not reach the "Lost" state, maintaining localization accuracy and preventing navigation failures.

The updated map is then deployed to the robots as shown in two specific workshop cases (see Figs. 6 and 7). Following the deployment of the new map, the *Localization Score* is reassessed. The results show a significant improvement in the localization score, indicating that the updated map aligns more closely with the real-world conditions.

Quantitative Comparisons on Localization Scoring System

The *Expert Assessment Score* is used as ground truth to evaluate the performance of the localization scoring systems. This score is assigned by an expert based on the reliability of the localization and is derived from a bag recording in cases where localization is found to be unreliable. The assessment considers map matching, the distribution of particle output by the

IEEE Robotics & Automation Magazine (RAM) paper, presented at ICRA 2026, Vienna, Austria. Cite as RAM paper.

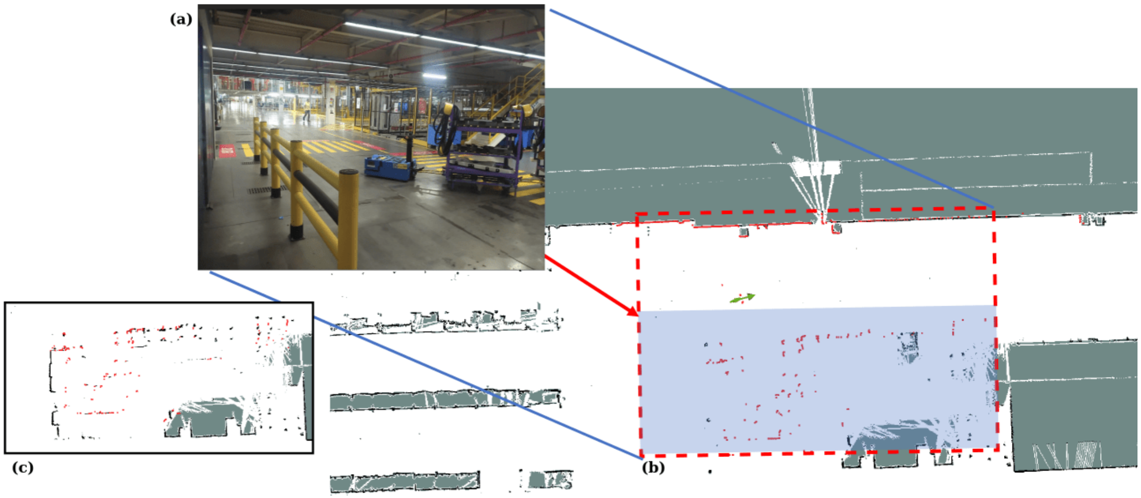

Figure 6. Real automotive production workshop employed for field tests: (a) Scene from the workshop, (b) occupancy grid map of the field, and (c) updated map according to *Localization Score* feedback.

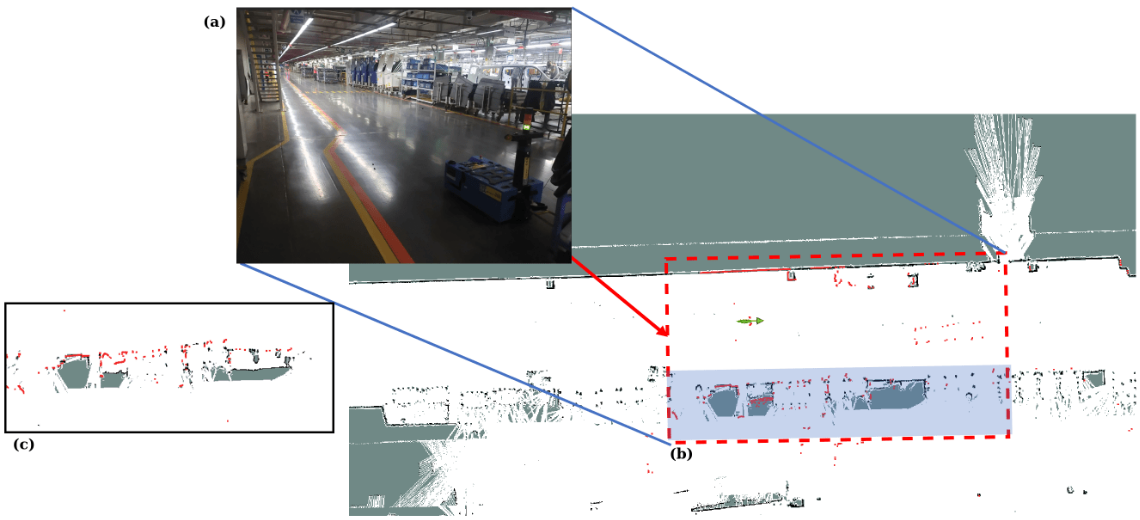

Figure 7. Real automotive production workshop employed for field tests: (a) Scene from the workshop, (b) occupancy grid map of the field, and (c) updated map according to *Localization Score* feedback.

AMCL algorithm, and verification through the camera feed from the robot. The evaluation employs the *F1*-score metric, an effective tool for assessing the consistency of a model. Here, "reliable" and "unreliable" states of localization systems' output are assumed as two classes. The overall localization scoring performance of the methods compared over the scenarios is demonstrated in Table 1. The results reveal that our *Localization Score* can detect reliable outputs and localization failures with more than 90% accuracy for even highly dynamic industrial scenes.

For our localization scoring system, the following values, summarized in Table 2, were used as reliable/unreliable class triggers. The "Lost" class is designated as *unreliable*, whereas other classes are considered *reliable* for all scenarios performed.

Table 1. Performance Scores Across Different Scenarios

Scenario	*F1*-Score of			
	ALS Score	NDT-MCL Score	Dynamic Localization Score	Localization Score (Ours)
1 in Fig. 2	0.5079	0.8980	0.3389	**0.9505**
2-a in Fig. 3	0.6282	0.8791	0.2531	**0.9664**
2-b in Fig. 4	0.8336	0.7443	0.3629	**0.9047**
3 Fig. 5	0.5559	0.7592	0.5049	**0.9182**

While the proposed localization scoring system demonstrates strong reliability in dynamic industrial environments, its applicability is currently focused on

Table 2. *Localization Score* Classification Based on Human Experts' Assessments

Investigated Scenario	*Localization Score* Range	*Localization Score* Class
1	0.74 - 1.0	Perfect
2	0.6 - 0.74	Good
3	0.55 - 0.6	Critical
3	0.5 - 0.55	Marginal
3	0.0 - 0.5	Lost

particle filter-based localization approaches and grid maps. Recent papers, such as Wilbers et al. (2019) [33], and the Robot Operating System (ROS) community continue to endorse these methods as state-of-the-art for pure localization methods. Notably, particle filter localization and grid maps remain the default map-based localization systems in the navigation stack of ROS1[2] and ROS2[3] frameworks. While SLAM systems are well-suited for simultaneous mapping and localization, their high computational load and limited reliability advantages in highly dynamic environments make them less ideal for real-time industrial applications.

6. CONCLUSION

The developed real-time localization scoring architecture effectively addresses the unique challenges posed by dynamic environments in automotive manufacturing. By leveraging metrics such as PF weight analysis, particle distribution, and map-measurement consistency, the *Localization Score* provides a reliable measure of localization confidence. Experimental results across various scenarios, from controlled environments to highly dynamic production floors, underscore the robustness and accuracy of the proposed system. The *Localization Score* not only facilitates adaptive and resilient navigation strategies but also contributes to the overall efficiency and reliability of AMR operations.

The primary limitation of the proposed *Localization Score* approach is its reliance on PF-based localization systems operating on grid maps. While these remain the industry standard for localization, expanding the scoring system to accommodate a broader range of localization techniques is a key future direction. Additionally, the method requires fine-tuning of weighting parameters for optimal performance. A learning-based adaptive architecture will be implemented to address this and further refine the scoring system, especially to automate the parameter tuning procedure. Lastly, while the approach has been validated in challenging industrial indoor environments, future research will explore its applicability to outdoor and other domains to enhance its scalability and adaptability.

[2]https://wiki.ros.org/navigation

[3]https://docs.nav2.org/

ACKNOWLEDGMENTS

This study is supported by a collaborative project, titled "Autonomous Navigation Software Development for Mobile Robots Used in Automotive Production Logistics", between Ford Otomotiv Sanayi A.Ş. (Otosan) and Istanbul Technical University. Ford Otosan is a leading automotive manufacturing company and has been among Turkiye's three largest exporters since 2004.

REFERENCES

1. A. K. Grover and M. H. Ashraf, "Autonomous mobile robots for warehousing and distribution industry: A step toward intralogistics 4.0," in *Digitization in supply chain management: trends, challenges and solutions*. World Scientific, 2024, pp. 153–183.
2. B. Thormundsson, "Global autonomous mobile robot market size 2016-2028," https://www.statista.com/statistics/1285835/worldwide-autonomous-robots-market-size/, 2023, accessed on 10 May 2024.
3. P. Y. Leong and N. S. Ahmad, "Exploring autonomous load-carrying mobile robots in indoor settings: A comprehensive review," *IEEE Access*, 2024.
4. C. R. Albrecht, J. Behre, E. Herrmann, S. Jürgens, and U. Stilla, "Investigation on robustness of vehicle localization using cameras and lidar," *Vehicles*, vol. 4, no. 2, pp. 445–463, 2022.
5. G. D. Tipaldi, D. Meyer-Delius, and W. Burgard, "Lifelong localization in changing environments," *The International Journal of Robotics Research*, vol. 32, no. 14, pp. 1662–1678, 2013.
6. R. B. Sousa, H. M. Sobreira, and A. P. Moreira, "A systematic literature review on long-term localization and mapping for mobile robots," *Journal of Field Robotics*, vol. 40, no. 5, pp. 1245–1322, 2023.
7. G. Fragapane, R. De Koster, F. Sgarbossa, and J. O. Strandhagen, "Planning and control of autonomous mobile robots for intralogistics: Literature review and research agenda," *European Journal of Operational Research*, vol. 294, no. 2, pp. 405–426, 2021.
8. P.-Y. Lajoie and G. Beltrame, "Swarm-slam: Sparse

IEEE Robotics & Automation Magazine (RAM) paper, presented at ICRA 2026, Vienna, Austria. Cite as RAM paper.

decentralized collaborative simultaneous localization and mapping framework for multi-robot systems," *IEEE Robotics and Automation Letters*, vol. 9, no. 1, pp. 475–482, 2024.

9. S. Nobili, G. Tinchev, and M. Fallon, "Predicting alignment risk to prevent localization failure," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1003–1010.
10. R. Valencia, J. Saarinen, H. Andreasson, J. Vallvé, J. Andrade-Cetto, and A. J. Lilienthal, "Localization in highly dynamic environments using dual-timescale ndt-mcl," in *2014 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2014, pp. 3956–3962.
11. N. Akai, "Reliable monte carlo localization for mobile robots," *Journal of Field Robotics*, vol. 40, no. 3, pp. 595–613, 2023.
12. A. Yilmaz, E. Sumer, and H. Temeltas, "A precise scan matching based localization method for an autonomously guided vehicle in smart factories," *Robotics and Computer-Integrated Manufacturing*, vol. 75, p. 102302, 2022.
13. L. Garrote, M. Torres, T. Barros, J. Perdiz, C. Premebida, and U. J. Nunes, "Mobile robot localization with reinforcement learning map update decision aided by an absolute indoor positioning system," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 1620–1626.
14. W. Wang, Q. Cao, X. Zhu, and M. Adachi, "An automatic switching approach of robotic components for improving robot localization reliability in complicated environment," *Industrial robot: an international journal*, vol. 41, no. 2, pp. 135–144, 2014.
15. A. Boukerche, B. Kantarci, and C. Kaptan, "Towards ensuring the reliability and dependability of vehicular crowd-sensing data in gps-less location tracking," *Pervasive and Mobile Computing*, vol. 68, p. 101248, 2020.
16. L. Ferranti, X. Li, J. Boutellier, and J. Kannala, "Can you trust your pose? confidence estimation in visual localization," in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2020, pp. 5004–5011.
17. J. Chen, C. Shi, and J. Chen, "Application of high precision localization in following service robots," in *2019 Chinese Automation Congress (CAC)*. IEEE, 2019, pp. 3736–3741.
18. G. G. Anagnostopoulos and A. Kalousis, "Can i trust this location estimate? reproducibly benchmarking the methods of dynamic accuracy estimation of localization," *Sensors*, vol. 22, no. 3, p. 1088, 2022.
19. D.-V. Nguyen, T.-K. Dao, E. Castelli, and F. Nashashibi, "A fusion method for localization of intelligent vehicles in carparks," *IEEE Access*, vol. 8, pp. 99 729–99 739, 2020.
20. M. Elsanhoury, P. Mäkelä, J. Koljonen, P. Välisuo, A. Shamsuzzoha, T. Mantere, M. Elmusrati, and H. Kuusniemi, "Precision positioning for smart logistics using ultra-wideband technology-based indoor navigation: A review," *Ieee Access*, vol. 10, pp. 44 413–44 445, 2022.
21. H. Jing, Y. Gao, S. Shahbeigi, and M. Dianati, "Integrity monitoring of gnss/ins based positioning systems for autonomous vehicles: State-of-the-art and open challenges," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 14 166–14 187, 2022.
22. C.-b. Moon, W. Chung, and N. L. Doh, "Observation likelihood model design and failure recovery scheme toward reliable localization of mobile robots," *International Journal of Advanced Robotic Systems*, vol. 7, no. 4, p. 24, 2010.
23. S. He, T. Song, P. Wang, C. Ding, and X. Wu, "An enhanced adaptive monte carlo localization for service robots in dynamic and featureless environments," *Journal of Intelligent & Robotic Systems*, vol. 108, no. 1, p. 6, 2023.
24. I. Hroob, B. Mersch, C. Stachniss, and M. Hanheide, "Generalizable stable points segmentation for 3d lidar scan-to-map long-term localization," *IEEE Robotics and Automation Letters*, 2024.
25. M. Demir, K. Fujimura, "Robust localization," Patent, 2022, patent No. 11506502, U.S. Patent and Trademark Office.
26. J. L. C. Villacrés, Z. Zhao, T. Braun, and Z. Li, "A particle filter-based reinforcement learning approach for reliable wireless indoor positioning," *IEEE journal on selected areas in communications*, vol. 37, no. 11, pp. 2457–2473, 2019.
27. A. Yilmaz and H. Temeltas, "Self-adaptive monte carlo method for indoor localization of smart agvs using lidar data," *Robotics and Autonomous Systems*, vol. 122, p. 103285, 2019.
28. I. H. Savci, A. Yilmaz, S. Karaman, H. Ocakli, and H. Temeltas, "Improving navigation stack of a ros-enabled industrial autonomous mobile robot (amr) to be incorporated in a large-scale automotive production," *The International Journal of Advanced*

Manufacturing Technology, vol. 120, no. 5-6, pp. 3647–3668, 2022.

29. A. Yilmaz, A. D. Vit, I. H. Savci, H. Ocakli, and H. Temeltas, "Reference cage architecture for autonomous docking of mobile robots in automotive production systems," *The International Journal of Advanced Manufacturing Technology*, vol. 129, no. 7, pp. 3497–3511, 2023.
30. S. Zhu, X. Zhang, S. Guo, J. Li, and H. Liu, "Lifelong localization in semi-dynamic environment," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 14 389–14 395.
31. C. M. Costa, H. M. Sobreira, A. J. Sousa, and G. M. Veiga, "Robust 3/6 dof self-localization system with selective map update for mobile robot platforms," *Robotics and Autonomous Systems*, vol. 76, pp. 113–140, 2016.
32. J. Weber and M. Schmidt, "Precise and reliable localization of mobile robots in crowds using ndt-amcl," in *2024 13th International Workshop on Robot Motion and Control (RoMoCo)*. IEEE, 2024, pp. 7–12.
33. D. Wilbers, C. Merfels, and C. Stachniss, "A comparison of particle filter and graph-based optimization for localization with landmarks in automated vehicles," in *2019 Third IEEE International Conference on Robotic Computing (IRC)*. IEEE, 2019, pp. 220–225.

Abdurrahman Yilmaz received his Ph.D. degree in Control and Automation Engineering at Istanbul Technical University (ITU), Turkiye. His current research areas include probabilistic approaches in robotics and autonomous robotic systems.

Umut Dumandag graduated from Istanbul Medeniyet University with a degree in Electrical and Electronics Engineering. He currently works as a Robotics Software Engineer at Bluepath Robotics. In his role, Umut focuses on developing and optimizing software for robotic systems.

Aydin C. Sari graduated from Yıldız Technical University with a degree in Electrical Engineering and obtained his MSc degree in Mechatronics Engineering from Istanbul Technical University. He has worked as a lead engineer, software architect, and manager at various UAV and AMR startups. He took on leadership roles at renowned autonomous vehicle companies, such as Ford Otosan and Argo AI. He is currently serving as the Robotics Software Manager at Bluepath Robotics.

Ismail H. Savci received his BS degree in Mechanical Engineering from Istanbul Technical University in 2007, and his MS and PhD degrees from Marmara University in 2009 and 2020, respectively. From 2008 to 2023, he worked at Ford Otosan in various roles, including senior engineer and test manager, where he established test laboratories in the Gölcük Plant. He has published 65 conference and journal papers, holds 21 patents, has led 3 international projects, and has published 3 book chapters. His expertise includes internal combustion engines and autonomous mobile robots. He is currently the CEO of Bluepath Robotics.

Hakan Temeltas received his Ph.D. degree in Robotics in 1993 at the University of Nottingham, UK. Currently, he is a professor at Istanbul Technical University. His major work has been in the field of sensor data fusion for navigation and control of mobile robots in several projects. His research interests, in general, are planning, modeling, control, and design of robotics and mechatronics systems.