Visual Localization through Fiduciary Markers and Particle Filter Drift Calculation

Jamin Goo
Undergraduate Student
The University of Texas at Austin, GDC
Austin, TX, 78705
+1 (512) 674-5556
jgoo@utexas.edu

Shivam Patel
Undergraduate Student
2300 Nueces St Apt 707
Austin, TX, 78705
+1 (319) 651-6181
shivampatel@utexas.edu

Danyaal Ali
Undergraduate Student
1009 West 26th Street Apt 203
Austin, TX, 78705
+1 (281) 943-9887
Danyaal.ali@utexas.edu

Mehrdad Darraji
Undergraduate Student
The University of Texas at Austin, GDC
Austin, TX, 78705
+1 (210) 719-7846
mehrdad_da@utexas.edu

ABSTRACT

The Segbots in the Building Wide Intelligence lab at the University of Texas at Austin currently use the Monte Carlo Localization algorithm (MCL) to keep track of the robot in relation to its map in paired with odometry. Due to its reliance on probabilistic model, the results are not always accurate. In this paper, we investigate this issue to quantify the transitional error of MCL and to compare it to our proposed solution of visual-based localization paradigm based on fiduciary markers using Augmented Reality tags. We performed experiments and noticed a high error rate in the accuracy of the particle filtering system contrasted by an improved accuracy by the visual localization system. This project's work provides a foundational basis to help justify modifying the current Segbots localization mechanism to a more accurate, vision-based system. In addition, the system developed by the group sets groundwork to overhaul the use of odometry and MCL for localization.

1. INTRODUCTION

The current localization system of the Segbots at the Building Wide Intelligence (BWI) lab uses an error-prone methodology in filtering the potential drift of the system's odometry. In specific, the Monte Carlo Localization (MCL) algorithm takes in sensory data from the robot's odometry system and attempts to correct this data based on general errors in the system. In response to this accumulated error-prone system, the group chose to quantify the error within this system. With a quantified error figure, the group could reject the current system of localization to ensure the integration of a more accurate and reliable localization system.

MCL relies on probability to try to find the position of the robot. It generates particles, each of which is a likelihood of robot's current location. Then, MCL determines the location of robot by examining the most dense region of particles [1]. While MCL gives close estimation to the actual position of the robot, it is still very time consuming and takes a long distance of exploration to reach a high level of accuracy. These mere estimations through the sensors are not always accurate since many places could give the same data as other places. Many members in the lab have experienced the inaccuracy when we had an experiment last semester where we tried to localize the robot twice by using a teleoperation command to allow the robot to explore. The results were not very good - only one or two out of thirty people recorded their result as the robot was able to accurately localize. This low percentage is the main objective that our project is trying to tackle.

In our approach, we proposed a landmark-based positioning system using Augmented Reality (AR) tags as fiduciary markers after careful examinations on other options including laser, ultrasonic, and Wi-Fi. Landmarks are passive objects in the environment which can be used to locate a robot with a high degree of accuracy when they are in the vision of a robot [2]. We decided to use landmarks because of the two main reasons: use of vision and no external installation. Vision is far more reliable compared to other solutions like laser or ultrasonic sensors in indoor environment. Since we use existing passive objects, there is no need to install extra hardware. and make the system easily extendable into other environments as long as the robot knows what to look for. The rationale behind using AR tags instead of landmarks in the BWI lab environment is to provide a layer of abstraction and to provide simplicity. Existing tools to detect AR tags can help in minimizing the complexity of the project and let us focus on the positioning system itself.

AR libraries, such as ALVAR, have great compatibilities with the ROS systems. The ‘ar_track_alvar’ package has the capability of generating AR tags, and identifying, and tracking the tag’s position. The reason why we use AR tags as our vision markers instead other possible alternatives, such as QR codes, is because we are only concerned with the detection of vision markers and its relative distance from the Segbot. The general visual depiction of the AR Tags used within this research project are outlined in Figure 0.

Figure 0. Example of Augmented Reality Tags

Almost every system has a differing coordinate frame; in this paper, we work with three different frames: map, robot, and Kinect camera. It is very difficult to work with multiple coordinate frames; however, in order to resolve this, we used ROS’s ‘tf’ package, which keeps track of all the possible frames and their orientation. One useful function that the package proves includes transforming coordinate frames. Transforming all of the coordinate frames in perspective of one object will allow the calculation process to work smoothly.

This paper’s findings will be beneficial to future projects concerning dead reckoning and simultaneous localization and mapping (SLAM), among many others. Having a more accurate positioning system for robots, through the use of fiduciary markers, will help yield positive results when using computer vision. Others will be able to use similar techniques when trying to autonomously localize the robot through similar objects and shapes in their current environment. Overall, this is especially helpful for it takes away the need of external hardware or costly material.

2. RELATED WORKS

I. *Indoor localization using laser scanner and vision marker*

Kim et al. tries to achieve efficient indoor localization algorithm using laser scanner and vision marker. While laser scanner is a widely used medium, laser itself cannot provide a comprehensive result as it returns inaccurate or unreliable data when an indoor environment is crowded. Hence, he uses vision sensor as a complement to increase the accuracy of localization in both dynamic and static environment. His methodology incorporates a collection of sequential data from laser emitting of laser beams and measuring the distance and angle. From the data set, he then calculates the likelihood function to execute the particle filter for localization algorithm. With the vision sensor, the robot grabs an image from its camera and then filters the image by searching for the black contour. After the search, it distincts the vision marker and the black contour using corner color information and this process calculates the angle of vision markers. Then, the robot maps 2D image coordinates to 3D coordinates. Lastly, with the 3D positions and map information, the robot calculates the likelihood function [3]. Since the robot can acquire two different likelihood functions, it allows robot to localize in any given environment as it can choose laser sensor for the static and vision marker for the dynamic environment.

Figure 1. Testing localization with vision markers [3]

This methodology highlights the need of incorporating multiple sensors for comprehensive localization technique. As the author points out, there are pros and cons for every approach. The success rate of laser sensor is depended on how static the targeted environment is. On the other hand, while vision marker allows the robot to achieve localization in this paper, it does not provide autonomous solution in a sense that we have to install vision markers rather than the robot finding out itself. Although the paper does not address how to consider both likelihood functions to accumulate one composite likelihood function, it tells us that the key of using multiple sensors for localization is the interaction of each data set acquired from sensors. Otherwise, localization would not be effective.

II. *Scanning Camera and Augmented Reality Based Localization of Omnidirectional Robot for Indoor Application*

Kundu et al. goes into detail on how inaccurate wheel odometry can be and how AR tags can help increase reliability. The paper points out how the slippage in this system can increase due to many external factors, including differences in ground reaction force and rotational movements. With this, the article explains how they use ARToolkit and how it affected their robot's performance. Kundu et al. identifies some potential difficulties in the process of gathering data from the tags that were proved to be true in our experiments. These include the restriction on the detection of the markers based on distance or angle from where the robot can view the tag. The study ended by showing how the absolute location technique improved the robot's performance. This gives our research the proof of concept it needs to be able to disrupt the current particle filtering system we have on our BWI segbots at this point in time. [4]

3. OUR APPROACH

3.1 Tag Chaining and Mapping

The group proposed a solution which creates a network of tags. In this approach, we compute the location of all tags in the network in terms of the one center tag. In our implementation, we denote the first tag that the robot sees as the center and we call the process of making the network as chaining.

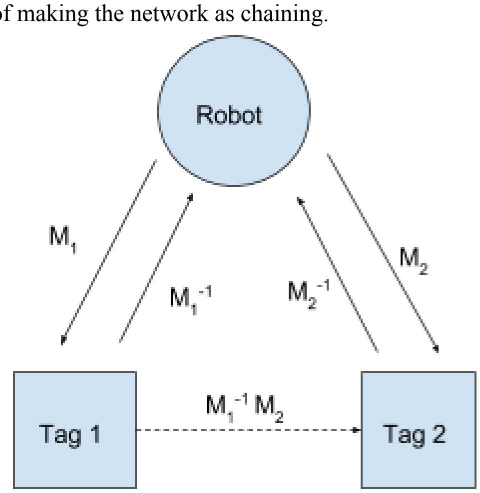

Figure 2. Interaction of Robot and Two Tags on Vision

Figure 2 above highlights how the robot interacts with two tags on its vision and the basics of our methodology. At any time, two tags are on the vision of the robot. As previously mentioned, ar_track_alvar provides the pose of a tag, which in this are M_1 and M_2. The inverse of each provide us the location of the robot. From this interaction, we re-establish the location of the Tag 2 in respect to Tag 1. As Figure 2 portrays, Tag 2's location can be restated as $M_1^{-1} M_2$. Now, the pose of Tag 2 is expressed in terms of Tag 1 and continuation of this process results the network of tags.

When looking at a tag with ar_track_alvar, the message that comes from the subscriber to system is a Pose object of the tag, relative to the camera. The program then converts the Pose object into the /base_link frame. After doing so, the program converts the internal information of the transformed Pose into a 4-by-4 matrix. The details of the newly created 4-by-4 matrix can be covered in Figure 3. This newfound matrix represents the position of the tag with respect to /base_link. As depicted in Figure 2, M_1 represents the matrix's information. Simply inverting the matrix gives you the opposite of the tag's original significance: the position of /base_link with respect to the tag. This information is critical in building our network of AR Tags and data collection.

$$\begin{bmatrix} \multicolumn{3}{c}{} & \text{pose.position.x} \\ \multicolumn{3}{c}{\text{Quaternion to 3x3 Rotation Matrix}} & \text{pose.position.y} \\ \multicolumn{3}{c}{} & \text{pose.position.z} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Figure 3. 4 x 4 Matrix to Represent the Pose of a Tag

The testing environment for this project was created by placing enough tags such that at any frame, the robot can see at least two tags at a time. We are going to use one straight horizontal corridor of the BWI Lab instead of entire circular lab to simplify the testing environment. Having two tags in the vision of the robot at all possible times will allow us to calculate the distance between them in relation to each other and in respect to the first seen tag and make a chain going down the wall more accurately. Currently, our algorithm looks for the first tag and assigns it to the position (0, 0, 0) to set this tag as the center and the position of other tags will be relative to the first tag. In order to find the position of the next tag, we came up with the formula:

$$T_n = T_{n-1} * M_{n-1}^{-1} * M_n$$

A single tag's position is found by multiplying the previous tag's position, the inverse matrix from the robot to the previous tag, and the matrix from the robot to the new tag.

The matrices we are using are a converted format from Eigen and ROS messages into Pure 4x4 Floating matrices. Our group uses rigid transformations to calculate the values and positions regarding the tag's relative positions.

This formula allows for an accurate representation of the new tag's position based in reference to the original tag. Previously, the research group had difficulty deriving the calculation to find the new tag's position based in reference to the original tag. By validating the mathematical background of the Matrix calculations, we are able to create an accurate, visual localization system that sets the groundwork for a truly visual dead reckoning system. In order to understand the math behind the new mapping system, the group was able to derive a brute-force solution of computing the Nth tag's relative position to Tag_0.

Once we found this brute-force solution, we simplified the solution to have an element that stores the build up of the Tag_{N-1}'s position. That way our solution dynamically grows instead of re-computing the new tag's location from the initial tag.

In the process of building our experiment, we need to calculate the robot's position at different moments in order to have consistent experimental design. In order to do so, we use an accumulation of the different elements aforementioned. The initial AR Tag is transformed to the perspective of the map. From there on, the robot moves to a new location where it can see the previous tag and the next one as well. In this location, the robot can calculate its new position based on a standard algorithm used in this experiment. The new ar_track_alvar message of the previous tag signifies the relative position of the previous tag to the robot. We can use the inverse of the relative information to tell us the position of the robot with respect to the previous tag. In addition, we know the location of the previous tag with respect to the map. With these two matrices, we can compute the location of the robot based solely off of visual indications and messages. This gives us the visually calculated robot position with respect to the map.

With the newfound location of the robot independent of the particle filter, we print out the location of both the visually calculated robot position as well as the robot location that the particle filter publishes to /amcl_pose. After printing this information, the robot needs to continue to build the map of AR Tags. To do this, the robot does not move; it also has the next tag in its visual frame. This means it also has the relative position of the next tag with respect to the robot. With this information, the program calculates the location of the next tag using ar_track_alvar's relative tag information and the robot's visually calculated position with respect to the map. This process is visually represented in Figure 4.

After finalizing the calculations at the robot's position, the robot moves to its next location. The next location is predetermined by the testing environment. The robot positions were preset so that the fluctuations in one-dimension are already known. That way the experiment knows the ground truth fluctuation in the one axis being measured. In this fashion, both the particle filter and the visual methods can be rigorously tested against the known truth the prove the better localization technology.

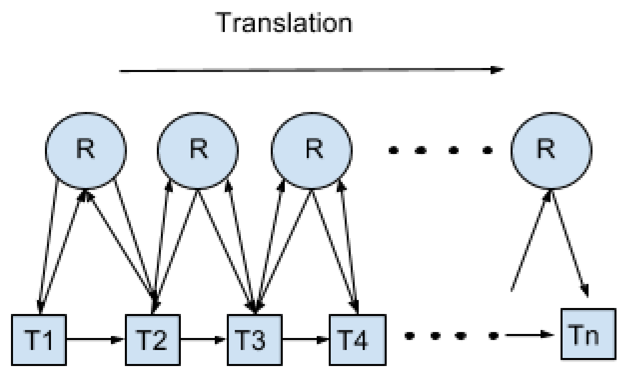

Figure 4. Diagram to Illustrate the Process of Chaining

The aforementioned process stands for the group's approach to the design process and experimentation of the research project.

3.2 Transformation

Our methodology heavily relies on transformation computation between different frame of reference. As previously mentioned, we retrieve the position of a tag using ALVAR in respect to the frame of Kinect.

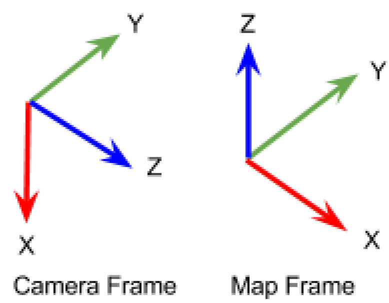

Figure 5. Orientation of Camera Frame and Map Frame

Figure 1 highlights the issue the group ran into with handling different frames and transformations using these frames. The transformation between camera frame and map frame is required in our methodology to calculate the pose of a tag. The problem is the difference in axes and its effect to the formula introduced in section 3.1. The matrix for the pose of the previous tag is in respect to map frame, but the matrix for the pose of the current tag is in respect to camera's frame. The matrix multiplication with the mixed coordinate system gave us inaccurate data. After series of attempts to coordinate the frames into one formula, we decided to rather transform all the incoming messages from ALVAR, which are in camera's frame into base_link, the robot's frame of reference. Such transformation still does not access the odometry and also makes the computation les complicated.

Another problem that the group ran into dealt with the different transforms within the segbot and ar_track_alvar's integration into the system. The messages that ar_track_alvar was displaying on any topic echo would have different coordinate axes compared to the normal and global coordinate axes of ROS. This forced us to transform the messages that came from ar_track_alvar into another tf frame to re-calibrate the different coordinate frame of the messages coming in. In fact, the team ran into many transform issues with the tf tree of the ROS system when converting the incoming AR message into another frame. The main resolution from this was to slow down our program and wait for an available transform between the different coordinate frames.

Besides the time lag from transforming, the specific transformations are critical to understanding how the project functions on a lower level, mathematical basis. The initial transformation follows the same procedure as Figure 6. We take the message from ar_track_alvar, which is respective to /nav_kinect_rgb_optical_frame, and transform it over to map. This absolutely places the tag in the map frame so our initial starting tag is in the map frame. After identifying this tag, each other tag is transformed only to the frame of base_link; this avoids using the particle filter in placing the new tags. After learning about the different transforms for each tag identification, the next element to understand is the Robot-Tag Dynamic when interacting with ALVAR.

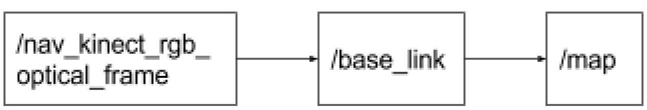

Figure 6. Sequence of transformations from AR Tag to Map Frame

4. MEASURING SUCCESS

In order to evaluate the accuracy of the proposed system, we first set up a testing environment. Our testing environment was set in way that tags are closely located such that the robot can always see two tags at a time and we explicitly mark the places that the robot should be. By marking where the robot should be at, our group restricted the effect of transition of robot only on X-axis of the map frame. To locate the robot at our desired location, we used keyboard teleop to twist the robot. Since the group was in charge of robot navigation, the data we present has inevitable human error.

For the experiment, we decided to use three sets of data per experiment trial: ground truth (set and measured by the team), visually localized data, and particle filter-localized data. These data sets allowed the research group to find the true error of the visually localized data and the particle filter data. By doing so, the group analyzed and concluded the better localization methodology between visual localization and particle filtering against the ground truth.

For a single experiment trial, the robot would be positioned at the same starting location and every position after the initial. Thus, the relative positions between any Nth position and the N-1th position should be the same throughout all of the trials.

When analyzing the datasets of the trials, the relative differences between each point should match the relative differences between the markers pre-set before the experiment. That way, we can analyze the data based on relative values, which also ignores the true initial starting transformation of the initial tag to the map frame.

In terms of our true data output, the group collected data that on average was able to depict the advantage of the visually located data over the particle filter. For each trial, the research group was able create three important visual representations of the data that depicted the error rate of the visual localization and particle filter localization.

A general diagram that signified the overall x-axis change of the experiment can be described in Figure 7. This graph can show the overall x-change and the relative values of the visual localization as well as the particle filter.

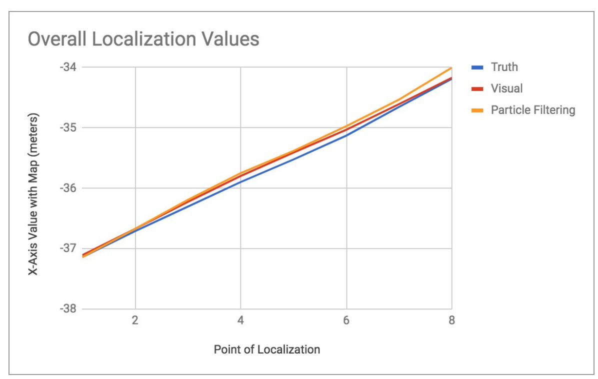

Figure 7. Overall Localization Values Chart on Set 1

With the general purpose of Figure 7, the group delved deeper into the purpose of the drift between the truth and the other two measurements. The group was able analyze the error of the incremental measurements of the two systems. In that, the individual error at each step was quantified to show the error of the visual system and particle filter. As seen in Figure 8, the individual error at each stage tends to grow for the particle filter, signifying the lack of reliability in the particle filter as the robot travels further distances. This general increase in individual error, in turn, explains the larger accumulated error of the visual localization technique over the particle filter, as seen in Figure 9.

Figure 8. Individual Residual Error Chart on Set 1

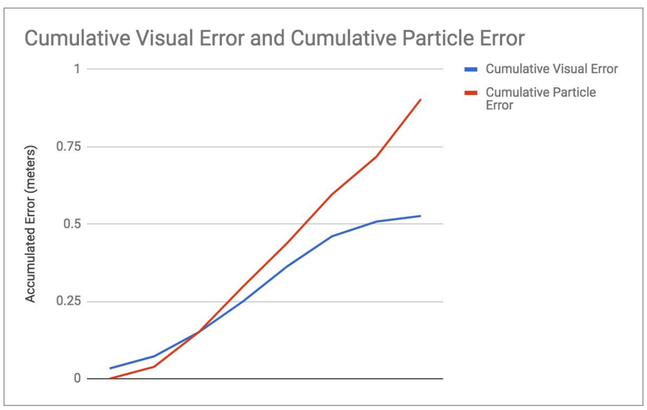

Figure 9. Cumulative Visual Error And Particle Error Chart on Set 1

Through Figures 7-9, the group was able to come to a declaration that the particle filter does in fact increase its error as the robot moves into further stages of the experiment; thus, the particle filter decreases its accuracy as the robot moves further from its initial localized position. On the other hand, the visually-based localization system maintains a lower error level throughout the experiment, signifying a more accurate localization system.

In another experiment, similar results were obtained as well. As described in Figures 10-12, the visually based localization system had a lower error rate once again.

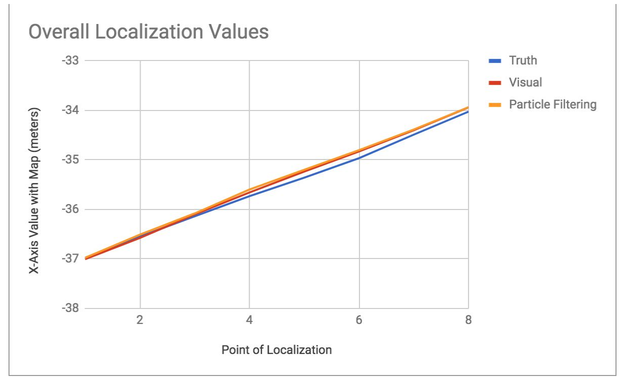

Figure 10. Overall Localization Values Chart on Set 2

Figure 11. Individual Residual Error Chart on Set 2

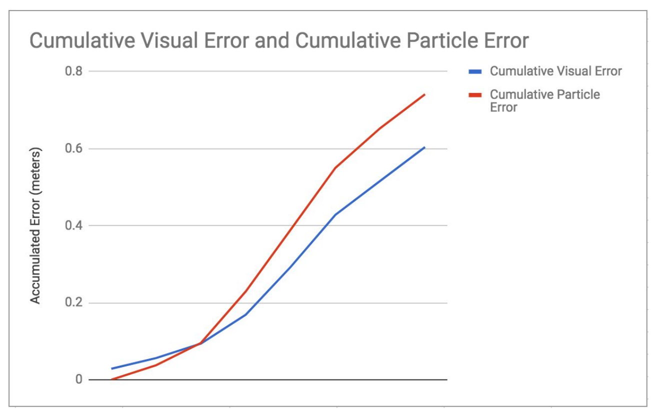

Figure 12. Cumulative Visual Error And Particle Error Chart on Set 2

Through these two different experiment occurrences, the augmented reality and visual-based localization system proved its success over the particle-filter's localization potential. In terms of hard data, the graphs are not able to show the true data points that signify the error of the two localization systems. With the data that correlated to the three main experiment trials, the error rate of the visual system was approximately 20.3 cm/meter traveled. On the other hand, the particle filter had an error of 26.5 cm/meter traveled. With these converted error rates, the visual localization system had a 16.04% increased performance rate compared to the particle filter.

This 16.04% improvement clearly indicates that a visually based system can perform significantly better than the current particle filter system. This improvement can allow for many different future works within the BWI Lab regarding a visually based localization system.

5. FUTURE WORKS

In this project, we used fiduciary markers as an alternative to the use of passive objects in an environment to focus on the effectiveness of landmark-based positioning system. Since our data has proven that the proposed localization system is more accurate than the particle filtering localization, the next step is to replace the markers with passive objects. The door tags could be the suitable replacement for the markers. The door tags have identification numbers that differ by their locations. The mapping of the entire map in terms of the door tag can output complete landmark based localization for the environment.

To increase the reliability and accuracy of the proposed system, there is a need to recalibrate ALVAR and Kinect. Kinect is the source of raw data and high level of calibration is paramount importance to acquire accurate data. Furthermore, customized ALVAR node can help the system to grab more reliable data using vision. Currently, we are using the general purpose node which basically reads and outputs any tag it sees. Since our methodology involves around the computation of two tags at a time, customization could help in acquiring what the system needs to calculate the location of a tag and thus increasing the reliability.

6. CONCLUSION

In conclusion, the research group was able to successfully create a rudimentary visually-based localization system that exceeded expectations and performed better than the built-in particle filter of the BWI-Lab robots.

The process to come from the initial project state to the ending state was very difficult, and the course of the project changed many times. However, the ending result maintained the goal to create a localization system through Fiduciary Markers and disprove the effectiveness and accuracy of the particle filter on the current BWI-Robots. With this conclusive outcome to our experiments, the BWI-Lab can move towards a visually based localization system with a firm basis.

In the end, as a final statement of this research project, the effective visual localization system proved the ineffectiveness of the particle filter and created a just rationale to change localization methodologies of the BWI-Lab to a more accurate and sophisticated system based on visual cues in the environment.

7. REFERENCES

[1] S. Thrun, D. Fox, F. Dellaert, and W. Burgard, “Particle filters for mo-bile robot localization,” inSequential Monte Carlo Methods in Practice,A. Doucet, N. de Freitas, and N. Gordon, Eds. New York: Springer-Verlag, 2001.

[2] Siegwart, Roland, Illah Reza Nourbakhsh, and Davide Scaramuzza. *Introduction to Autonomous Mobile Robots*. 2nd ed. Cambridge, MA: MIT, 2011.

[3] B. Kim, B. Choi, E. Kim and K. Yang, “Indoor localization using laser scanner and vision marker for intelligent robot,” 2012 12th International Conference on Control, Automation and Systems, JeJu Island, 2012, pp. 1010-1012.

[4] Ananda Sankar Kundu, Oishee Mazumder, Ankit Dhar, Prasanna K. Lenka, Subhasis Bhaumik, Scanning Camera and Augmented Reality Based Localization of Omnidirectional Robot for Indoor Application, In Procedia Computer Science, Volume 105, 2017, Pages 27-33, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2017.01.183.