Hierarchical 6D Pose Estimation Strategy for Mobile Manipulator

1st Hyeonwook Song
Department of Electrical Engineering
Hanyang University
Seoul, Republic of Korea
hwhy2024@hanyang.ac.kr

2nd Yang-Jin An
Department of Electrical Engineering
Hanyang University
Seoul, Republic of Korea
emdydqortkgh@hanyang.ac.kr

3rd Chang Mook Kang
Department of Electrical Engineering
Hanyang University
Seoul, Republic of Korea
kcm0728@hanyang.ac.kr

***Abstract*—This paper presents a hierarchical 6D pose estimation strategy for automated wheel replacement using a collaborative robot on a 1:10 scale platform. The proposed perception pipeline employs a coarse-to-fine approach, integrating YOLO v11-based PnP for initial alignment with CHT-PCA refinement for precise nut engagement. Experimental results demonstrate a 95% success rate—significantly outperforming the 54% baseline—confirming the system's robustness and applicability in unstructured environments**

***Index Terms*—Collaborative robot, Wheel replacement automation, Deep learning vision, YOLO v11-Pose, Perspective-n-Point (PnP), Circular Hough Transform (CHT), Principal Component Analysis (PCA), Integrated end effector design**

I. Introduction

With the maturation of collaborative manipulation and learning-based perception, mobile manipulation is becoming viable for maintenance tasks requiring precision and repeatability. This paper targets automated wheel replacement by coupling a deep learning vision pipeline with a single collaborative arm and a compact, integrated end-effector that combines an impact driver, RGB camera, and gripper on a uFactory Lite6. In the intended deployment, an autonomous mobile robot (AMR) first navigates to the vicinity of a target wheel; in this work, experiments assume the AMR has completed that approach, and the collaborative arm performs loosening, removal, and reinstallation of the wheel.

Recent studies have demonstrated that integrating deep learning-based machine vision with robotic arms enhances the reliability and efficiency of industrial inspection and manipulation. Mao et al. [1] developed a robot arm system equipped with a deep learning network for rim defect inspection, achieving robust object recognition and seamless automation on a practical robotic platform. Li et al. [2] proposed a 3D hand-eye calibration method for collaborative robot arms, improving spatial calibration accuracy and enabling more reliable manipulation in unstructured manufacturing scenarios.

The perception stack operates in two stages: YOLO v11-Pose detects keypoints of the wheel and nuts to estimate the initial 6-DoF pose via the PnP algorithm. This allows the robot to align with the wheel axis, after which a geometric refinement stage using the Circular Hough Transform (CHT) and Principal Component Analysis (PCA) estimates a precise 6-DOF pose for nut engagement and wheel handling.

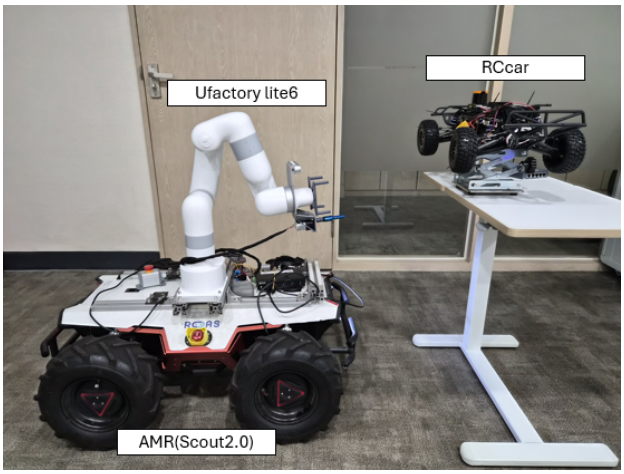

Fig. 1. Mobile manipulator platform used in this study: a uFactory Lite6 collaborative arm mounted on an AMR-sized wheeled base with an integrated end-effector (impact driver, camera, gripper) for automated wheel removal and installation; the AMR brings the arm near the target wheel, and the arm executes loosening, removal, and replacement

A custom end-effector mounted on the Lite6 executes the full task sequence—nut loosening with the impact driver, wheel grasping with the gripper, wheel removal, and reinstallation—without tool changes or multiple manipulators. Validation on a 1/10 scale RC car platform demonstrates fully automated nut recognition and wheel exchange with improved success rate and efficiency, indicating practicality for scaled-up, real-world mobile manipulation workflows.

We developed a compact integrated end effector combining an impact driver, camera, and gripper to enable single-arm collaborative robot automation of the wheel replacement process. A two-stage vision pipeline using YOLO v11 for detection and CHT-PCA for precise pose estimation significantly enhanced manipulation accuracy. This reliable and flexible system was validated on a scaled platform, demonstrating its effectiveness in dynamic industrial environments.

II. System Architecture

The proposed system is based on the uFactory Lite6 collaborative robot (6-axis) and a 1/10 scale RC car platform, as illustrated in Fig. 3. The end effector was custom designed

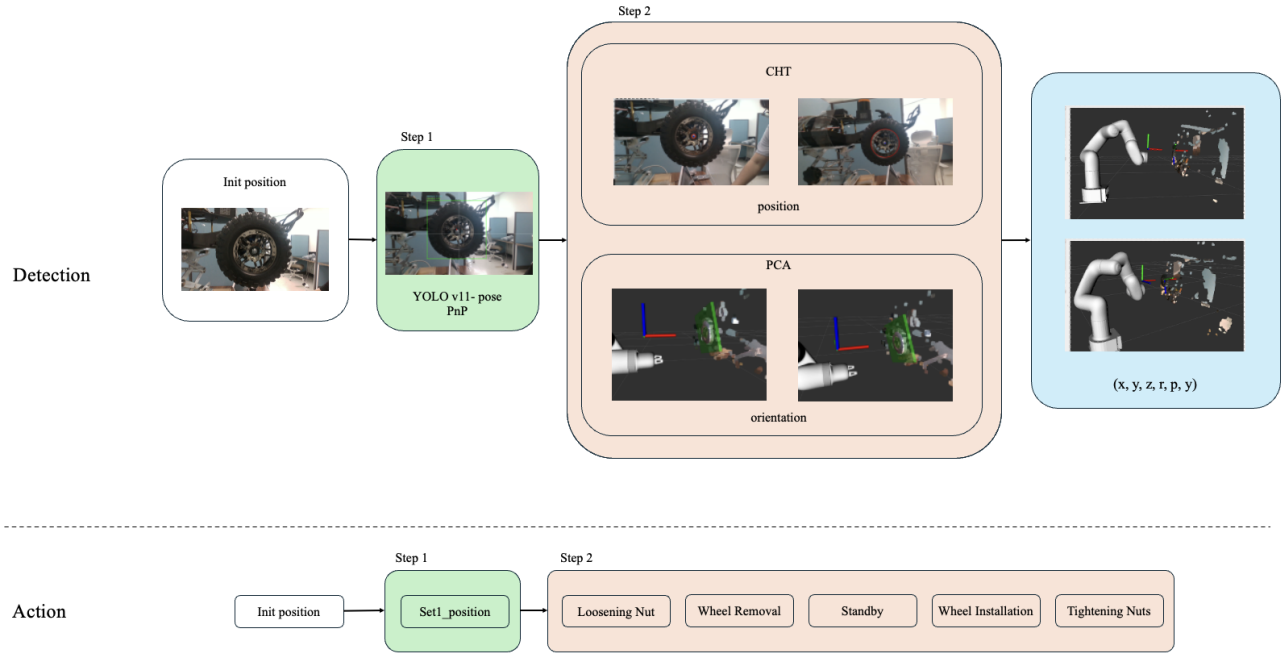

Fig. 2. System architecture showing a two-step detection and action process: initial positioning using YOLO v11 detection for coarse localization (Step 1), followed by precise recognition using CHT and PCA algorithms for accurate position and orientation detection (Step 2), enabling sequential robotic actions including Loosening Nut, Wheel Removal, Standby, Wheel Installation, and Tightening Nut.

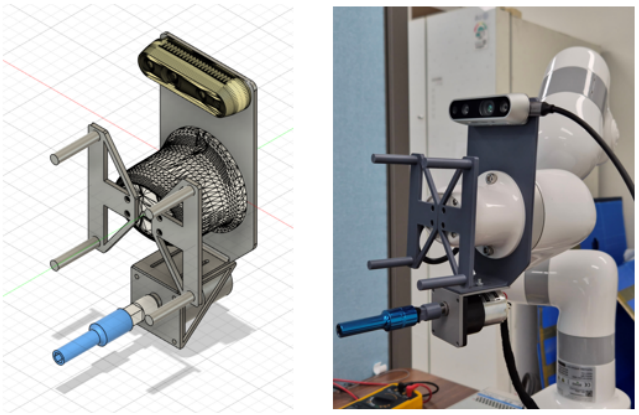

Fig. 3. Hardware platform overview: (left) CAD design of the integrated end effector system, (right) implemented system with uFactory Lite6 collaborative robot and 1/10 scale RC car platform

and fabricated to integrate a DC motor-driven impact driver, an RGB camera, and a gripper as a single unit. As shown in the CAD design (Fig. 3, left), the end effector features a compact, multi-functional design that consolidates all necessary tools for wheel replacement operations. The implemented system (Fig. 3, right) demonstrates the practical integration of these components mounted on the robot arm. This configuration enables the robot to perform all wheel replacement processes, including nut loosening, wheel gripping, removal, and replacement, while maintaining the flexibility and safety characteristics essential for collaborative robotics applications.

The end effector design incorporates three key components strategically positioned for optimal functionality: the RGB camera provides visual feedback for the perception pipeline, the custom-designed gripper enables secure and adaptive grasping of the wheel for precise handling, and the impact driver delivers the necessary torque for nut manipulation. The modular design allows for independent operation of each component while ensuring spatial efficiency and minimizing interference during complex manipulation tasks. The vision system employs an Intel RealSense D435 camera, which offers both RGB and depth sensing capabilities, enhancing spatial awareness and enabling precise 3D pose estimation within the ROS2 framework. This integration supports real-time sensor data processing and robot control, ensuring robust synchronization between perception and manipulation.

III. Method

A. Vision and Perception

The vision and perception module operates in two sequential stages to ensure accurate and robust recognition of wheels and nuts for autonomous manipulation, as demonstrated by the system architecture in Fig. 2.

1) Step 1 (Wheel and Nut Detection using YOLO v11): The automated wheel replacement process initiates from the

collaborative robot's home position. The end effector's RGB-D camera captures the workspace, and the system employs the YOLO v11-Pose model trained to detect keypoints corresponding to the wheel center.

Unlike standard 2D detection, the proposed system leverages the depth sensing capabilities to enhance pose estimation accuracy. For each detected 2D keypoint, the corresponding depth value is extracted from the depth map to derive its 3D spatial coordinate in the camera frame. Using these observed 3D points and the known 3D geometric model of the wheel, the system solves the Perspective-n-Point (PnP) problem to estimate the coarse 6-DoF pose (T_{cam}^{wheel}) of the target wheel.

Based on this depth-integrated pose estimation, the robot controller generates a trajectory to align the camera's optical axis with the wheel's rotation axis. The robot then moves to a 'set1 position,' maintaining a distance of 20–30 cm perpendicular to the wheel surface. This alignment, facilitated by accurate depth measurements, minimizes perspective distortion and maximizes the robustness of the subsequent CHT-PCA precision recognition in Step 2.

For the experimental implementation, we collected a comprehensive dataset comprising 442 images and annotated keypoints for wheel centers and nuts to train the YOLO v11-Pose model. Fig. 4 illustrates the training performance and convergence behavior.

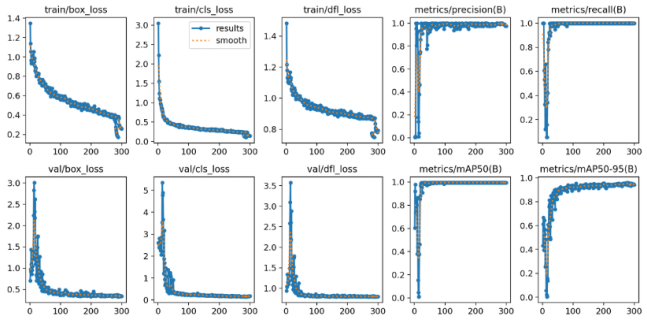

Fig. 4. YOLO v11 training graph showing loss curves and performance metrics over 300 training epochs

As shown in Fig. 4, the training metrics demonstrate stable convergence with decreasing loss values across all categories for both training and validation sets. The precision and recall metrics, along with mAP50 and mAP50-95 scores, show consistent improvement throughout the training process, ultimately achieving precision and recall values exceeding 0.95 and mAP50 scores approaching 1.0. The close alignment between training and validation curves indicates minimal overfitting, validating the effectiveness of our custom dataset for wheel and nut detection tasks.

2) Step 2 (Nut Pose estimation using CHT and PCA): Following the strategic positioning to the set1 location determined in Step 1, the system proceeds with a secondary high-precision recognition process that employs a combined CHT-PCA algorithmic approach to achieve enhanced spatial accuracy. The CHT is applied to the localized image patch to detect and extract the circular features of both the wheel rim and individual nuts, providing precise 2D spatial coordinates within the image plane. However, considering that the wheel axis may not be perfectly perpendicular to the robot's coordinate frame due to vehicle positioning variations, PCA is subsequently employed to determine the three-dimensional orientation of the wheel assembly. The PCA algorithm analyzes the geometric distribution of detected edge and contour points to extract the roll, pitch, and yaw (RPY) angular values, characterizing the wheel's spatial orientation relative to the robot's reference frame. By integrating the spatial coordinates obtained from CHT with the orientation information derived from PCA, the system generates a comprehensive 6-DOF pose estimation that significantly enhances the precision and reliability of subsequent robotic manipulation tasks, thereby improving overall task completion accuracy in challenging operational environments.

Algorithm 1 Wheel Pose Estimation using CHT-PCA

Input: RGB image I, depth image D, camera intrinsics (f_x, f_y, c_x, c_y)

1) **Preprocessing:**
Extract wheel ROI from I using YOLO detection. Convert ROI to grayscale: $I_{\text{gray}} = \text{cvtColor}(I, \text{BGR2GRAY})$. Apply Gaussian blur: $I_{\text{blur}} = \text{GaussianBlur}(I_{\text{gray}}, \sigma = 1.0)$. Extract edge points $P = \{(u_i, v_i)\}$ within wheel region.
2) **Circular Hough Transform (CHT):**
Detect circle parameters (u_c, v_c, r) from I_{blur} using HoughCircles with minRadius $= 50$, maxRadius $= 120$.
3) **Depth-based 3D Projection:**
Compute mean depth: $d = \text{mean}(D[u_c - 1 : u_c + 2, v_c - 1 : v_c + 2])$. Calculate 3D center:

$$z = \frac{d}{1000}, \quad x = \frac{(u_c - c_x)z}{f_x}, \quad y = \frac{(v_c - c_y)z}{f_y}$$

Wheel center: $\mathbf{p}_{\text{center}} = (x, y, z)$.
4) **PCA for Wheel Plane Orientation:**
Collect 3D edge points P_{3D} within band $(r-3) < \text{distance} < (r+3)$ and $D[u, v] > 0$. Compute mean: $\bar{\mathbf{p}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{p}_i$. Compute covariance matrix:

$$\Sigma = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{p}_i - \bar{\mathbf{p}})(\mathbf{p}_i - \bar{\mathbf{p}})^T$$

Perform SVD: $\Sigma = USV^T$. Extract plane normal: $\mathbf{n} = U[:, -1]$ (eigenvector of smallest eigenvalue). Compute roll-pitch-yaw (r, p, y) from $\mathbf{n}$.

Output: Wheel pose $\{x, y, z, r, p, y\}$

In particular, a dataset consisting of 100 images that capture a variety of 1/10 scale model car wheels was constructed to quantitatively evaluate the nut recognition accuracy of the CHT algorithm. For each image, the recognition results were compared with manually annotated ground truth and analyzed using the Intersection over Union (IOU) metric. Additionally, two representative real RC model wheels were examined in dedicated experiments to verify performance in real-world scenarios.

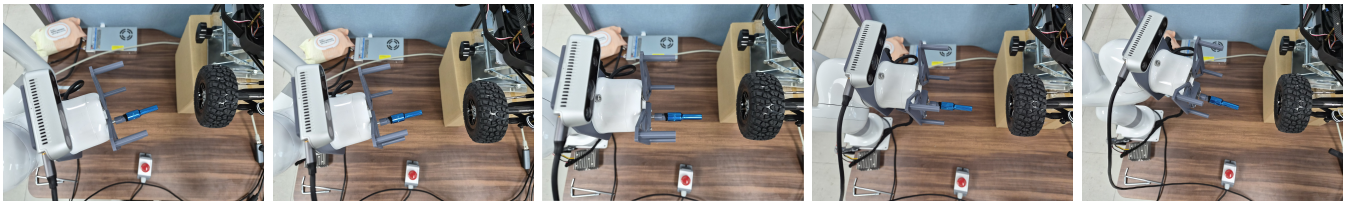

Fig. 5. Experimental setup for evaluating pose invariance under yaw variation: The robotic arm and the target wheel were positioned at $-10°$, $-5°$, $0°$, $5°$, and $10°$ yaw (from left to right). Each configuration was tested with 20 independent trials.

B. Manipulation and Task Scenario

Based on the perception information, the collaborative robot follows an optimized trajectory to approach the target. Using the integrated end effector, it sequentially loosens the nuts and grips, removes, and replaces the wheel. The overall system workflow consists of detection, precise position correction, impact control, gripper operation, and wheel replacement, as illustrated in Fig. 2.

Wheel Removal Scenario: The wheel removal process follows a systematic two-stage approach where Step 1 utilizes YOLO v11 detection to identify the target wheel and calculate the optimal set1 position for precise recognition, followed by robotic arm movement to this strategic location. In Step 2, the combined CHT and PCA algorithms perform high-precision localization to extract three critical coordinate sets: the set position for each nut (impact driver approach point), the nut position (exact 3D coordinates and orientation), and the wheel position (gripper engagement point). All detected coordinates are transformed and stored within the robotic arm's coordinate frame, enabling the system to save data for each target location during the operation. The robotic arm then executes the sequential tasks accordingly.

1) Loosening the Nuts: The robot arm moves the end effector to the wheel's set position. The integrated impact driver engages, loosening the nut based on the precise nut position and orientation data. This process is repeated for each nut until all are loosened.
2) Wheel Removal: Once all nuts are loosened, the gripper on the end effector activates. Using the calculated wheel position, it securely grips the wheel and carefully removes it from the axle.
3) Standby: The robot arm holds the removed wheel and moves to a designated standby position, waiting for the wheel to be presented for installation.
4) Wheel Installation: The robot arm, with the new wheel in its gripper, returns to the axle's position. It aligns the wheel and places it onto the hub.
5) Tightening Nuts: The robot arm retrieves the nuts, positions them over the bolts, and uses the impact driver to sequentially tighten each nut, ensuring they are securely fastened.

This systematic workflow, guided by the two-stage vision and perception system, ensures a robust and reliable wheel replacement process, from precise nut location to final installation.

IV. Experimental Results

A. Experimental Setup

To thoroughly evaluate the robustness of the proposed two-stage vision system, experiments were conducted under four orientation conditions. In Cases 1–3, roll, pitch, and yaw were independently varied across five discrete angles $(-10°, -5°, 0°, 5°, 10°)$, while the remaining two angles were fixed. Case 4 introduced randomized orientation, with all RPY values uniformly sampled within the range of $(-10°, 10°)$.

At each orientation, a dedicated experimental run was performed by randomly initializing the starting position of the robotic arm within its kinematic workspace and the camera field of view. In addition, for every run, the relative orientation between the end-effector gripper and the wheel was explicitly varied according to the defined conditions, thereby testing the system under diverse alignment scenarios.

Each case consisted of 25 independent trials, resulting in a total of 100 experiments per comparative method (YOLO v11-only vs. YOLO v11 + CHT-PCA). This protocol ensured a comprehensive assessment of pose-invariant performance. Figure 5 illustrates representative setup images, where roll and pitch were fixed while yaw was varied to evaluate orientation robustness.

To additionally verify the performance of the 6-DoF pose estimation algorithm using CHT-PCA, a simulation environment closely resembling the real setup was constructed in NVIDIA Isaac Sim, and the algorithm was validated in simulation accordingly.. The simulated scene instantiated the actual robot model and camera, and the algorithm's estimation accuracy was evaluated by varying XYZ coordinates and RPY orientations consistent with the real experimental protocol.

B. CHT-based nut recognition result

To verify the performance of the CHT algorithm, we used a dataset comprising 100 images of diverse 1/10 scale RC car wheels and evaluated nut recognition results. For each wheel type and image, the detected result was compared to the manually labeled ground truth using the IOU (Intersection over Union) metric. Furthermore, two real RC model wheels were selected and subjected to the same evaluation in actual operational conditions, enabling direct analysis of recognition accuracy in real environments.

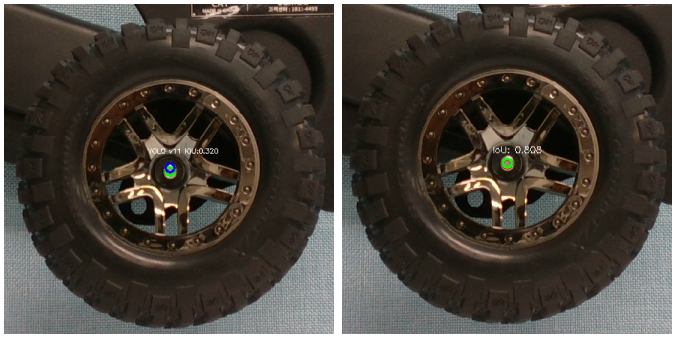

Fig. 6. Visual comparison of recognition results: Initial YOLO v11-Pose detection (left) and the proposed precise CHT-PCA refinement (right)

TABLE I
COMPARISON OF NUT RECOGNITION ACCURACY (IOU)

Method	Mean IOU
YOLO v11-Pose	0.320
CHT-PCA	0.808

As summarized in Table I, the initial detection using YOLO v11-Pose yielded a mean IOU of 0.320. While sufficient for coarse localization, this relatively low score highlights the limitations of deep learning-based regression in capturing the exact boundaries of small, circular objects. In contrast, the proposed CHT-PCA refinement significantly improved the mean IOU to 0.808. This substantial increase demonstrates that integrating geometric circle fitting with learning-based detection is essential for achieving the high-precision localization required for stable tool engagement.

C. 6D-pose estimation Result

To circumvent the difficulty of obtaining precise 6D ground-truth wheel poses in the robot-base frame under real conditions, the accuracy of the 6D pose estimation algorithm was evaluated in NVIDIA Isaac Sim, where ground-truth poses are exactly known and controllable, as illustrated in Figure 7. The wheel's orientation was held fixed while the robot's XYZ and RPY were matched to the real experimental setup, enabling a like-for-like simulation of the acquisition geometry; accuracy was then assessed by comparing the CHT-PCA estimates against the simulator's wheel pose in the robot-base frame. As shown in Table II, per-axis accuracies surpassed 90% on all coordinates, with roll slightly lower at 88.4% due to the limited observability of roll from PCA-derived features. Importantly, within the task scenario considered, small roll-angle errors do not propagate to operation failure provided that the remaining translational and rotational axes are accurate, thereby preserving functional performance; excluding roll, the consistently high accuracy across other axes supports the practical validity of the proposed method for the automated wheel-change task

Fig. 7. Isaac Sim experimental environment (left) and recognition results in simulation (right)

TABLE II
PER-AXIS MEAN ACCURACY(%):XYZ IN MM, RPY IN DEGREES

Approach	X	Y	Z	Roll	Pitch	Yaw
CHT-PCA	92.8	97.3	97.6	88.4	95.4	96.3

D. Task Success Rate

Table III presents the comparative performance results for both approaches across 100 experimental trials.

TABLE III
PERFORMANCE RESULTS

Approach	Success Rate	Avg. Task Time (s)
YOLO v11-Pose	54%	80.2 $\pm$ 8.7
Proposed (PnP + CHT-PCA)	95%	52.8 $\pm$ 6.2

As shown in Table III, the baseline approach utilizing only YOLO v11-Pose with PnP achieved a moderate success rate of 54%. While capable of bringing the end-effector to the vicinity of the wheel, it frequently failed during precision-critical phases, such as exact nut engagement and wheel alignment, due to the inherent jitter and estimation errors of the keypoint detection. In contrast, the proposed integrated system achieved a 95% success rate, representing a significant 41 percentage point improvement in task reliability. Furthermore, the precise alignment provided by the CHT-PCA refinement streamlined the operation, reducing the average task time by approximately 27.4 seconds.

E. Analysis

These results demonstrate that the proposed two-stage recognition process is critical for high-precision manipulation tasks. The analysis highlights a key limitation of relying solely on deep learning-based pose estimation (YOLO v11-Pose): while effective for *coarse* localization and initial approach, it lacks the sub-millimeter precision required for mechanical assembly tasks like nut threading.

By integrating the CHT-PCA algorithm as a refinement stage, the system effectively compensates for the residual errors of the PnP solution. This hierarchical pose refinement strategy combines the robustness of deep learning for global detection with the geometric accuracy of classical computer vision for local alignment. The dramatic increase in success rate (from 54% to 95%) confirms that this hybrid approach

ensures the reliable 6-DoF pose estimation necessary for seamless nut tightening and wheel handling.

While the proposed hierarchical strategy demonstrates high reliability in standard scenarios, specific failure modes were identified under edge conditions. Notably, when the initial angular misalignment between the camera optical axis and the wheel surface exceeded 10 degrees, the performance of the second-stage refinement significantly degraded. In these high-obliquity cases, the circular features of the wheel rim and nuts appear increasingly elliptical due to perspective distortion. Consequently, the Circular Hough Transform (CHT), which is optimized for geometric circles, frequently failed to converge to the correct center coordinates. This estimation error had a cascading effect; the inaccuracy in the wheel center estimation propagated to the nut detection phase, causing misalignment of the impact driver and reducing overall task success. These findings indicate that while the hybrid pipeline improves precision, its robustness relies on the initial alignment remaining within a tolerance threshold, highlighting a critical area for future optimization against large deviations.

V. Conclusion

This paper introduced a robust two-stage vision system for automated wheel replacement, integrating YOLO v11-Pose for coarse approach and CHT–PCA for precise 6-DoF pose refinement. Experimental validation demonstrated that this hybrid pipeline raised the task success rate from 54% to 95% and reduced mean execution time by over 27 seconds compared to a PnP-only baseline. The results confirm that while deep learning effectively handles initial localization, the integration of geometric algorithms is essential for ensuring the high precision required for nut engagement and wheel alignment. Although performance degradation was observed under large initial misalignments ($> 10^{\circ}$), the proposed system proves the efficacy of combining learning-based and geometric perception for industrial mobile manipulation. Future work will focus on enhancing tolerance to large deviations and integrating visual-language–action (VLA) models to further improve task generalization.

Acknowledgment

This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.RS-2020-II201373, Artificial Intelligence Graduate School Program(Hanyang University)) and Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.RS-2025-25443597).

References

[1] W.-L. Mao, Y.-Y. Chiu, B.-H. Lin, C.-C. Wang, Y.-T. Wu, C.-Y. You, and Y.-R. Chien, "Integration of Deep Learning Network and Robot Arm System for Rim Defect Inspection Application," Semantic Scholar, May 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:249005304R.

[2] L. Li, L. Wan, V. Krueger, and X. Zhang, "3D Hand-Eye Calibration for Collaborative Robot Arm," arXiv preprint arXiv:2504.21619, Oct. 2024.

[3] T. P. Nguyen, H. Nguyen, H. Q. T. Ngo, "Development of the Robotic Motion Controller for a Wheeled Manipulator," Measurement and Control, Mar. 2025. doi:10.1177/00202940241260178.

[4] X. Zhang, X. Hu, H. Li, Z. Zhang, H. Chen, "Research on six-joint industrial robotic arm positioning error compensation algorithm based on motion decomposition and improved CIWOA-BP neural network," Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, Jul. 2024. doi:10.1177/09544062241264706.

[5] E. S. Kesuma, P. H. Rusmin and D. A. Maharani, "Pallet Detection and Distance Estimation with YOLO and Fiducial Marker Algorithm in Industrial Forklift Robot," 2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Bali, Indonesia, 2023, pp. 270-275, doi: 10.1109/ICAIIC57133.2023.10066999

[6] Y. Lin, C. Wu, R. Wang, K. Chen, J. J. Lee, and D. Xu, "YOLO-6D+: Single Shot 6D Pose Estimation Using Privileged Silhouette Information," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, Jun. 2022, pp. 338-347. doi:10.1109/CVPR52688.2022.00039

[7] B. Tekin, S. N. Sinha, and P. Fua, "Real-Time Seamless Single Shot 6D Object Pose Prediction," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, Jun. 2018, pp. 292-301. doi:10.1109/CVPR.2018.00038

[8] C. Wang, D. Xu, Y. Zhu, R. Martín-Martín, C. Lu, L. Fei-Fei, and S. Savarese, "DenseFusion: 6D Object Pose Estimation by Iterative Dense Fusion," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, Jun. 2019, pp. 3343-3352. doi:10.1109/CVPR.2019.00345

[9] K. Park, T. Patten, and M. Vincze, "Pix2Pose: Pixel-Wise Coordinate Regression of Objects for 6D Pose Estimation," Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, Oct. 2019, pp. 7668-7677. doi:10.1109/ICCV.2019.00778

[10] K. Samarawickrama, G. Sharma, A. Angleraud, and R. Pieters, "6D Assembly Pose Estimation by Point Cloud Registration for Robot Manipulation," arXiv preprint arXiv:2312.02593, Dec. 2023. doi:10.48550/arXiv.2312.02593.

[11] C.-J. Lin, P.-J. Lin, and C.-H. Shih, "Vision-based Robotic Arm Control for Screwdriver Bit Placement Tasks," Sensors and Materials, vol. 36, no. 3, pp. 1003–1018, Mar. 2024. doi:10.18494/SAM4739.

[12] G. G. Oladipupo, "Development of a Robotic System for Automatic Wheel Removal and Fitting," *IEEE SEM*, vol. 10, no. 5, pp. 21-28, May 2022.

[13] G. Lee, J.-S. Kim, S. Kim, and K. Kim, "6D Object Pose Estimation Using a Particle Filter With Better Initialization," *IEEE Access*, vol. 11, pp. 11451–11462, 2023. doi:10.1109/ACCESS.2023.3241250.

[14] Y. Cohen, A. Biton, and S. Shoval, "Fusion of Computer Vision and AI in Collaborative Robotics: A Review and Future Prospects," Applied Sciences, vol. 15, no. 14, article 7905, Jul. 2025. doi:10.3390/app15147905.

[15] R. Khanam and M. Hussain, "YOLOv11: An Overview of the Key Architectural Enhancements," arXiv preprint arXiv:2410.17725, Oct. 2024.

[16] L. Duarte, P. Neto, and J. Norberto Pires, "Applications and design methodology of mobile robotic manipulators: A review," Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, Aug. 2024. doi:10.1177/09544062251347212.

[17] N. Núñez-Calvo, A. J. García-Cerezo, and J. A. Gómez-de-Gabriel, "Enhancing accuracy in Mobile Manipulators: Challenges and solutions," Robotics and Computer-Integrated Manufacturing, 2025. In press.

[18] H. Zhang, S. Choi, and J. Bohg, "Mobile Manipulation in Unexplored and Unstructured Environments," arXiv preprint, 2024. [Online]. Available: https://openreview.net/pdf?id=27NXYoh9kJ

[19] D. G. Psychogios, G. E. Beligiannis, and N. A. Panagiotopoulos, "Model-Based Digital Twin for Collaborative Robots," In Proceedings of the 11th International Conference on Model-Driven Engineering and Software Development (MODELSWARD), 2024, pp. 199–210

[20] O. V. S. Runola and J. Latokartano, "Opportunities and Challenges of Collaborative Robotics in Human-Centric Automotive Assembly," University of Vaasa, 2024. [Online]. Available: osuva.uwasa.fi

[21] S. Chen, P. Abbeel, and D. Pathak, "A Survey of Robotic Navigation and Manipulation with Embodied AI," arXiv preprint, 2025.