Calibration-free Localization for Mobile Robots Using an External Stereo Camera

Shinya Yasuda, Taichi Kumagai, and Hiroshi Yoshida
System Platform Research Laboratories, NEC Corporation
1753 Shimonumabe, Nakahara-ku,
Kawasaki, Kanagawa, Japan
Email: {s-yasuda@ce, t-kumagai@bq, h-yoshida@jh}jp.nec.com

Abstract—We propose a calibration-free method for localizing mobile robots using a stereo camera installed on the ceiling. We use both optical images and depth data obtained from the stereo camera to localize the robots. This method achieves 2.3 mm precision at the center of the field of view and 30 mm near the edge when the stereo camera is installed on an approximately 3.2-meter high ceiling, where the field of view is approximately 5 m × 3 m. The accuracy, i.e., the average distance between the localization result by the proposed method and the ground truth, was obtained at 28 mm. This value is an improvement over that obtained from the localization using only the optical image (42 mm). We also show experimentally that it is possible to perform automatic control of a mobile robot on the basis of the location information measured by the proposed method.

I. INTRODUCTION

Mobile robots, e.g., cleaning robots, service robots that convey something, and so on, have become much more common. For these robots, it is important to accurately determine the direction that they have to move to carry out their tasks. For example, a cleaning robot needs to move around a room to clean the floor. A service robot track an established route to convey things. A common way to achieve point-to-point conveyance is to use a physical guide such as magnetic tape. Such service robots have been used especially in factories, where a conveyance route is rarely changed. However, once the conveyance route changes, robots operators should remove the guide and prepare a new conveyance route. Also, a robot would be unable to avoid an obstacle it encounters unless a detour path is prepared.

From the above reasons, localization methods without physical guides have attracted much attention. A number of recently developed conveyance robots work with simultaneous localization and mapping (SLAM) technology [1], [2]. They can avoid obstacles and calculate a detour path by themselves. Nevertheless, frequent layout changes or luggage temporarily placed on the floor confuses the robots and causes localization to fail. This is partly because robots may confuse a temporary obstacle with a structural object such as a pillar. In short, applying existing systems to environments is difficult where drastic changes often occur. When one wants to employ robots in such environments, it becomes important to find obstacles for the robots as early as possible and to calculate another route to the destination.

To solve the aforementioned problems, we propose a calibration-free localization method for mobile robots using a stereo camera that is installed on the ceiling. The stereo camera can capture pictures and depth data simultaneously. The fixed camera enables distinguishing a temporarily placed object from a structural object more easily than a robot’s built-in sensors. By this feature, it is possible that the system detects obstacles on the planned path even when they are too far away to be detected by the robot’s build-in sensors. The main contributions of this work are an indoor localization method using a stereo camera installed on a ceiling and the achievement of 2.3 mm (30 mm) precision of localization when a robot is captured at the center (near the edge) of the field of view using the experimental prototype.

The rest of the paper is organized as follows. In Sec. II, we introduce previous works and how the proposed method differs from them. In Sec. III, we propose the system to localize mobile robots and explain the reasons for using a stereo camera. We also introduce the hardware we use in our prototype such as the stereo camera and robots. In Sec. IV, we present the localization method for the robots using the stereo camera and evaluate the precision and accuracy. In Sec. V, we show an example of controlling a robot on the basis of the method. Section VI concludes the paper.

II. RELATIONSHIP WITH PREVIOUS WORK

Localizing mobile robots precisely is a challenging task and many approaches have been proposed. As mentioned before,
one of the most precise methods that is installed in commercially available robots is SLAM technology, which realizes approximately 30–100 mm. Nevertheless, as aforementioned, applying SLAM technologies to localize mobile robots is difficult especially when the environment changes frequently. Furthermore, interference problems have been reported that cause a decrease in precision in light detection and ranging (LiDAR) systems [3].

There are several robust methods that handle changes to environments using landmarks. For example, in Ref. [4], the authors proposed a method to localize on the basis of QR codes placed on the floor. However, unless the QR codes were kept clean, they became difficult to read. Thus, putting such codes or markers on the ceiling, as the authors of Ref. [5] did in their study, is a good idea. Similarly, in Ref. [6], ceiling pattern recognition was proposed. As another approach, localization by impulse radio ultra-wideband signal is used for tracking [7], [8]. However, the above methods lack precision compared with that of commercial guideless conveyance robots. Additionally, the methods cannot find obstacles on the planned path on their own.

III. LOCALIZATION BY STEREO CAMERA

A. Merits of using stereo camera

To overcome such difficulties, we propose a localization method using an external stereo camera installed on a ceiling, as illustrated in Fig. 1. As will be described, we detect the top board of the robots and localize on the basis of the depth data obtained from the stereo camera. How to improve the precision of localization is presented in Sec. IV. As the robots in the system use an external sensing device, the robots do not necessarily have expensive sensors and/or processing units, thereby reducing the financial cost in making the robots.

Additional merits are possible with this architecture. For example, it enables detecting the existence of obstacles even when they are placed out of sight from the robot. A stereo camera at a fixed place can easily detect obstacles on the floor. On the basis of the information about the obstacles, we can calculate the shortest available route to avoid the obstacles to get to the desired location. For another example, we can design a cooperative conveyance system like the one illustrated in Fig. 2, where two robots sandwich an existing dolly, wagon, or cart to carry it without makeover. An image of the robots working is also shown in Fig. 3. To implement this system, it is required that the localization should be performed with a few centimeters precision at most to catch the cart accurately. This system is developed by localizing not only the robots but also obstacles and things to carry in a commensurate way. Although a robot could be localized using optical image recognition data obtained from an external RGB camera instead of from depth data [9], items to be carried are difficult to localize precisely using only this kind of camera when their heights are unknown—especially when the items are placed near the edge of the field of view.

This paper focuses on the details of the localization method and the problems to be solved from the next section on. Note that we will not go deeply into the details of the cooperative conveyance system. The system will be completed as future work, and the details will be published elsewhere.

B. Stereo camera as environment sensor

A stereo camera consists of two optical cameras with spatial separation. The captured objects in sight are compared using two images obtained from each camera as if they were human eyes. With this architecture, we can determine the three-dimensional coordinates of the objects. The depth data correspond to each pixel in a dot-by-dot way, so we can localize many objects using a single stereo camera without the landmarks that are often used [10].

The success of this architecture depends on achieving sufficient precision of localization using a stereo camera with a high mounted position. Commercial conveyance robots without physical guides work with 30–100 mm precision. Thus, we need to achieve at least this level of precision to control mobile robots. The precision of the stereo camera is known to decrease quadratically as the distance between the camera and the object increases. This means the desired precision can be simply achieved to lower the height of the stereo camera. However, this approach deteriorates the availability or convenience of the system because the lower mounting height...
Fig. 4. (a) Cropped optical image of robot. The camera was installed on a 3.2-meter high ceiling. (b) Depth image corresponding to the area shown in Fig. (a). The areas with red (blue) color are far from (near) the stereo camera.

provides a smaller field of view, requiring more stereo cameras to cover the assumed working area of the robots.

We show an example picture and depth data in Figs. 4 (a) and (b), respectively. The precision of the depth data obtained using a stereo camera installed on a 3.2-meter high ceiling is too low to detect or localize robots. In depth data, we cannot see any sharp rectangles. The edges are too blurry. For a still object, the situation is a bit better. This is because we can apply a temporal filter, i.e., the exponential moving average for each pixel of depth data. We know the object does not move, we thus can eliminate temporal fluctuation and localize it. However, this approach cannot be applied to moving objects, including mobile robots.

C. Hardware used in the system

In this section, we introduce the hardware, i.e., robots and stereo cameras, to be used in this experimental system. A photo of the robots is shown in Fig. 5. The size of one robot is 356 mm wide, 535 mm long, and 425 mm tall. The underbody of the robots is made commercially by Vstone Co.,Ltd. and consists of a chassis, a battery, motors, and wheels. We attached a board to the front of the robot, which was installed for a future experiment involving cooperative conveyance. The robot’s body can turn up to 45 degrees to enable it to take a curving route as seen with the (orange) robot on the right in Fig. 3. A laptop computer was put inside the robot to communicate with the controller and regulate the motors.

The speed of the wheels can be controlled independently. If we provide speed that is equal in magnitude and reverse in direction to each wheel, the robot rotates on site. Note that the rotation center does not coincide with the geometric center of the robot’s top board. The rotation center is located 90 mm closer to the front board from the geometric center of the top board.

An inside part of the rectangle on the top board emits white light using built-in LEDs. As described later, we use this light to localize the robot. The circle part is a physical switch (button). The button is designed as an access to the robot directly by people around the robot. The button does not emit light.

For the stereo camera, the RealSense D435 made by Intel Corporation is used. As shown in Fig. 6, the stereo camera has two infrared (IR) cameras. From the disparity, we can calculate the distance between the cameras and the object. The D435 can project an IR pattern onto the objects. Using the projection, the D435 can calculate the distance to uniformly looking objects having few characteristic points. Although the D435 also has a standard optical camera that can take RGB images, we do not use it because of its narrow field of view.

The resolution of the IR and depth images is 1280 × 720 pixels. Each pixel has associated point cloud data that are expressed as three-dimensional coordinates \((x, y, z)\). The origin of the depth data is the location of the stereo camera.

IV. LOCALIZATION OF ROBOTS AND EVALUATION

A. Localization method

The method to localize the robots is divided into two stages. The first stage detects the robots from the IR optical images. The second extracts the location in three-dimensional coordinates of the robot from the depth data. We separate the localization into these two stages because the optical (bright-
Fig. 7. Snapshot of localization application. The red and blue rectangles, arrows, and locations with green numerics are overlayed.

Fig. 8. Robot configuration for evaluation of precision. Robot 1 and 2 were set on the edge and near the center of the field of view, respectively.

ness) information shows higher precision (less deviation) than depth data.

The robot is detected with brightness information. The top board of the robot emits light, so that the robot can be detected as a rotated rectangle that circumscribes the brightest area of the taken IR image. Thus, we can obtain where the robots are in pixel units. We also can determine the orientation of the robots because the long sides of the rectangle are parallel to the direction of the robot. The button on the top board does not emit light, which means it is darker than the rest of the top board. We make use of these characteristics to detect the board side of the robot.

After detecting the robot in pixels, we obtain its physical location in meters using depth data. The stereo camera enables us to obtain the physical location of each pixel of the IR optical images. However, we may fail to obtain some portion of the depth data on the basis of the configuration or position of the robots. To prevent this effect and to improve localization precision, we collect all depth data in the detected rectangle and determine the average to obtain the geometric center of the robot in meters.

We show a working snapshot of the localization in Fig. 7. In this example, we localized two robots at a time. The brightness of the robots was brighter than that of the reflected light, so we can distinguish the robots from other things.

B. Position dependence on localization precision

In this section, we report the evaluation results of localization. We put the stereo camera on a 3.2-meter ceiling, which is the highest part of our laboratory. In this case, an area of approximately $5 \times 3 \text{ m}$ was covered in the field of view.

First, we set Robot 1 on the edge of the field of view and Robot 2 near the center as shown in Fig. 8. The localization results of the $y$-coordinate are shown in Fig. 9 as a function of camera frame number. In this example, the localization was performed at 30 frames per second. As shown in Fig. 9, we could obtain a higher precision when we set the robot closer to the center than to the edge. The standard deviation of the measured $y$-coordinate was calculated as 16 mm and 0.5 mm for Robot 1 and Robot 2, respectively. Note that we rounded the measurement results off to the nearest millimeter and that a rounding error was included. The standard deviation of the Euclidean distance around the averaged point was 30 mm (2.3 mm) on the edge (center) of the field of view.

C. Comparison with other methods

In this section, we compare the proposed method with two existing methods. One is a pixel-based localization, where the position of the robot in millimeters is calculated by multiplying a scale factor by the position in pixels. The other is localization by using a total station, which is a widely used surveying instrument. In this comparison, we move a robot manually by using a remote controller. The origin of the coordinates is set as the center of the field of view of the stereo camera. The robot starts from (1252, 52). The scale factor of the pixel-based localization is set to 3.37 because the start point should be the same among the methods employed here.

The total station consists of an theodolite and an electro-optical distance measuring instrument, which enables us to track the position of the reflection prism within an error of 1 mm. Due to the precision, we set the position obtained by the total station as the ground truth. To enable easier comparisons, we install the reflection prism directly below the geometric center of the robot’s top board that emits light as shown in Fig. 10. The total station that we used was the iX-1005 made by Topcon Corporation. Although a total station can achieve high accuracy and precision, the tracking
fails when the robot moves too fast or there is an obstacle between the instrument and the reflection prism. We thus carefully move the robot at a constant speed of 0.1 m/s. When the total station works successfully, we can obtain the three-dimensional displacement of the robot relative to the position of the total station. We shift the data so that the start point is identical for the other methods.

The localization accuracy is evaluated as the averaged distance from the contour obtained by the total station. The contours obtained in the evaluation experiment are shown in Fig. 11. Note that the rotation and geometric centers of the robot’s top plate are different, which results in the circular sectors seen in Fig. 11 in each corner. For each point of the total station data, the minimum distance to the other contours is calculated. We take the average over the calculated distance to obtain the accuracy. As a result, the averaged error and standard deviation was obtained as 28±21 mm and 42±29 mm for the proposed method and the pixel-based localization, respectively. From the above result, the proposed method performed better performance in localizing conveyance robots than the pixel-based method without any calibration by using the commercially available stereo camera.

V. AUTOMATIC CONTROL OF MOBILE ROBOT

Here, we demonstrate automatic control of the mobile robot with the proposed localization method. This section presents the results of using one robot. The planned route consists of four waypoints as shown in Fig. 12. The communication link was an IEEE 802.11n wireless LAN. We sent the robot its current location \((x_c, y_c)\), orientation \((\theta)\), and destination \((x_t, y_t)\) simultaneously in a UDP packet 30 times a second. For simplicity, we use the rotation center of the robot as the current location. The rotation center is 90 mm ahead of the geometric center. Thus, we calculate \((x_c, y_c)\) as

\[
\begin{align*}
x_c &= x_G + 90 \cos \theta \\
y_c &= y_G + 90 \sin \theta,
\end{align*}
\]

where \((x_G, y_G)\) is the geometric center. Once the robot received the location information, it calculated its motors’ speed to reach its destination with a circular contour, i.e.,

\[
\begin{align*}
v_R &= \max \left( \frac{d - \ell \sin \phi}{d} v, 0 \right) \\
v_L &= \max \left( \frac{d + \ell \sin \phi}{d} v, 0 \right),
\end{align*}
\]

Fig. 12. Supposed contour in demonstration of automatic control of a robot.

Fig. 13. Illustration of mathematical symbols.

\[
\begin{align*}
\end{align*}
\]
where $v_R$ is the moving speed of the right wheel, $v_L$ is that of the left wheel, $v$ is the maximum moving speed of the robot, $\ell$ is the spacing between the two wheels, $d$ is the distance between the robot and the target position, and $\phi$ is the angle between the orientation of the robot and the target position. The mathematical symbols are illustrated in Fig. 13. At every corner, the robot rotated on site to face the next corner. The robot then maintained its input speed until the next speeds were input to the motors. The moving speed of the robot [v in Eq. (1)] was 30 m/minute. When the robot reached within 0.6 m of each waypoint, the speed $v$ was reduced relative to the distance to the waypoint. The robot was considered to arrive at the waypoint when the distance between the robot and the waypoint became less than 0.05 m. The rotation was considered to be finished when the absolute difference between the orientation of the robot and the goal angle became less than 0.02 radians, which approximately equals 1.15 degrees. The control mechanism is summarized in Fig. 14.

The measured contour is shown in Fig. 15. The robot started from the point near (−1000, −500) and went around the rectangle course five times. The dashed red and blue lines in Fig. 15 are the assumed rectangle and the measured contour, respectively. In this experiment, the robot stopped at every corner for at least 100 frames so that we could take an average over 100 frames to obtain the robot’s stopping point with high precision after the experiment.

The mean error between the stopping point and the assumed stopping position was found to be 28 mm. We also calculated the root mean square error of the whole contour and the ideal contour except for the path between the start point and the first corner. The value obtained was 17 mm. On the basis of these obtained values, we conclude our system achieved movement that was as precise as commercial conveyance robots.

VI. SUMMARY AND FUTURE WORK

This paper introduced a calibration-free localization method for robots using a stereo camera installed on the ceiling. The method consists of two stages. First, we detect the bright part of the IR optical images. Second, we extract the depth data corresponding to the area that was detected as being a robot in the first stage and determine the average of the extracted depth data.

In this paper, we showed that the precision was 2.3 mm near the center and 30 mm near the edge of the field of view with a stereo camera installed on a 3.2-meter high ceiling. For the accuracy, i.e., the difference between the proposed method and the ground truth, we evaluated it to be 28 mm on average, which is better than that of the pixel-based localization. We also showed that the automatic control of the robot on the basis of the localization method was possible, and we achieved a 28 mm stopping error in the experiment.

For future work, we would like to develop a cooperative conveyance system, in which a stereo camera is used to detect and localize a cart to be conveyed whose height is not known in advance, as well as the robots.

REFERENCES