

Real-Time Marker-Based Monocular Autonomous Docking in Semi-Unstructured Indoor Environments*

Sebastian Chinchilla¹, Takumi Saito², Ryosuke Oikawa³, Tomoaki Yamada⁴, Naoto Toshiki⁵,
Satsuki Yamane⁶, Jose Salazar⁷, Ankit A. Ravankar⁸ and Yasuhisa Hirata⁹

Abstract—

Environmental changes can severely disrupt the docking sections of transportation systems in factories. Such disruptions not only halt the production line but also necessitate human intervention for loading and unloading, posing safety risks. To address this, we introduce an autonomous docking system designed for resilience in semi-unstructured environments, especially when faced with varying light conditions and physical alterations. Our solution involves an advanced ArUco marker system. By integrating the yaw orientation of the marker with angular velocity data from an inertial measurement unit, we achieve a homogeneous matrix reconstruction. This enhanced marker data then guides a reference trajectory for docking, controlled by a PI Pure Pursuit mechanism. To further improve marker recognition, we have incorporated reflective and anti-reflective materials and modulated the marker's white margin. Our test results indicate a significant improvement in the detection range of a 21cm marker expanded from 60cm to 2m. Moreover, our system ensures a docking precision of $\pm 31\text{mm}$ and $\pm 1.82^\circ$.

I. INTRODUCTION

Intelligent autonomous robots and vehicles have been recently used to replace or aid humans in the transportation and distribution of cargo within factories [1]. However, there is a delicate part of the distribution process that possesses a series of complicated issues: autonomous docking. Specifically, docking is the process in which a mobile machine couples with fixed equipment to exchange loads. It is a challenging task for autonomous robots due to static, safety, and dynamic issues.

In terms of the static challenges, we can mention the accuracy, available space for motion, and shutter configuration invariability. First, Regarding the accuracy, the robot is required to dock at a given position and orientation within a required accuracy. In general, in many applications, the robot

is required to arrive in a pose perpendicular to the shutter and follow a determined path, usually a straight line, with millimetric precision. Second, concerning space, robots must maneuver within a confined space while achieving a correct pose at the goal point. This is a consequence of the fact that around the docking there are surrounding equipment, walking areas for humans, and transportation zones for other robots and vehicles so the available area for such a process is very limited. Third, concerning shutter configuration, adding a global guidance system on the shutter side is very restrictive. Whereas, the inclusion of signal emitters such as infrared sources is limited to the existence of available power sources, the incorporation of passive markers is hindered by the absence of sufficient space to locate relatively large patterns on the surface area of the equipment. As a result, it is preferable to keep all the sensors on board of the robot, and all global reference systems as compact as possible in the shutter side.

Another attention area are the dynamic constraints within the factory environment. Due to safety reasons, the motion of autonomous robots is bounded between a minimum and a maximum speed value, as well as a minimum docking time. The problem that arises from these conditions is a compromise between the completion of the trajectory and the speed of response. To arrive at the docking goal with the correct pose in the allotted time, a fast response is desired. However, in many cases, this leads to either violating the established maximum speed, not arriving at the docking on time, or arriving in an incorrect orientation, all resulting in accidents or failure of the docking.

Furthermore, there is the semi-unstructured nature of the docking area. On the one hand, the shutter itself is structured (the docking area) meaning the shutter shape and its area are predetermined and fixed so that it is possible to assume such equipment is static. On the other, the surroundings of the shutter are unstructured. Nearby the shutter, moving objects constantly pass by at an unpredictable rate. For example, electric carts and other robots move around the shutter to perform their tasks depending on the production demand of that day. Therefore, autonomous robots must be tolerant to constant small and slow changes in the environment, such as the relocation of pallets and equipment [2]. Additionally, they need to adapt to fast-moving dynamic objects, such as humans and forklifts. Finally, they need to be robust under extreme changes of lighting conditions, since most factories operate round the clock [3]. On top of that, solutions must be feasible and affordable.

¹S. Chinchilla sebastian.chinchilla@toyota-ej.co.jp,
²S. Saito takumi.saito@toyota-ej.co.jp,

³R. Oikawa ryosuke.oikawa@toyota-ej.co.jp,

⁴T. Yamada tomoaki.yamada@toyota-ej.co.jp,

⁵N. Toshiki naoto.toshiki@toyota-ej.co.jp, and

⁶S. Yamane satsuki.yamane@toyota-ej.co.jp, are with the Project F of the Robotics & AI Development Department of the TOYOTA MOTOR EAST JAPAN, INC., Susono, Japan.

⁷J. Salazar j.salazar@srd.mech.tohoku.ac.jp,

⁸R. Ankit ankit@srd.mech.tohoku.ac.jp, and

⁹Y. Hirata hirata@srd.mech.tohoku.ac.jp, are with the Department of Bioengineering and Robotics, Tohoku University, Sendai, Japan.

While the shutter itself is structured meaning that the shutter shape and its area are predetermined and fixed so that it's possible to assume such equipment is static. Whereas, the surroundings of the shutter are unstructured. First, the environment around the shutter is constantly changing and its motion is variable, either at a fast pace (such in the case of forklifts or other robots) or at a slow rate (such as pallets or other equipment that occasionally are moved). Second, there are variable lighting conditions.

Aiming to reduce the health risks of humans in the transportation and distribution tasks at factories, we propose an autonomous docking system for semi-unstructured environments. This paper proposes two main contributions. First, we developed a methodology for detecting a global reference in varying lighting conditions. Based on the ArUco fiducial marker, we enhanced the robustness of the detection of the marker by using reflective and non-reflective materials rather than the traditional paper-based marker. Specifically, we propose a technique for modulating the ratio between the light-absorbing and reflection portions so that the marker can provide a measurement reference in a wide variety of lighting conditions. Second, we developed a lightweight real-time monocular method for enhancing the raw noisy measurements of the fiducial marker. This was achieved by fusing the raw orientation obtained from the fiducial marker, fusing with the orientation measurement of an IMU, and retro-feeding the corrected orientation to the subsequent homogeneous transformations that generate the global reference. Based on these developments, we implemented an autonomous docking system that allows the robot to dock at a given position and goal while dealing with velocity and spatial restrictions.

II. RELATED WORKS

Autonomous mobile robots use a variety of sensors to navigate their environment. Traditional methods include mapping with LiDARs and computer vision for the analysis of global references to navigate in structured environments. In addition, fusion-based techniques enhance such methods through the incorporation of additional sensors such IMUs to navigate in semi-structured and semi-static environments.

Fiducial Markers are visual patterns used in localization systems that provide position and orientation measurements. Among them, the most widely used are ArUco [4], ARTag [5], [6], AprilTag [7], and STag [8]. While AprilTag provides the more accurate orientation measurements [9], STag excels at position measurements [8]. However, both require intensive computations, making it difficult to use them in real-time in many navigation applications [10]. ARTag requires the least computational power but it has the lowest detection rate. Amongst them, ArUco possesses the best trade-off between measurement and computing power. It provides accurate position and orientation measurements while being lightweight for real-time implementations [11].

Although it provides accurate measurements, several factors affect the ArUco Marker performance. First, in terms of accuracy, the measurement error is proportional to the

distance [12] and orientation [13] to the marker, and inversely proportional to the marker. Consequently, the best results are obtained when the marker size is large enough to be detected by the camera and it is in proximity to the camera sensor. Second, lighting conditions greatly affect the recognition rate of the marker [14]. Both bright and dark environments degrade the identification of the marker due to reflections and shadows casted on the marker. Consequently, these markers present challenges when used in tasks that necessitate 24/7 operations.

Recent strategies for enhancement involve employing multiple markers, which typically necessitate the use of several cameras or high-resolution observation. These multiple markers can be bidimensional or tri-dimensional. Planar markers include grids of markers [15] and nested markers [16] (small markers inside bigger ones), while tri-dimensional markers include multiple surface structures such as dodecahedrons [17] and cubes [18]. For the assessment of these markers, cameras with lower resolutions (640×480 p) have been employed in tasks such as motion capture and surveillance [19]. In contrast, single-lens cameras necessitate a minimum resolution of 1280×720 p, and for millimetric accuracy, resolutions reaching up to 4096×3000 p are essential [17]. However, all of these solutions generate a trade-off with computing power, which prevents them from being used effectively in the navigation of autonomous robots. For this reason, they are mostly relegated to surveillance tasks [19], motion capture [18], and tracking of small mechanical motions in strain experiments [17].

An alternative approach to improve precision involves integrating methodologies with inertial systems. Such combinations have enabled the utilization of ArUco markers in real-time navigation for UAVs (Unmanned Aerial Vehicles). The main goal is to use the marker as a reference for landing and docking to either a mobile base [20], [21], a fixed base [22], the floor [22], or the roof [11]. However, in practice, the camera and the marker are parallel, which circumvents the problems related to the difference in orientation between the marker and the camera. Additionally, they have been tested only in outdoor and indoor environments under good lighting for recognition.

III. FACTORY CONDITIONS

To develop autonomous navigation for semi-structured settings, we base our design on real-world scenarios and their inherent conditions. Specifically, our study focuses on the docking prerequisites for the car component distribution system at a factory under the Toyota Motor East Japan, Inc. umbrella. Nevertheless, due to confidentiality constraints, the experiments documented herein were conducted in a laboratory setting using a replicated robot, as opposed to the genuine industrial robot or factory infrastructure. 1(a) showcases the replicated robot, while the shutter is depicted in Fig.1(b).

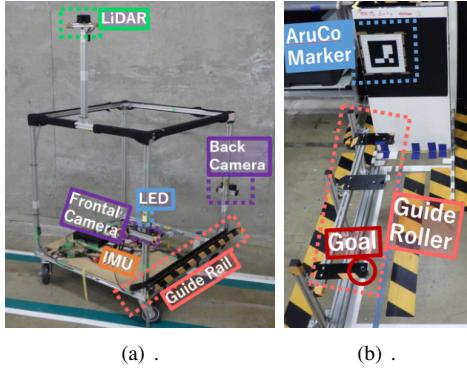


Fig. 1. Components of the system: (a) the robot and its guide rail, (b) the shutter and its guide roll

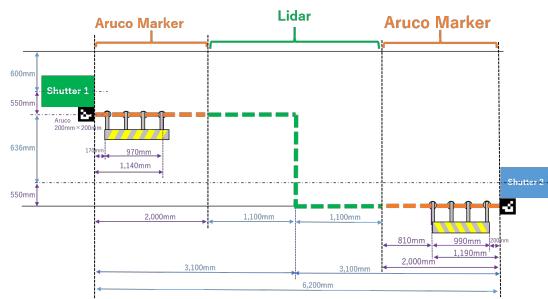


Fig. 2. Transportation and distribution areas in the factory.

A. Semi-Unstructured Conditions

Our proposed distribution system can be divided into two parts: The shutter area, and the transportation area, as seen in Fig.2. In transportation, the robot moves a load of components from a loading area (shutter 1) to an unloading area (shutter 2). This portion is very predictable and static, so a LiDAR based mapping system is sufficient to control and navigate the robot from one shutting area to the other. The second area is the shutters, where the docking is performed to exchange the vehicle components. This is a particularly challenging area due to their unstructured conditions.

The docking, or shutter, areas are subjected to multiple dynamic changes throughout the day. For example, surrounding the shutters, there is a temporal storage area for pallets and carts full of components. Next, a multitude of equipment is temporarily installed and removed, which affects the layout surrounding the robot. Because of these reasons, the LiDAR navigation, which depends on a static map, suffers a considerable loss of localization accuracy that leads to docking failures. This situation prevents the LiDAR system from being used within a distance of 2m away from each shutter.

A third situation limits the use of computer vision approaches. Since the factory operates around the clock, the lighting conditions vary from obscure to luminous. Specifically, the measured light intensity varies from 7 to 90 lux. Such conditions reduce the recognition rate of landmarks or markers that might be used as global references. While the

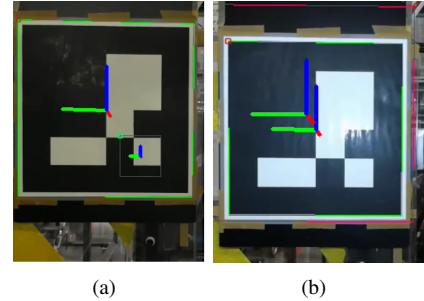


Fig. 3. Lighting conditions: (a) shadows casted under obscure settings, (b) Reflections caused due to bright natural or artificial lights.

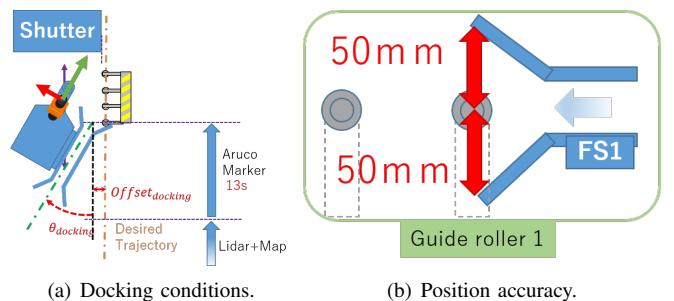


Fig. 4. Docking mechanism of the shutters.

first one casts shadows, the second one produces undesired reflections, both distorting the images captured from the used references (Fig. 3).

B. Docking Accuracy requirements

The actual mechanical docking system is composed of two parts. While, in the shutter, there is a roller guide that extends from it to 1.12m (Fig. 1(a)), on the robot side there is a guide rail that has an aperture of 100mm (Fig. 1(b)). The docking task of the robot consists of introducing the fixed guard rail inside the on-board guide connector. To do so, it is necessary to arrive at the initial docking point with an accuracy of $\pm 50\text{mm}$ (Fig. 4(a)) and $\pm 6^\circ$ (Fig. 4(b)). As long as the robot arrives at the goal within that range of error, the robot just needs to move forward to guarantee a successful docking. The mechanical design of the guide and rail mechanism allows for a final passive correction of the pose so that the robot arrives at the shutter in a straight orientation. On the contrary, if the robot does not manage to comply with such accuracy, the guide roll fails to enter the guide connector, failing the docking procedure.

C. Safety Restrictions

Due to safety reasons, the velocities and docking time of the robot are bounded. In the case of the velocities, the maximum velocity is clamped at 0.625m/s to reduce the damage to equipment and people during an imminent collision. In terms of the docking time, the robot must dock within 20 seconds, so that it does not become a slow obstacle to other robots, vehicles, or employees.

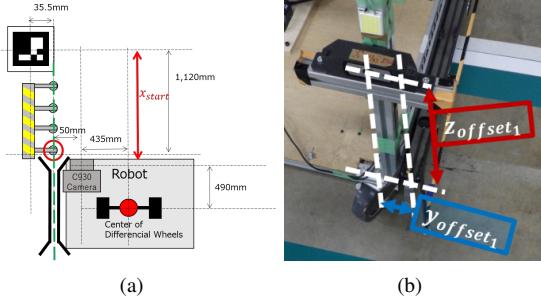


Fig. 5. Space availability for additional equipment: (a) measurements, (b) camera location.

D. Spatial Restrictions

In addition, there are restrictions to the available space in the robot and the shutter. These limitations come from the fact that in other areas there are already existing sensors and devices related to the post-docking procedure. Regarding the shutter, the space for installing a global reference, such as a marker, is limited to $21 \times 21\text{cm}^2$. Furthermore, the center of such marker is offset concerning the ideal trajectory of the roller guide by 35.5mm, as seen in Fig. 5(a). Relating to the robot, the vacant space for a camera is at a vertical offset of 15cm and a horizontal one of 5cm with respect to the guard rail (z_{offset_1}), as shown in Fig. 5(b).

IV. SYSTEM OVERVIEW

A. Control Concept

The general idea of the autonomous docking is presented in Fig. 6. Initially, we have a fiducial marker that can be used as a global reference or landmark. Then, through this reference, we calculate the location of the goal, whose position is fixed with respect to the marker, and based on its position and orientation we build an ideal reference trajectory. This trajectory is a straight line perpendicular to the marker and parallel to the docking goal and docking rollers. Furthermore, we chose a goal point in the trajectory line that is within a virtual circular perimeter built around the robot. As the robot moves, the goal point is updated, allowing it to get closer to the docking goal (WP_2) while correcting its orientation.

B. System Integration

The docking system was implemented through the integration of the three main blocks shown in Fig. 7. It includes the Raw ArUco (RA), ArUco Measurement Enhancement and Localization (AMEL), and the ArUco Control Block (ACB). Their purpose is to obtain the global measurement, perform corrections and move the robot based on this information, respectively.

1) *Global reference:* We use an ArUco marker [23] as the global reference. It was chosen because it provides a measurement of its position and orientation with respect to the camera in real time. To obtain such measurement, first, we capture a raw image with a web camera, and subsequently estimate its position and orientation through the

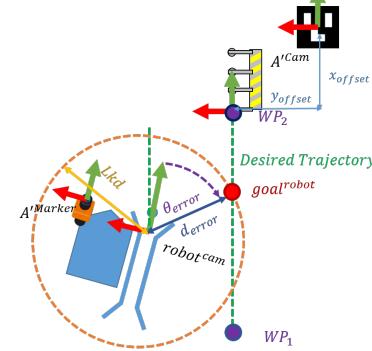


Fig. 6. Concept of the navigation with an ArUco marker.

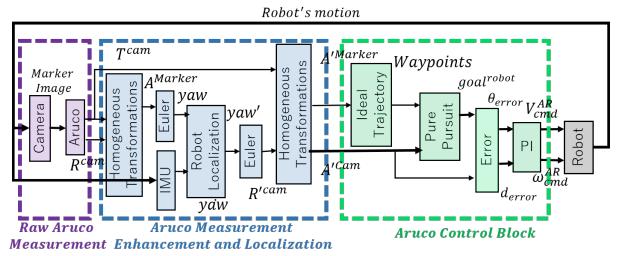


Fig. 7. Block diagram of the navigation system.

use of the ArUcoPose algorithm [23] within the RA block. Having the purpose of maximizing the speed processing of the algorithm, we capture the images at 25FPS with a low resolution of $640 \times 480\text{p}$.

In order to simplify the system, we assume that the measurements are bidimensional (x and y) and that the only orientation between the marker and the camera is the yaw angle. Since the measurements provided by the marker are tri-dimensional, we nullify the z , roll, and pitch measurements.

2) *Enhancement of the Global Reference Measurements:* The ArUco marker's accuracy is affected by the distance and angle from the marker to the camera. As shown in Fig 5(a), there is a planar offset between the marker and the docking goal. Next, at the starting point of motion, the camera is even farther away from the marker, and with an offset orientation with respect to the plane of the marker. Both situations greatly affect the marker measurements. Given that orientation is most susceptible to noise, we suggest a measurement correction for the AMEL block.

We take the raw pose orientation and position of the marker and convert it into a homogeneous transformation matrix as seen by the reference frame of the camera A^{Marker} . From there, we extract the yaw orientation with an Euler function. Through the use of an extended Kalman filter (EKF), we fuse and smooth such angular measurement with the yaw angular velocity ($\dot{\gamma}_{yaw}$) of an inertial measurement unit (IMU). The resulting fused angle (γ_{yaw}') is turned into a rotation matrix with an inverse Euler function combined with the original raw translation vector, and converted

with homogeneous transformation to generate the corrected transformation matrices with the measurements of the robot position and orientation seen by the camera (A'^{cam}) and by the camera of the robot (A'^{Marker}).

The proposed reconstruction and enhancement of the transformation matrices with EKF fusion is utilized for three reasons. First, a simple EKF of the x, y, and yaw from the camera reference frame data is still very noisy at the angle measurement. Consequently, when transitioning to the marker's reference frame, the measurements exhibit discontinuities, manifesting as abrupt shifts. Second, filtering and fusing of the x, y, and yaw from the marker, the yaw from the IMU, and even the odometry of the wheels, still results in discontinuous measurements. Lastly, in the camera's reference frame, the raw ArUCo measurements for x and y are already continuous; only the yaw angle exhibits discontinuity. Consequently, we opted to employ the filtering and fusion methodology solely to fuse the yaw and subsequently reconstruct the transformation matrices. This approach yielded more favorable outcomes.

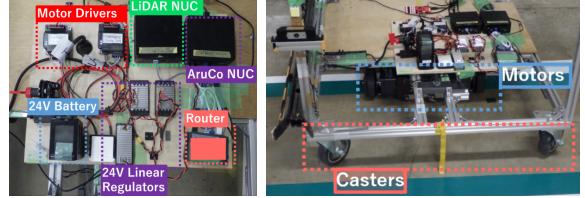
3) Enhancement of the Recognition Rate of the Marker: ArUco markers are usually used directly printed on paper and used that way. For protection, the marker is laminated as well. However, both cases present reflections when the luminosity of the environment is high, causing the black portion of the marker to look white. In contrast, when the brightness of the surroundings is low, the white color of the marker appears darker in the image of the camera. As a consequence of both phenomena, the recognition rate of the marker gets reduced. To increase the recognition rate of the marker, we implemented two countermeasures.

First, we use a marker with a white reflective material and a black anti-reflective cloth. This way the white portion is visible under low luminosities while the black one is not affected by reflections.

Second, we widened the white margin of the marker to the minimum required by the darkest environment. In general, the bigger the black portion of the marker, the better the accuracy is obtained. Similarly, the wider the white margin gets, the more robust the recognition rate becomes. Nevertheless, since the maximum size of the marker is limited to 21cm due to the previously mentioned space restrictions of the shutter, there is a compromise between the affordable margin and marker sizes.

Our solution consisted of locating the camera at the farthest place from the marker (the starting point) during the hours with the lowest ambient luminosity. During that phase, we experimented with various margin sizes and ultimately chose the slimmest one that ensured a 100% recognition rate. Subsequently, we confirm its robustness by corroborating the recognition rate during different hours and illumination conditions around the clock.

4) Trajectory Control: The ACB block creates an ideal trajectory based on the orientation of the robot. It creates a straight line that extends on the negative portion of the x-axis of the reference frame of the marker. The point in such line that must be followed by the robot is determined with the



(a) Controllers and drivers. (b) Differential wheels with casters.

Fig. 8. Control hardware.

Pure Pursuit algorithm (PP) [24]. In short, PP draws a circle around the robot reference frame, whose intersection with the ideal line is used as the robot's motion goal. In case there are two points where the circle touches the line, the one closest to the docking goal is chosen. The resulting goal point of the PP is compared with the current robot's pose to estimate the orientation and position error. Finally, such error is turned into linear and angular velocity commands through a PI controller.

5) Hardware: The robot is a custom differential drive robot whose controllers are shown in Fig. 8(a). Its differential wheels (Fig. 8(b)) are operated through an Orientalmotor BLV-R-type driver. The controllers for the ArUco and LiDAR system are each in a NUC11 i7 processor, which is interconnected with a wireless router. Regarding the ArUco marker system, we are using a C930 web camera at 30FPS with a 640×480 p resolution, as shown in Fig 1(a), and a RT-USB 10-axis IMU as the inertial measurement unit. Finally, the power source consists of a 24V linear regulator that is fed with a single 24V battery.

6) Software: The development environment is Ubuntu 20.04, in which we use ROS to interconnect the different software blocks through wireless. The ArUco raw measurements are performed with the ArUco library, in OpenCV within Python 3. The EKF fusion is performed with the Robot Operating System (ROS) robot-localization [25] package. Finally, the homogeneous and Euler transformations are executed with functions from the ROS tf package. Finally, the ACB block is custom software written in Python 3.

V. EXPERIMENTS

A. Experimental Setup for the Marker Selection

In order to determine the best marker for our application in the factory, we are testing two different types of markers with four different margin sizes under a variety of lighting conditions. All of the markers were initially printed on paper, though they were covered in different materials afterward. In the first case, we laminate the paper marker with a mat transparent plastic, so that reflections are reduced. As a result, both, the white margin and black portion of the marker have a mat finish (Fig. 9(b)). We call the resulting marker a mat-mat marker (MM). In the second case, after laminating the markers, we cover the black portions with an anti-reflective cloth, whereas a reflective sticker is glued to the white portion. We refer to such combination as the white-reflective black-anti-reflective marker (WRBA)(9(c)).

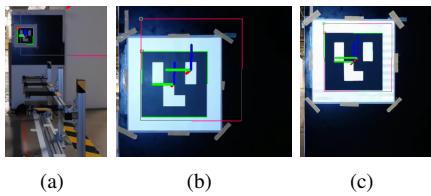


Fig. 9. Markers with different materials (white margin: 3cm); (a) Measurement setup, (b) White: laminated mat black: laminated mat, (c) White: reflective, black: anti-reflective.

In addition, we are testing both marker types with four white margin thickness values. Namely, 1.5cm, 2cm, 2.5cm, and 3cm. It is important to mention that all of the markers have a total area of 21 cm x 21 cm. Consequently, the black portion of the marker measures 18 cm x 18 cm, 17 cm x 17 cm, 16 cm x 16 cm, and 15 cm x 15 cm, respectively. Since the ArUco measurements degrade as the marker shrinks, we did not want to reduce the black portion of the marker beyond 15cm. Similarly, we do not want to reduce the marker's margin beyond 1.5cm because, in a preliminary experiment with a paper-based marker, a margin of 0.5cm was not recognized in our test setup. Consequently, we increased such value three times having the goal of increasing the minimum recognition rate.

In terms of the luminosity, we tested 4 values: 7lux, 30lux, 70 lux, and 90lux. The first two correspond to measurements performed in the late afternoon and early night, whereas the next two coincide with the morning and early afternoon.

The setup for the recognition experiment is shown in Fig. 9(a). We locate the marker on the shutter part of the robot in the starting position, so that there is a perpendicular distance between the marker and the camera of 2m. Then, we capture images at 25FPS for one minute for a total of 1500 frames. Finally, we calculate the recognition rate as the percentage of frames that were recognized.

B. Result of the Marker Selection

The obtained recognition rates are shown in the Fig. 10. The best performance was obtained with the WRBA marker at 2.5cm and 3cm of margin, which were successfully detected without problems in all of the lighting conditions. In contrast, margins of 2cm and 1.5 preset a recognition rate of just 52% and lower. MM markers performed well above 30lux but failed the recognition with margins of 2.5cm and lower at 7lux.

Since the WRBA marker was robust at 3cm and 2.5cm under all of the illuminances, we selected such marker material. In addition, to have a security margin for the recognition, we chose a marker with a 3cm thickness.

C. Experimental Setup for the Raw Marker Accuracy

To determine the actual dynamic measurement accuracy of the selected WRBA marker, we performed a rotational experiment. We located the robot at the starting position indicated in Fig. 5(a) at different distances from the marker (x_{offset}) and then proceeded to rotate the robot $\pm 10^\circ$ around

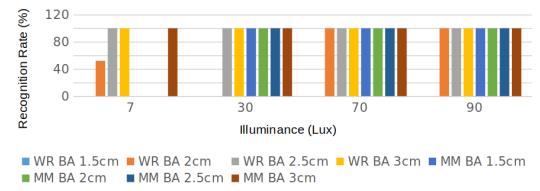


Fig. 10. Recognition rate.

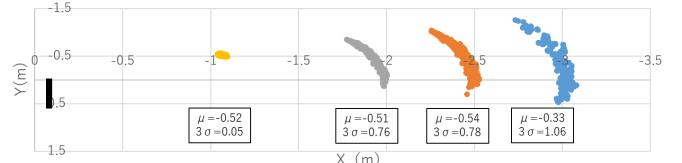


Fig. 11. Raw dynamic accuracy of the WRBA marker (21cm×21cm). Rotation at the axis of the robot.

its axis. When estimating the position of the center of the robot from the marker reference frame, it is expected to observe that it is at a fixed location while rotating. We located the robot at four different positions, to determine how far such observation holds. Since the results of the recognition experiments provided recognition rates of 100% from 7 to 90lux, we only performed experiments at an indoor luminance of 30lux.

D. Results for the Raw Marker Accuracy

In fact, our observations indicated that the robot remained stationary, except at the point closest to its rotation relative to the marker, as depicted in Fig. 11. All other distant reference points result in an error that has a standard deviation bigger than our target accuracy. In other words, we can only obtain accurate measurements as long as the camera is within a range of 60cm from the marker. However, this is not acceptable for our system, since we need to obtain a reliable measurement in the range from 1.12m to 2m. As a result, we are required to increase the default accuracy of the ArUco markers.

E. Experimental Setup for the Docking Tests

To test our proposed system, we performed a docking experiment. We located the robot at three different locations, each with a different orientation, as shown in Fig. 12. The selected left and right starting points represent the maximum offsets at which the LiDAR system would hand over the robot to the developed docking system. In other words, our system expects to initiate its motion within an area of $\pm 100\text{mm}$ and $\pm 10^\circ$. In addition, the central line is included as well to verify that the robot can move in a straight line as close as possible. Lastly, each starting point is tested three times, for a total of nine experiments.

F. Results of the Docking Tests

Figures 13, 14, and 15, present the position of the marker with respect to the camera reference frame, the position of the camera with respect to the marker reference frame, and

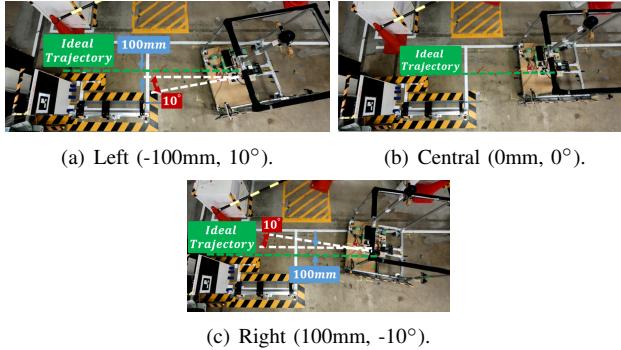


Fig. 12. Starting points and orientations for the docking experiments.

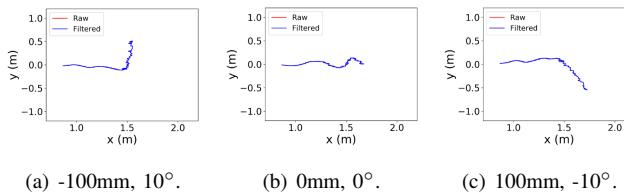


Fig. 13. Position results from the camera reference frame. The robot moves from the right to the left of the graph.

the yaw orientation between the camera and the marker, respectively, of one sample of each starting position. In these images, the red lines represent the raw ArUco measurements, whereas the blue lines indicate the output of our fusion filtered with matrix reconstruction.

As can be noted from the graphs, the noise of the yaw angle is greatly reduced, which translates into a relatively smooth and continuous trajectory measurement in the marker reference frame. Measurements from the camera reference frame are not filtered in our current scheme so that they coincide with the raw values.

Fig. 16 contains the accuracy results. The standard devia-

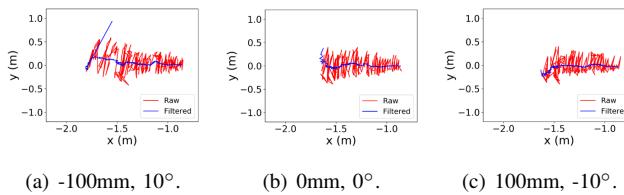


Fig. 14. Position results from the marker reference frame. The motion starts from the left of the graph.

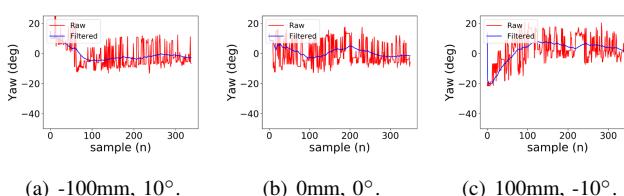


Fig. 15. Orientation results from the robot reference frame.

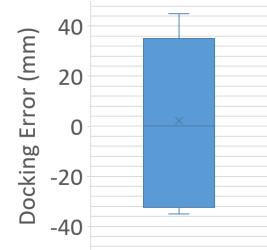


Fig. 16. Docking accuracy.

tion of the arrival point has a value of $\pm 31.3\text{mm}$ and $\pm 1.82^\circ$, which complies with the $\pm 50\text{mm}$ and $\pm 60^\circ$ requirement. Furthermore, the maximum values are as well within this range, which guarantees that even in the worst-case scenario the system is going to be capable of docking.

VI. DISCUSSION

The recognition experiments demonstrated that the WRBA marker is robust in environments with illuminations ranging from 7lux to 90 lux. Although the recognition test was performed statically at a distance of 2m, the docking tests showed that the recognition rate of such a marker is still 100% even under dynamic conditions. However, the docking was performed at about 30lux. A future complete confirmation of the spectrum of dynamic recognition rates would imply performing the docking under different lighting conditions. Despite that, the static recognition suggests that the dynamic one would be robust as well.

Regarding the autonomous docking, the resulting accuracy demonstrates that the system is capable of docking within the goal of accuracy. Furthermore, the fact that the goal accuracy is better than the raw ArUco measurement accuracy indicates that the designed AMEL system indeed corrected and enhanced those measurements. Nevertheless, it remains to determine what is the independent contribution of the IMU and the reconstruction of the homogeneous matrix in the process. Additionally, as a future comparison, we need to corroborate the percentage of accuracy enhancement in comparison to a classical EKF or UKF with IMU fusion.

It is important to mention that some improvements can be made to the system to increase its accuracy. First, the only data that was fused in our scheme was the yaw angle. Once the homogeneous matrices have been reconstructed with the yaw angle, a second fusion block can be added to fuse the x and y measurements as well as the IMU and wheel odometry of the robot.

An additional possible improvement comes in the form of an update to the version of the ArUco Library. At the moment, due to the simplicity of its implementation, we utilized the ArUco Library version that comes with OpenCV 4.5. This version uses the *estimatePose* function to extract the orientation and position from frames that contain the marker. However, more recent updates employ the *SolvePnP* function to solve the perspective-n-point problem, which is expected to provide more reliable results. Combined with our

approach, the docking accuracy could increase beyond our current results.

About the chosen sensors, it is important to mention that we chose a monocular approach rather than a multicamera for three reasons. The first one is cost. A multiple-camera approach might generate more accurate results, but it does increase the required implementation cost. The second reason is computational power. When using a single camera, the system can work at a frequency of 25 Hz. However, when using two cameras at the same time, the maximum frequency drops to 10Hz, which is not enough to comply with the required real-time response of the system. The third reason is space availability and safety. Cameras can only be located in areas where there are no other sensors and that not might represent a collision risk.

VII. CONCLUSIONS

In this paper, we presented an approach to perform an autonomous docking through the use of fiducial markers. We enhanced the marker in two ways. In terms of recognition rate, we made the marker robust to environments with varying lighting conditions, particularly from 7 to 90 lux by covering the white portions of the marker with reflective material and the black ones with an anti-reflective cloth. Second, we proposed a method to determine the minimum allowable white margin of the marker to aid the recognition of the marker. In the case of a 21cm marker, the minimum margin is 2.5cm, so we utilized a 3cm margin to provide a safe marker recognition. The second enhancement to the marker was through the implementation of the AMEL block, which reduces the noise of the orientation measurement and then reconstructs the global reference seen from the reference frame of the marker. The correction of the angle allowed us to extend the usable dynamic range of the marker from 60 cm to 2m to obtain an accuracy of $\pm 31\text{mm}$ and $\pm 1.82^\circ$, granting the autonomous robot the ability to dock automatically. Finally, we designed a trajectory controller that uses the corrected ArUco measurements. It creates a reference line that is followed with the help of the Pure Pursuit and PI algorithms. The combination of the enhanced ArUco measurements and recognition with the trajectory controller enables a docking system that can dock despite offsets between the marker and the camera angle and position, and any changes of light in the environment.

REFERENCES

- [1] F. Gebert, “Development of an autonomous mobile mapping robot by combining the navvis vlx with the boston dynamics spot,” Ph.D. dissertation, Hochschule für Angewandte Wissenschaften München, 2023.
- [2] J. Qian, V. Chatrath, J. Servos, A. Mavrinac, W. Burgard, S. L. Waslander, and A. P. Schoellig, “Pov-slam: Probabilistic object-aware variational slam in semi-static environments,” *arXiv preprint arXiv:2307.00488*, 2023.
- [3] N. Abdelaziz and A. El-Rabbany, “An integrated ins/lidar slam navigation system for gnss-challenging environments,” *Sensors*, vol. 22, no. 12, p. 4327, 2022.
- [4] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, “Automatic generation and detection of highly reliable fiducial markers under occlusion,” *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014.
- [5] A. M. Romanov and A. A. Tararin, “An automatic docking system for wheeled mobile robots,” in *2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus)*. IEEE, 2021, pp. 1040–1045.
- [6] M. Fiala, “Designing highly reliable fiducial markers,” *IEEE Transactions on Pattern analysis and machine intelligence*, vol. 32, no. 7, pp. 1317–1324, 2009.
- [7] E. Olson, “Apriltag: A robust and flexible visual fiducial system,” in *2011 IEEE international conference on robotics and automation*. IEEE, 2011, pp. 3400–3407.
- [8] B. Benligiray, C. Topal, and C. Akinlar, “Stag: A stable fiducial marker system,” *Image and Vision Computing*, vol. 89, pp. 158–169, 2019.
- [9] J. Wang and E. Olson, “Apriltag 2: Efficient and robust fiducial detection,” in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 4193–4198.
- [10] M. Kalaitzakis, B. Cain, S. Carroll, A. Ambrosi, C. Whitehead, and N. Vitzilaios, “Fiducial markers for pose estimation: Overview, applications and experimental comparison of the artag, apriltag, aruco and stag markers,” *Journal of Intelligent & Robotic Systems*, vol. 101, pp. 1–26, 2021.
- [11] A. J. Gaona, C. D. Pose, J. I. Giribet, and R. Bunge, “Prototyping of a multirotor uav for precision landing under rotor failures.”
- [12] K. Tsujita, “A feasible study on the model predictive control for docking approach of small spacecraft using thrusters and a control moment gyro,” in *2023 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*. IEEE, 2023, pp. 696–701.
- [13] A. Bontemps, G. Godi, E. Fournely, R. Moutou-Pitti, and J. Gril, “Implementation of an optical measurement method for monitoring mechanical behaviour,” *Experimental Techniques*, pp. 1–14, 2023.
- [14] R. Schouten, O. Arslan, and A. Isleyen, “Accuracy of single camera pose estimation with aruco fiducial markers.”
- [15] M. Žuži, J. Čejka, F. Bruno, D. Skarlatos, and F. Liarokapis, “Impact of dehazing on underwater marker detection for augmented reality,” *Frontiers in Robotics and AI*, vol. 5, p. 92, 2018.
- [16] T.-D. Do, N. Xuan-Mung, and S.-K. Hong, “Vision-based target pose estimation with multiple markers for the perching of uavs,” *arXiv preprint arXiv:2304.14838*, 2023.
- [17] G. Čepon, D. Ocepek, M. Kodrič, M. Demšar, T. Bregar, and M. Boltežar, “Impact-pose estimation using aruco markers in structural dynamics,” *Experimental Techniques*, pp. 1–12, 2023.
- [18] H. Sarmadi, R. Muñoz-Salinas, M. Berbís, and R. Medina-Carnicer, “Simultaneous multi-view camera pose estimation and object tracking with squared planar markers,” *IEEE Access*, vol. 7, pp. 22 927–22 940, 2019.
- [19] P. Oščádal, D. Huczala, J. Bém, V. Krys, and Z. Bobovský, “Smart building surveillance system as shared sensory system for localization of agvs,” *Applied Sciences*, vol. 10, no. 23, p. 8452, 2020.
- [20] C.-W. Chang, L.-Y. Lo, H. C. Cheung, Y. Feng, A.-S. Yang, C.-Y. Wen, and W. Zhou, “Proactive guidance for accurate uav landing on a dynamis platform: A visual–inertial approach,” *Sensors*, vol. 22, no. 1, p. 404, 2022.
- [21] E. Narváez, A. A. Ravankar, A. Ravankar, T. Emaru, and Y. Kobayashi, “Autonomous vtol-uav docking system for heterogeneous multirobot team,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–18, 2020.
- [22] J.-C. Lee, C.-C. Chen, C.-T. Shen, and Y.-C. Lai, “Landmark-based scale estimation and correction of visual inertial odometry for vtol uavs in a gps-denied environment,” *Sensors*, vol. 22, no. 24, p. 9654, 2022.
- [23] R. Muñoz-Salinas and S. Garrido-Jurado, “Aruco: A minimal library for augmented reality applications based on opencv,” *Sensors*, vol. 14, no. 11, pp. 21 072–21 089, 2014.
- [24] R. C. Coulter, “A robust path following algorithm for inside/outside vehicle navigation,” in *Proceedings of the American Control Conference*, 1992, pp. 1–7.
- [25] T. Moore and D. Stouch, “A generalized extended kalman filter implementation for the robot operating system,” in *Proceedings of the 13th International Conference on Intelligent Autonomous Systems (IAS-13)*. Springer, July 2014.