

A Low-Cost Robotic Docking System with Monocular Camera and ArUco Markers*

1st Junseok Oh

School of Electronics and Electrical Engineering
Kyungpook National University
Daegu, Republic of Korea
gsu04295@nate.com

2nd Minyoung Kim (Corresponding Author)

Department of Electronics Engineering
Kyungpook National University
Daegu, Republic of Korea
minykim@knu.ac.kr

Abstract—This study proposes an autonomous charging docking system for transportation robots using a monocular camera and ArUco markers. Transportation robots equipped with only a monocular camera face limitations in accurately measuring the distance and angle to the charger. To address this issue, a system was designed to predict depth and angle by training a regression model on the projected rectangular size changes of three ArUco markers. The proposed system recorded a depth error of 1.18 cm and an angle error of 3.11°, demonstrating superior performance compared to the conventional SolvePnP method (58.54 cm, 6.64°). In docking tests, it successfully achieved charging docking with a distance error within 2 cm and an angle error of 3.07°. This study validates the potential for transportation robots to perform stable and precise charging docking in various environments.

Index Terms—Monocular Camera, ArUco Marker, Regression Model

I. INTRODUCTION

Efficient logistics operations are a critical factor in determining the competitiveness of modern manufacturing, and various robotic technologies are being introduced to achieve this goal. An Autonomous Mobile Robot (AMR), as shown in Fig 1(A), utilizes high-cost sensors such as LiDAR and Depth Cameras to precisely perceive its surroundings and navigate paths. Additionally, an AMR can calculate the distance and angle to a charger using these sensors, enabling automatic charging docking. However, the high cost of these advanced sensors presents a limitation, making it difficult to equip all robots with such capabilities.

As a cost-effective alternative, transportation robots equipped with a monocular camera, as shown in Fig 1(B), have been proposed. These robots can follow an AMR and perform similar functions; however, they face a significant technical challenge in that monocular cameras alone cannot provide distance information about the surrounding environment. This limitation is particularly pronounced when attempting to automate docking, as essential information such as X-axis distance (Depth), angle (θ_b), and Y-axis distance (K) cannot be measured. Figure 1(C) visually illustrates these critical parameters—X-axis distance, angle, and Y-axis distance—required during the docking process.

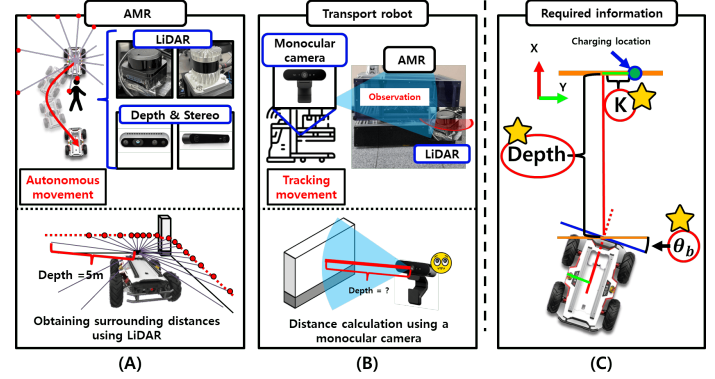


Fig. 1. Differences Between AMR and Transportation Robot for Automatic Charging: (A) AMR Sensors, (B) Transportation Robot Sensors, (C) Information for Automatic Charging

This study proposes a docking system that accurately estimates the distance and angle between a robot and a charger using a monocular camera and three ArUco markers. The proposed system is designed based on the characteristic that the size and deformation pattern of the rectangle surrounding the ArUco marker change depending on the distance and angle. Fig. 2 illustrates the overall flowchart of the system.

As shown in Fig. 2(A), the data required for training were collected using a monocular camera and 2D LiDAR, including the side lengths of the ArUco marker at various distances and angles. These data were used to develop an X-axis distance prediction model and an angle prediction model. The trained models calculate the X-axis distance and angle by utilizing the size of the ArUco marker and observed distance information, which are then used to enable the robot to perform docking.

Additionally, as shown in Fig. 2(B), the size of the rectangle observed by the monocular camera was used as input data to design a regression algorithm-based two-step model. This model first predicts depth and then calculates the angle, enabling precise docking without relying on expensive LiDAR, as illustrated in Fig. 2(C).

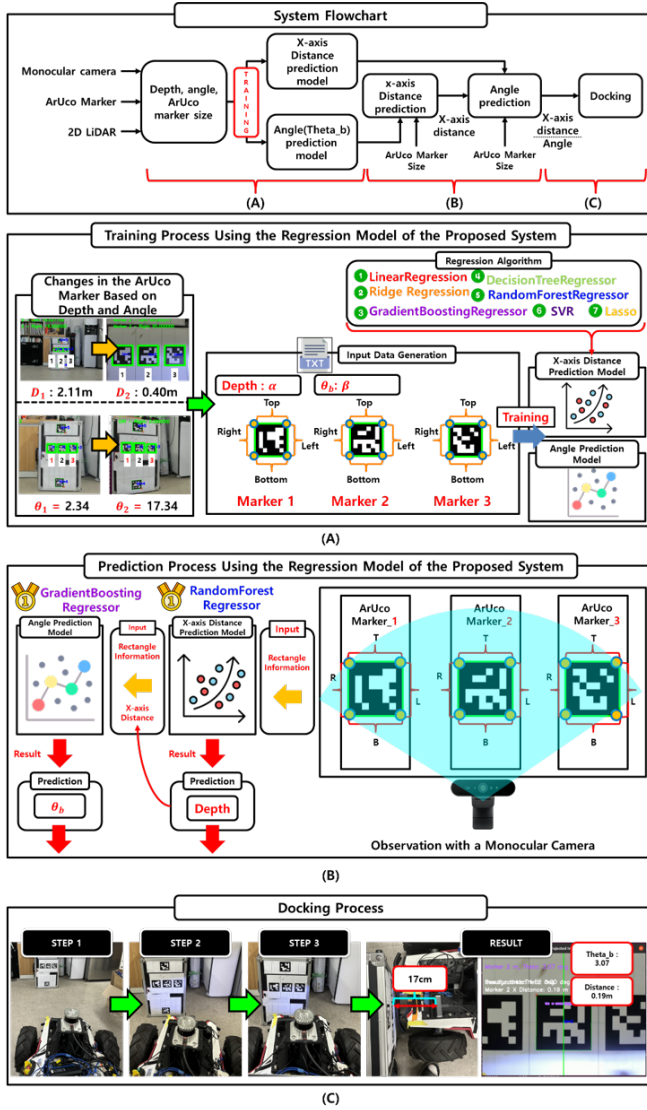


Fig. 2. Flowchart of the Proposed System

II. COMPARISON BETWEEN EXISTING SYSTEMS AND THE PROPOSED SYSTEM

A. 3D Pose Estimation Based on SolvePnP Using a Monocular Camera and an ArUco Marker

A monocular camera and an ArUco marker are widely utilized as cost-effective solutions for estimating the position and orientation of objects in 3D space. The SolvePnP algorithm is a key method that leverages these two devices to estimate the relative position and orientation between the camera and the marker. The fundamental principle of SolvePnP is illustrated in Fig. 3. When an ArUco marker is placed perpendicular to the camera, it is projected as a perfect rectangle in the image. However, when the marker is tilted, it appears as a distorted rectangle due to the perspective projection effect. This distorted shape is analyzed using the coordinates of the four corners of the marker, which serve as the basis for calculating the Yaw, Pitch, and Roll angles. Additionally,

SolvePnP utilizes the Z-axis value of the translation vector to estimate the distance between the camera and the marker. This enables the accurate acquisition of depth information using only a monocular camera, making it a cost-effective alternative to expensive depth cameras or LiDAR. However, as the distance between the camera and the marker increases, the projected size of the marker decreases, leading to reduced resolution and signal-to-noise ratio (SNR), which in turn lowers the accuracy of pose and distance estimation.

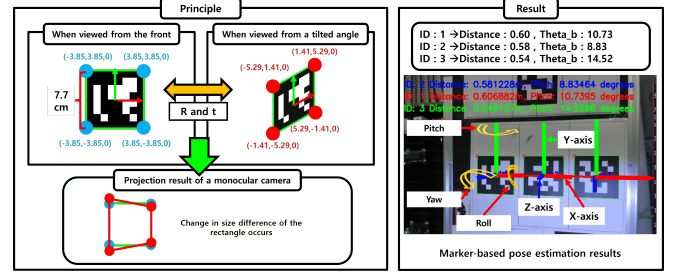


Fig. 3. Depth and Pose Estimation Method Based on SolvePnP Using a Monocular Camera and ArUco Markers

B. Method for Obtaining the Pose Between the Camera and the Marker Using a Monocular Camera and an ArUco Marker in This Study

In this study, a system was implemented to predict the X-axis distance and angular information to the charger using a monocular camera and three ArUco markers. ArUco markers, based on unique patterns, minimize the effects of lighting changes and background colors, offering advantages that overcome the limitations of conventional methods [3]. Due to these characteristics, ArUco markers were adopted in this study.

The proposed system analyzes the side lengths (Top, Bottom, Left, Right) of the rectangle surrounding the ArUco markers projected onto the monocular camera to estimate the X-axis distance (Depth) and angle (Θ_{ab}) to the charger. As shown in Fig. 4(A), the side lengths of the rectangle decrease inversely with the distance (Depth) between the camera and the charger. Additionally, as illustrated in Fig. 4(B), the deformation pattern of the rectangle changes with the angle (Θ_{ab}) due to the perspective projection effect. Based on these characteristics, a regression model was trained on the patterns of changes in distance and angle. Using the observed side lengths as input data, the model was developed to predict the X-axis distance (Depth) and angle (Θ_{ab}).

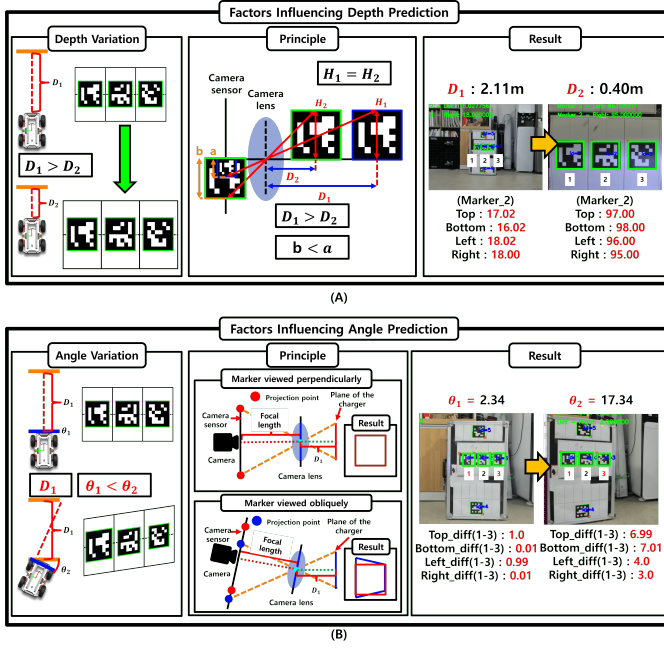


Fig. 4. Rectangle Size Changes Around the ArUco Marker by Distance and Angle: (A) Changes by Distance, (B) Changes by Angle

Predicting angles using only the size of the rectangle is challenging because the rectangle's size is influenced by the interaction between distance and angle. To address this issue, a method of fixing the distance during angle prediction model training was introduced. As shown in Fig. 5, the distance was segmented in 5 cm intervals, and angle data corresponding to each segment were used for training to focus on analyzing the changes in rectangle size caused by angle variation. This approach significantly improves the accuracy of the prediction model.

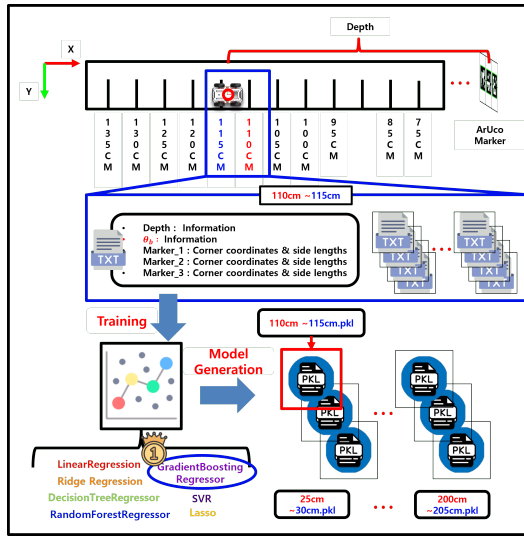


Fig. 5. The Process of Fixing Distance to Develop an Angle Prediction Model

In addition to the X-axis distance (Depth, D) and angle

(Θ_{ab}) required for charger docking, the Y-axis distance (K) must also be calculated. As shown in Fig. 6, the Y-axis distance (K) is computed using the triangulation method, which requires the pixel distance, depth information (D), and camera focal length (f) [2]. The triangulation method, when combined with conventional data-driven approaches, provides more accurate docking information.

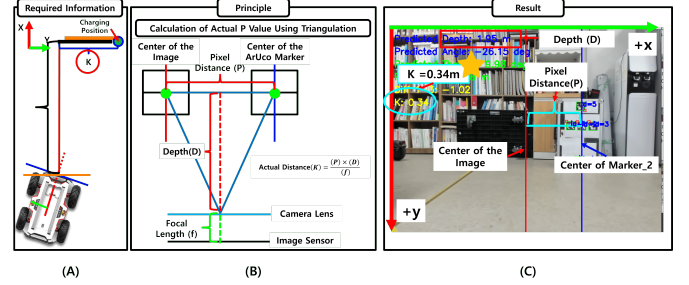


Fig. 6. The Process of Calculating the K Value Using Triangulation

III. DATA ACQUISITION PROCESS AND REGRESSION MODEL TRAINING RESULTS

To build accurate prediction models for the X-axis distance (Depth) and angle (Θ_b), data were collected and trained across various distances and angles. Manual data collection is time-consuming and makes it difficult to ensure data consistency, so an efficient and precise data collection method using a 2D LiDAR and a monocular camera was implemented. Figure 7(A) visually represents the information obtained from each sensor: the monocular camera provides the projected image of the charger, while the LiDAR offers coordinate data to the charger. To accurately determine the charger's position by combining the information from each sensor, it is necessary to integrate the coordinate systems of the monocular camera and LiDAR, as shown in Figure 7(B). For this integration, internal and external parameters are required.

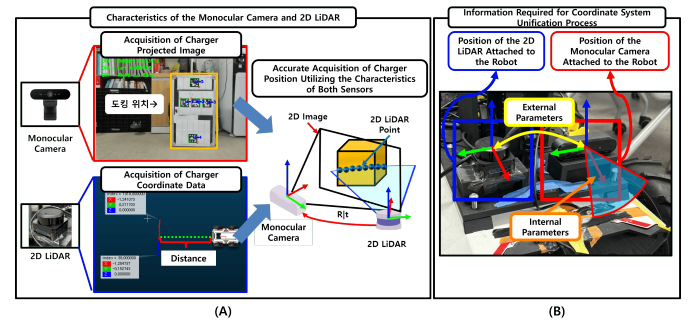


Fig. 7. Characteristics of the Monocular Camera and 2D LiDAR and Information Required for Coordinate System Unification Between the Two Sensors: (A) Information obtainable from each sensor, (B) Internal and external parameters

A. Coordinate System Unification Between the Monocular Camera and 2D LiDAR

To unify the coordinate systems, the internal and external parameters of the camera were calculated. Fig. 8(A) shows

the process of extracting internal parameters (focal length, principal point coordinates, lens distortion coefficients, etc.) using a checkerboard. Fig. 8(B) illustrates the process of calculating the transformation relationship (rotation matrix R and translation vector T) between the LiDAR and the camera using the SolvePnP algorithm. Through this process, as shown in Fig. 8(C), LiDAR data were projected onto the monocular camera image, enabling the data from both sensors to be interpreted within a unified framework.

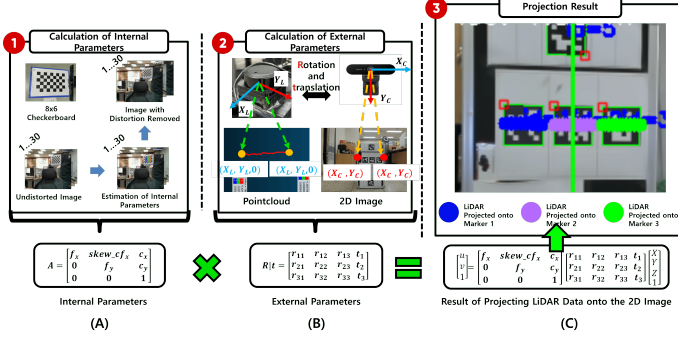


Fig. 8. Unifying Coordinate Systems Between Monocular Camera and 2D LiDAR: (A) Internal Parameter Calculation, (B) External Parameter Calculation, (C) 2D LiDAR Projection onto Image

B. Method for Data Acquisition (X-axis Distance (Depth), Θ_b , Rectangle Side Lengths)

To collect X-axis distance (Depth) and angle (Θ_b) data, a 2D LiDAR was utilized. As shown in Fig. 9(A), the LiDAR provides the X-axis distance closest to the coordinates of the center of Marker 2. As illustrated in Fig. 9(B), the angle (Θ_b) is calculated based on the leftmost and rightmost points of Marker 2's coordinates. Through this process, as shown in Fig. 9(C), data on the side lengths (Top, Bottom, Left, Right) of the ArUco marker were collected. The X-axis distance (Depth) between the camera and the marker, the angle (Θ_b), the side length information (Top, Bottom, Left, Right), and the corner coordinates of each ArUco marker were all stored in a txt file for use in model training.

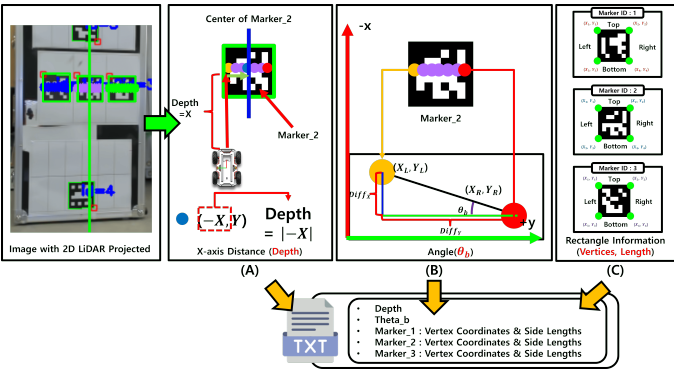


Fig. 9. Obtaining X-axis Distance, Angle, and Rectangle Side Lengths Using LiDAR Data: (A) X-axis Distance, (B) Angle, (C) Side Lengths of Rectangles 1, 2, and 3

C. Training and Testing Results of Distance and Angle Prediction Models

Various regression algorithms (Linear Regression [4], Ridge [5], Lasso [6], Decision Tree [7], Random Forest [8], SVR [9], Gradient Boosting [10]) were used to train prediction models for predicting X-axis distance (Depth) and angle (Θ_b). After splitting the data into training and testing sets in an 8:2 ratio, the performance of each algorithm was evaluated, and the best model was selected.

- X-axis distance (Depth) prediction: The Random Forest model showed the highest performance, with a Mean Squared Error (MSE) of 0.0018 and an R^2 score of 0.9886.
- Angle (Θ_b) prediction: For angle prediction, as shown in Figure 10, the X-axis distance was divided into 5 cm intervals, and individual models were trained for each interval. The optimal model varied by distance. Across all intervals, the models demonstrated stable prediction performance with an average MSE of 1.5185 and an R^2 score of 0.915.

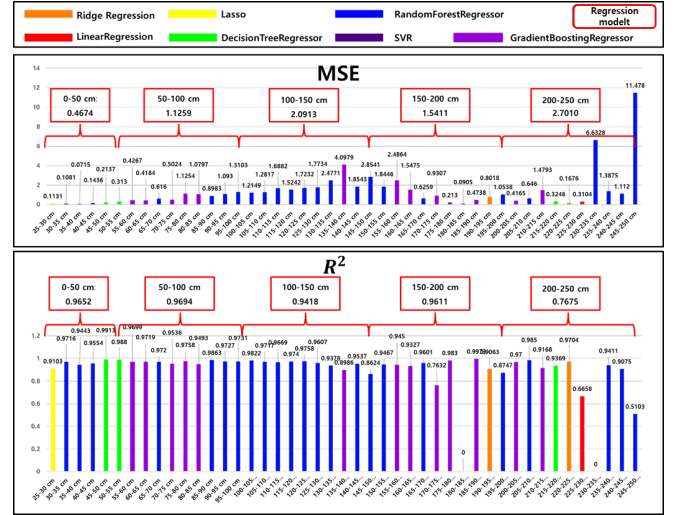


Fig. 10. Learning results of the angle (Θ_b) prediction model

Based on the evaluation results from the test data, as shown in Figure 11(A), the X-axis distance (Depth) prediction using the Random Forest model produced results similar to the actual values. For angle (Θ_b) prediction, the optimal model varied depending on the distance, and predictions were made using the most suitable model for each range, yielding results close to the actual values. For example, as shown in Figure 11(B), predictions were made using Ridge regression for the 190–195 cm range, Random Forest regression for the 85–90 cm range, Lasso regression for the 25–30 cm range, Gradient Boosting regression for the 155–160 cm and 170–175 cm ranges, and Decision Tree regression for the 45–50 cm range, all of which produced results similar to the actual values.

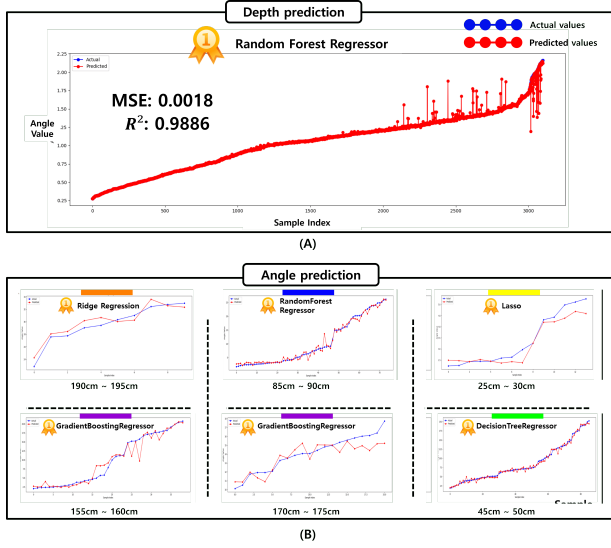


Fig. 11. Model test results, (A) X-axis distance (Depth) test results, (B) Angle (Θ_{ab}) prediction test results

D. Comparison of performance between the proposed system and the existing system

The proposed system outperformed the existing SolvePnP method in both depth and angle prediction.

- X-axis Distance (Depth): The proposed system achieved an average error of 1.18 cm compared to 58.54 cm for SolvePnP (Fig. 12(A)).
- Angle (Θ_{ab}): The average error was 3.11° for the proposed system, significantly lower than SolvePnP's 6.64° (Fig. 12(B)).

This demonstrates the system's superior accuracy and practicality, maintaining high precision even at long distances.

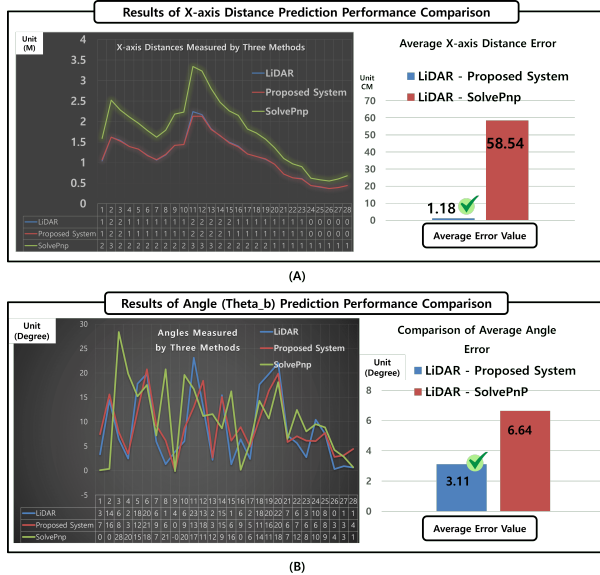


Fig. 12. Performance Comparison of the Existing and Proposed Systems Using LiDAR Data: (A) X-axis Distance Prediction, (B) Angle Prediction

IV. DOCKING METHOD AND PROCESS TO THE CHARGER

A. Docking Method Based on the Position of the Charger and the Robot

The docking process is divided into a total of eight types, as shown in Fig. 13, depending on the relative position and orientation of the robot and the charger. These docking methods are determined by the robot's angle (Θ_{ab}), Y-axis distance (K), and an auxiliary point (Point A). Point A serves as an intermediate target point for docking, designed to ensure that the robot can approach the charging position stably.

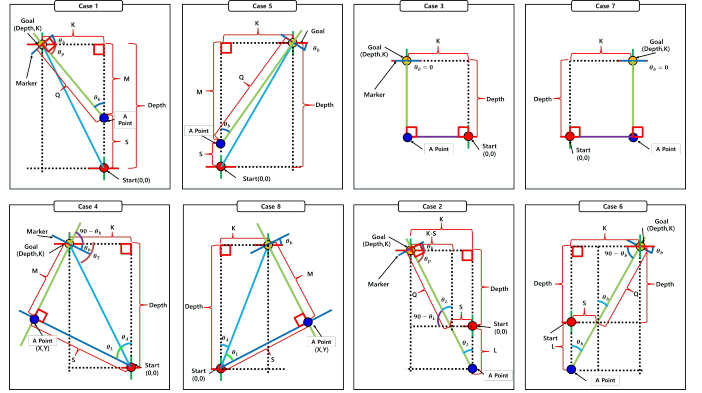


Fig. 13. Various Docking Methods Based on the Positions of the Robot and the Charger

B. Docking Process and Results for Case 1

The docking process for Case 1 is illustrated in Fig. 14 and consists of the following three stages:

- 1) Calculate: The values for Depth (distance), Θ_{ab} (angle), and K (Y-axis distance) are calculated, and the intermediate target point, Point A, is set for docking.
- 2) *Moving₁*: The robot moves straight to Point A.
- 3) *Moving₂*: After reaching Point A, the robot rotates by the calculated Θ_{ab} angle.
- 4) *Moving₃*: Finally, the robot moves straight to the docking position and completes docking.

The final post-docking state showed an X-axis distance (Depth) of 19 cm and an angle (Θ_{ab}) of 3.07° , confirming successful docking. With a fixed distance of 17 cm between the robot's front and the camera, the actual docking error was measured as 2 cm. The angular error was only 3.07° , demonstrating high precision.

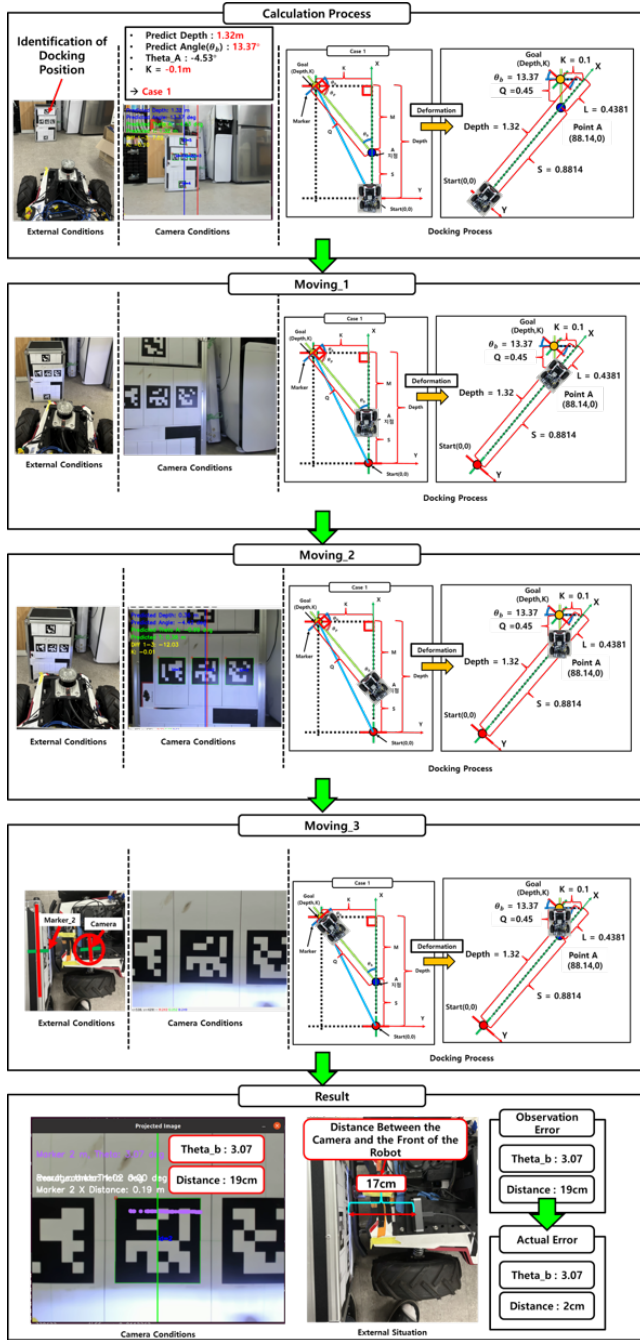


Fig. 14. Docking Process and Results for Case 1

V. CONCLUSION

This study proposed a low-cost robotic charging system that automates docking with a charger by precisely predicting the X-axis distance (Depth) and angle (θ_b) and calculating the Y-axis distance (K) using a monocular camera and ArUco markers. By analyzing the deformation characteristics of the projected rectangle surrounding the ArUco marker, a model was developed to estimate depth and angle without the need for expensive sensors. This demonstrates the feasibility of enabling transportation robots to autonomously charge without

human intervention using only a monocular camera.

Performance comparison results showed that the proposed system achieved an average error of 1.18 cm in depth prediction and 3.11° in angle prediction, significantly outperforming the conventional SolvePnP method (depth: 58.54 cm, angle: 6.64°) in precision. Notably, the proposed system supports various docking cases, and testing confirmed successful docking with an accuracy of 2 cm in distance and 3.07° in angle.

This study validates the practicality of a monocular camera-based docking system and makes a significant contribution to the adoption of autonomous robots in manufacturing and logistics. Future research will focus on improving the accuracy and generalizability of the prediction model by incorporating more diverse data and scenarios.

ACKNOWLEDGMENT

This research was supported by the Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE) (Project No. P0020536, The Competency Development Program for Industry Specialists), by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) (Project No. 2022R1A2C2008133), and by the Core Research Institute Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (Project No. 2021R1A6A1A03043144).

REFERENCES

- [1] Y. Wang, Z. Zheng, Z. Su, G. Yang, Z. Wang, and Y. Luo, "An improved ArUco marker for monocular vision ranging," in Proc. Chinese Control and Decision Conf. (CCDC), Hefei, China, 2020, pp. 2915–2919, doi: 10.1109/CCDC49329.2020.9164176.
- [2] I. V. J. Lau, A. I. Sacil, and J. B. G. Ibarra, "Triangulation method for camera-based distance measurement with application for the blind," in Proc. Int. Conf. on Information Communication and Software Engineering (ICICSE), Beijing, China, 2024, pp. 80–85, doi: 10.1109/ICICSE61805.2024.10625678.
- [3] 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, doi: 10.1016/j.patcog.2014.01.005.
- [4] A. N. Tikhonov, "Solution of incorrectly formulated problems and the regularization method," Soviet Mathematics Doklady, vol. 4, pp. 1035–1038, 1963.
- [5] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," Technometrics, vol. 12, no. 1, pp. 55–67, 1970, doi: 10.1080/00401706.1970.10488634.
- [6] R. Tibshirani, "Regression shrinkage and selection via the lasso," J. Roy. Statist. Soc. B, vol. 58, no. 1, pp. 267–288, 1996.
- [7] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, Classification and Regression Trees. Monterey, CA, USA: Wadsworth and Brooks/Cole, 1984.
- [8] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [9] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," Statistics and Computing, vol. 14, no. 3, pp. 199–222, 2004, doi: 10.1023/B:0000035301.49549.88.
- [10] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001, doi: 10.1214/aos/1013203451.