System for Autonomous Management of Retail Shelves using an Omnidirectional Dual-arm Robot with a Novel Soft Gripper

Nijil George*[1], Somdeb Saha[1], Shubham Parab[1], Vismay Vakharia[1]
Rolif Lima[1], Vighnesh Vatsal[1] and Kaushik Das[1]

***Abstract*— Managing shelves in retail stores includes restocking, rearrangement and replenishment of products. As these are some of the most labor-intensive activities, there has been widespread demand from retailers for automation in this domain. However, major challenges still remain in perception, navigation and manipulation while implementing an autonomous robotic system for this purpose. We present a system aimed at addressing some of these challenges through novel approaches. In terms of perception, we have developed a transformer-based local anomaly detection algorithm that can identify misplaced items without the need for a central database. Navigation of the omnidirectional mobile base is performed through stereo vision and LiDAR sensors. Finally, identifying grasping and manipulation as one of the key shortcomings of present robotic systems in this domain, we have developed a customized soft robotic gripper targeted at retail objects. It has compliant cable-driven fingers, and a palm configuration that can be adapted in real-time based on the target object's geometry. Coupled with a conventional two-fingered gripper in a dual-arm setup, this system is equipped to handle most objects encountered in a retail setting. We describe the underlying hardware and algorithms for each component of the system, evaluating their individual performance. We then evaluate the whole system in a mock retail setup, demonstrating promising results for autonomous management of shelves.**

I. Introduction

In recent years, there has been a notable surge in demand for automation in retail store shelf management. This demand stems from the need to alleviate bottlenecks in the otherwise streamlined retail logistics operations, that have seen significant advancements in automation for warehousing and transportation. This has been coupled with labor shortages due to ageing demographics in industrialized nations and the effects of the COVID-19 pandemic [1]. Therefore in the near future, retail automation would primarily be focused on labor-intensive activities within stores, such as item replenishment, item rearrangement, and addressing toppled items. These tasks have predominantly relied on manual intervention due to the intricate demands of perception, manipulation, and navigation required for autonomous system operation [2].

Numerous attempts have been made at automating retail stores, typically involving only perception and navigation, using a mobile robot with a camera [3] moving through a store to track inventory and identify objects that require human intervention [4], [5]. Some solutions also include robotic arms [6], though they have focused on remote workers, where the same tasks as done by a worker would be performed

[1]The authors are with TCS Research, Tata Consultancy Services Ltd., Bengaluru, Karnataka - 560066, India. *Corresponding Author: george.nijil@tcs.com

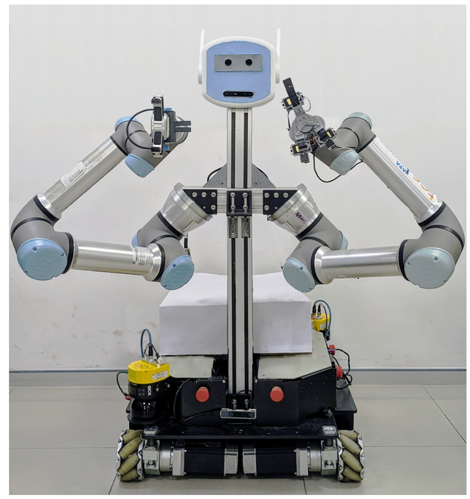

Fig. 1. GriffinX: An in-house built omnidirectional mobile robot with dual-arm collaborative manipulators

using a robot as a tele-existence interface [7]. While these works promise to alleviate some of the labor bottlenecks in retail, they are ultimately still reliant on repeated human intervention. A comprehensive, fully autonomous solution is not currently available on the market.

Our proposed system aims to tackle multiple challenges encountered in retail store automation, thereby enabling humans to focus on more creative and cognitively demanding tasks. In this study, we deploy the system to address the detection and removal of misplaced items, involving separate modules for perception, manipulation and navigation, designing the system to suit the unique demands of retail store operations. Perception proved to be particularly challenging, as conventional methods for identifying objects that need intervention in retail stores rely on product recognition [8] and planogram compliance [9]. Supervised deep learning methods used for this purpose struggle due to the dynamic nature of item designs and rapidly evolving inventory [10]. To overcome this, we employ a fully unsupervised method capable of detecting misplaced items through local contextual cues.

In terms of manipulation, given the diverse array of items encountered in retail stores, we opt for a dual-arm setup equipped with two distinct types of grippers. Existing solutions have relied on learning dexterous behaviors for conventional parallel-jawed rigid grippers [11], or have involved custom-designed grippers based on vacuum-suction [12] and soft materials [13]. We introduce a novel soft gripper design that allows for handling large variety of object types. It consists of three cable-driven soft fingers, each actuated independently. The configuration of the fingers in its palm is adaptable in real-time based on the shape of the target object. With a conventional parallel-jawed gripper on one arm to handle rigid prismatic objects, having the novel soft gripper on the other arm allows for the handling of irregular and fragile objects, thus covering a wide spectrum of target items using a single setup.

For navigation purposes, we introduce an omnidirectional mobile base that was designed and fabricated in-house, facilitating seamless movement within a retail environment.

A. Contributions and Overview

The system consists of an omnidirectional mobile base with two manipulators mounted on top. Each manipulator has a different type of gripper as end-effector, enabling the system to handle a large spectrum of objects which are encountered in a retail store. It has an RGB-D camera for perception, the images taken by the camera are used by the perception algorithms to detect and localize anomalies present in the retail store shelves. Fig. 1 depicts our prototype and Fig. 5 shows the block diagram of the system.

The key contributions of this paper are:

- A novel reconfigurable soft robotic gripper design adaptable to a large variety of retail items
- A modular omnidirectional mobile base tailored for retail stores
- System integration of multiple methods from our previous works and from the literature

II. System Description

A. Mobile Robot with Dual Arms

Retail store environments commonly have flat, smooth floors with narrow aisles. Considering these requirements, an in-house designed and fabricated omni-directional platform is used as the mobile base. It provides agility, efficiency, and reliability in autonomous navigation within retail environments. The chassis is constructed from durable materials like aluminum composite, housing integrated compartments for electronic components, batteries, and sensors. A set of Mecanum wheels [14] are coupled with high-torque BLDC motors providing precise motion control of the base platform. It enables movement in forward, backward, diagonal, and lateral directions, as well as in-place rotation, thereby offering exceptional maneuverability. This agility is particularly advantageous in retail store environments, where space between aisles is often constrained. The mobile base is equipped with an NVIDIA Jetson Xavier TX2 as an onboard computer to perform critical control and navigation tasks by using a suite of sensors. For obstacle avoidance and navigation, LiDAR, Real-Sense depth camera D415, and T265 self-tracking cameras are used. The speed of the robotic base is engineered to optimize both efficiency and safety; it is quick enough to navigate from one position to another while ensuring the safety of customers and employees.

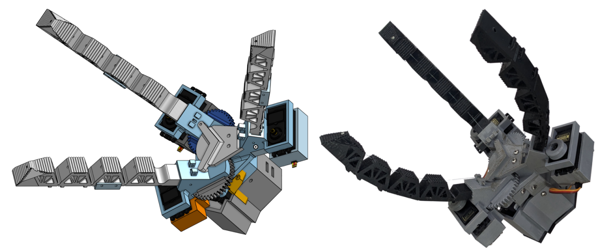

Fig. 2. CAD model and fabricated reconfigurable three-fingered soft robotic gripper

For manipulation in the retail store, two Universal Robots UR5e Cobots (collaborative robots), each offering 6 degrees-of-freedom (DoF) are mounted on the mobile platform. The manipulators are mounted in a symmetrical configuration at 45 degrees relative to the base. One manipulator is equipped with an OnRobot RG2 gripper, designed for grasping a variety of standard objects while the other one has a custom-built soft gripper mounted for grasping delicate arbitrarily shaped items.

For efficient motion planning of the mobile base platform employs ROS Navigation framework for traversing from one point to another. Whereas, the motion planning of the manipulators uses a Model Predictive Controller in Cartesian space from our previous work [15]. The trajectory obtained from the MPC controller is translated to joint space using the open-source TracIK [16] library to solve the inverse kinematics problem. This ensures that the arms reach their desired end-effector poses along a planned path. The combination and configuration of sensors enable the robot to perceive its surroundings, make decisions, and carry out tasks with minimal human intervention.

B. Perception

The perception subsystems consists of an RGB-D camera, a shelf-detection module, an anomaly detection module and an object localization module. We have used an Intel D415 RealSense Depth Camera for capturing the RGB-D video of the scene. Shelf-detection is done using ArUco markers placed along the boundary of the shelf racks. The anomaly detection module is adapted from [17] by training it on the RP2K retail dataset [18]. The object localization module uses the output from the anomaly detection module and the depth information from the RGB-D image to determine the location of the anomalous object in the 3D robot coordinate frame. This 3D coordinate is used by the robotic manipulator to grasp the target anomalous object.

C. End-Effector

The end-effector plays a pivotal role in the efficacy of a manipulation system, significantly influencing its overall success rate. To enhance the system's versatility in handling a diverse range of objects, we integrate two distinct types of grippers as end-effectors for the two manipulators: a two-fingered rigid gripper and a novel three-fingered soft gripper. While the OnRobot RG2 gripper serves as the rigid gripper,

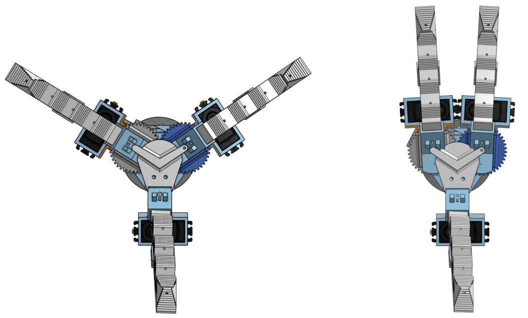

Fig. 3. Gripper configuration for spherical (left) and cylindrical (right) grasping

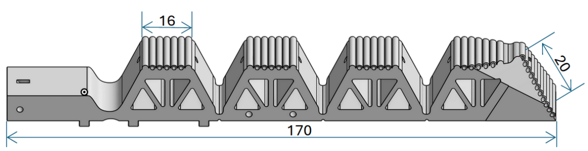

Fig. 4. Soft finger design (dimensions in mm)

its efficacy is limited in grasping the varied objects encountered in retail environments. To address this, we develop a bespoke three-fingered soft gripper capable of dynamic reconfiguration, depicted in Fig. 3. This reconfigurability enables the gripper to seamlessly transition between spherical and cylindrical grasps, enhancing its adaptability to a wide array of objects commonly found in retail settings.

The soft gripper's pliable nature allows it to securely grasp delicate or compliant objects, such as farm produce and chip packets, without causing damage—items that are typically difficult for rigid grippers to grasp. Fabricated using a 3D printer, the gripper comprises a rigid palm made of Polylactic Acid (PLA) and three soft fingers crafted from Thermoplastic Polyurethane (TPU-95A). Each finger is individually actuated via cables connected to their respective servo motor, with one finger fixed at the base and the others movable relative to the palm. The reconfiguration is enabled by a fourth servo motor. Modular in design, the gripper allows for easy replacement of worn-out fingers and customization with alternative cable actuated designs, ensuring adaptability to diverse object types. Fig. 2 shows the CAD model and the fabricated soft gripper.

Fig. 4 illustrates the finger design, featuring four segments with flexible portions facilitating deformation upon cable actuation. The cable is tethered through the holes provided in the segments with a mounting hole at the tip of the finger. The tip of the finger is made slanted to allow the fingers to be used for pinch grasping. The dimensions mentioned in the figure were deduced by considering small to medium sized items normally found in retail stores. This design can be customized for objects of different sizes by scaling the fingers appropriately.

III. Methods

A. Anomaly Detection

We have extended our prior research [17] to address anomaly detection within racks. In real-world scenarios, multiple racks may appear within a single image frame, necessitating the use of a shelf-rack detection framework prior to the anomaly detection module to isolate rack regions. To achieve this, we employ ArUco markers [19] with sequential ID numbers for individual shelf rack identification. Leveraging the ArUco marker detector in the OpenCV library [20], we detect these markers; however, we encountered intermittent marker detection issues in video frames. To mitigate this, we have integrated a Kalman filter [21], [22] to stabilize detected ArUco marker positions.

Within each shelf-rack, objects are detected using an object detector, and the resulting bounding boxes are used to crop the image into smaller windows, each containing a single object. These cropped images are then fed into the anomaly detection module to identify anomalous objects based on feature variations. The anomaly detection module outputs bounding boxes of the anomalous objects in each rack, from which we compute centroid positions in the image coordinate frame. Subsequently, we determine depth information corresponding to these centroid positions in the RGB-D image.

We conduct camera calibration [23] to obtain the camera intrinsic and extrinsic matrices with respect to the robot coordinate frame. These matrices are used to compute the 3D coordinates of target centroids in the robot coordinate frame.

B. Autonomous Navigation

The mobile robot base is equipped with four Mecanum wheels. This design facilitates omnidirectional movement without the need for a conventional steering system. It operates as a holonomic mechanism, capable of traveling in any direction regardless of its orientation [24]. Eq. 1 represents the mathematical model of inverse kinematics of the Mecanum-drive mobile base to obtain angular velocities of the wheels (ω_i) from the desired body velocities: linear (v_x, v_y) and angular ω_z.

$$\begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \\ \omega_4 \end{bmatrix} = \frac{1}{R_w} \begin{bmatrix} 1 & -1 & -(l_x + l_y) \\ 1 & 1 & (l_x + l_y) \\ 1 & 1 & -(l_x + l_y) \\ 1 & -1 & (l_x + l_y) \end{bmatrix} \begin{bmatrix} v_x \\ v_y \\ \omega_z \end{bmatrix} \tag{1}$$

The desired body velocities for the mobile base are given by the navigation module. The navigation module consists of four sub-modules: odometry, localization, mapping and path planning. The measurements from the IMU and tracking camera are integrated with data from the wheel encoders to achieve reliable and accurate odometry. The LiDAR scan along with the generated odometry is utilized to build a map capturing the spatial layout and landmarks within the retail environment. A particle filter based localization technique, Adaptive Monte Carlo Localization (AMCL) [22]

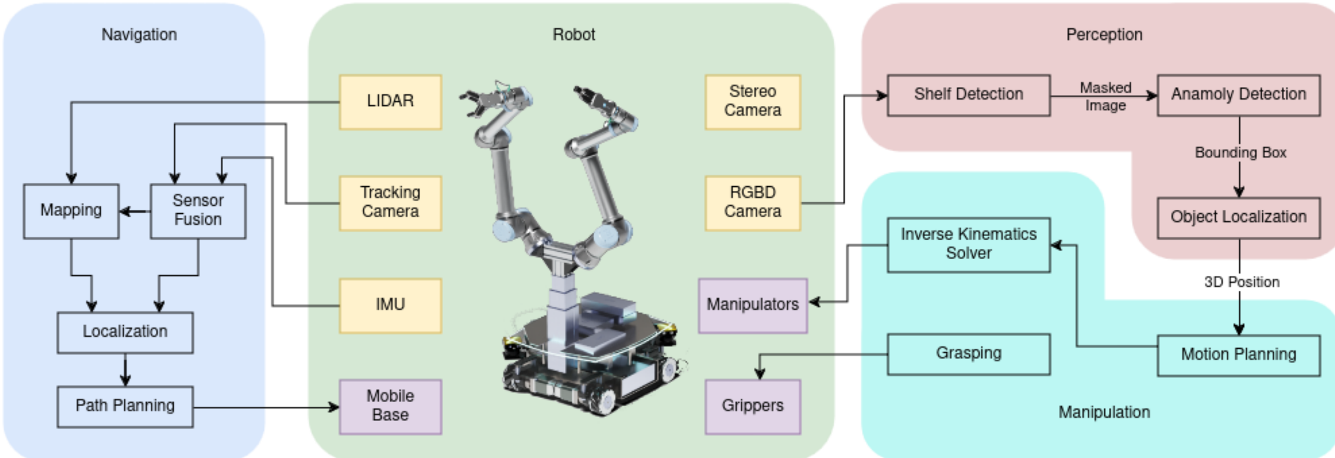

Fig. 5. Overall system architecture

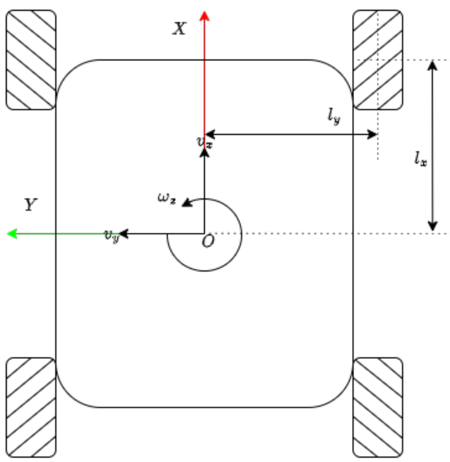

Fig. 6. Schematic diagram of 4-Mecanum wheel omnidirectional drive

is employed to accurately estimate the mobile base's position and orientation within the map. In order to navigate from one position to another, A* based holonomic path planning is performed on the map. Local planning algorithm, Dynamic Window Approach generates commands for the mobile base to execute while avoiding dynamic obstacles. All the modules use the ROS Navigation (Nav2) implementation of algorithms.

C. Manipulation

1) Motion planning of manipulator arms: The 3D coordinates of the target objects, obtained from the anomaly detection module is taken as the goal position for the manipulator and a Model predictive contoller (MPC) is used to plan the path of the end-effector to move to the goal position. MPC controller is used due to it's ability to provide optimal trajectories while simultaneously satisfying constraints corresponding to collision with environment, and with robot's body [15]. MPC formulation used for the robot control is as given below:
Cost function: A standard quadratic cost function is used for driving the state $\mathbf{x}$ corresponding to the desired goal position $\mathbf{x_d}$ while also minimixing the velocity $\mathbf{u}$ as given below

$$J(\mathbf{x},\mathbf{u}) = \min_{\mathbf{x},\mathbf{u}} \sum_{i=0}^{N} (\mathbf{x_i} - \mathbf{x_d})^T \mathbf{Q} (\mathbf{x_i} - \mathbf{x_d}) + \mathbf{u_i}^T \mathbf{R} \mathbf{u_i} \quad (2)$$

where, $\mathbf{x}_i$ and $\mathbf{u_i}$ are the i^{th} state and control element in the trajectory, $\mathbf{Q}$ and $\mathbf{R}$ are positive semi-definite and positive definite matrices respectively used to weight the state and control cost.
System Model: A point mass model is used to model the motion of each end-effector independently, this is not a hard requirement and even a double integrator can be used, since the trajectory obtained by solving the resulting optimal control problem using MPC methodology is used as a reference for the inner-loop high gain controller, which render the inner loop system as identity system to outer-loop controller. Thus the propagation of the state $\mathbf{x}$ at each discrete time step t is given as

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \mathbf{u}_t \delta \quad (3)$$

where, $\delta = T/N$, with T being the time horizon and N being the number of steps in the horizon.
Constraints: The end effector's position is constrained to prevent collisions with both the robot body and items on the shelf. This is achieved through collision avoidance constraints, implemented using planners and ellipsoidal constraints. [15]

A planner constraint is defined for preventing collision with objects such as end of the shelf, floor and is defined as

$$\mathbf{A}\mathbf{x} \geq \mathbf{b} \quad (4)$$

where, the rows of matrix $\mathbf{A}$ comprise the unit normal to the plane that we do not want end-effector to cross, and the elements of $\mathbf{b}$ correspond to the calibrated limits. Example, to prevent the end-effector from colliding with the floor, one of the rows of the $\mathbf{A}$ will be chosen as $[0, 0, 1]$ and the corresponding value of $\mathbf{b}$ will be chosen as the height of the floor from the base of the robot.

Similarly, to avoid collision with smaller objects, an ellipsoidal constraint is used to prevent the end-effector from entering the enclosed ellipsoid volume around objects in the workspace.

$$(\mathbf{x} - \mathbf{x}_o)^T \mathbf{RMR}^T (\mathbf{x} - \mathbf{x}_o) > \alpha \tag{5}$$

where, $\mathbf{x}_o$ is the centroid of the ellipsoid, $\mathbf{R}$ is the rotation matrix defining the orientation of the ellipsoid. $\mathbf{M}$ is a diagonal matrix and α is a scalar value, chosen empirically to enclose the object in the ellipsoid.

In addition to this, the velocity of the end-effector is also limited with in the fixed range using the bounding constraint as below

$$\mathbf{u_{min}} \leq \mathbf{u} \leq \mathbf{u_{max}} \tag{6}$$

Upon solving the optimal control problem, a collision-free trajectory for the end-effector is obtained within the specified time horizon. The first point of this trajectory is then passed to the inverse kinematics solver, which generates eight possible sets of joint angle solutions for each arm. The solution closest to the robot's current joint configuration is selected. Additionally, the robot must adopt a desirable posture, ensuring that the elbow joint is consistently oriented away from the robot's body. To achieve this, weighted least squares (7) is employed to select the most suitable joint positions from the set of solutions provided by the inverse kinematics solver.

$$\theta = arg \min_{\theta} \sum_{i}^{N} \; ||\theta - \theta_l||^2_{\mathbf{W}_l} + ||\theta - \theta_d||^2_{\mathbf{W}_d} \tag{7}$$

where, $|| \cdot ||_{\mathbf{W}_i}$ represents weighted norm of difference in the IK solution θ, θ_l and θ_d represent the current robots joint position and desired joint configuration respectively with weight matrix $\mathbf{W}_l$ and $\mathbf{W}_d$ respective errors.

2) Grasping: For the two-fingered rigid gripper, the grasp pose required for successfully grasping the target object is estimated using grasp pose generator [25]. For the three fingered soft gripper, we have adapted our previous work [26] for grasp pose generation. Because the efficacy of the grasp is dependent on the design of the soft-fingers it is essential to model the finger each time the design is changed, we have followed the method mentioned in [27].

IV. Results

We tested the individual subsystems, as well as the system as a whole in a mock retail environment. The perception subsystem's performance mirrored the performance which was obtained in [17]. Fig. 7 shows the output of the perception module to an image from a mock retail setup.

The navigation subsystem was able to properly localize the mobile base to be within ± 5 cm error using the onboard sensors when compared to the ground truth provided by a motion capture system.

The soft gripper was extensively tested for its efficacy to grasp retail items. We have included objects of different shapes and sizes, trying to imitate the variety of objects normally encountered in a retail setting, including fruits, vegetables, plastic bottles, tin cans, cuboidal tetra-packs and chips packets. The soft gripper was able to grasp all of the items when used in a top-down approach provided the dimension of the object along the grasp closure direction of the fingers was less than the width of the gripper in open position (18 cm). The compliant nature of the fingers allowed the gripper to conform itself to the shape of the object being grasped and it either led to a pinch grasp or a form closure grasp depending on the size of the object. Sideways grasps proved to be much more challenging because of the inability of the gripper to attain form closure when the objects are too small for the dimension of the soft-fingers. We estimated that for an object with a dimension less than 7 cm along the grasp closure direction, the gripper is not able to attain a successful grasp. However, this limitation can be overcome by scaling the design of the soft-fingers accordingly. Fig. 8 and Fig. 9 depict the successful grasps of common retail items from top-down and sideways approaches respectively.

Fig. 7. Shelf detection module masking out the part of the image outside the rack (blue tint). Anomaly detection module localizing the anomaly within a rack (red bounding box)

The system was tested in a mock retail setup, which is a simplified version of a real-world retail store, where items are kept sufficiently apart for ease of grasping. Additionally, ArUco markers are pasted along the periphery of each shelf rack, allowing for masking of the non-rack parts of the input image. The schematic model of the same is illustrated in the Fig. 10 and Fig. 11 shows the system interacting with the setup.

The map of the above setup is generated according to the method described in Sec. III-B, shown in Fig. 12. In the figure, the sequence and the path followed by the robot is pointed out using numbers and red-dotted line respectively. Here, the robot starts from position 1, which is considered as a docking area, moves towards position 2, where the first rack is present. The perception sub-system takes in the image from the RGBD camera, performs shelf detection and returns the masked image. This image is passed to the anomaly detection module which returns the 3D position of the anomalous object in the robot's local coordinate frame. This 3D position is used by the manipulation subsystem to generate the commands for the manipulators and the

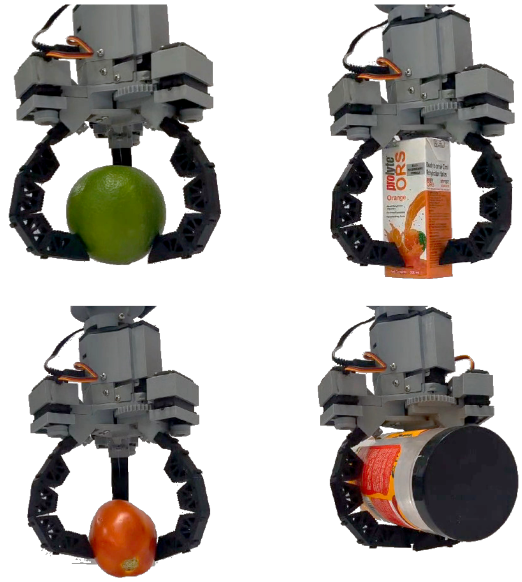

Fig. 8. Top-Down grasp of retail objects using soft gripper in spherical configuration (left-top, right-top, left-bottom) and cylindrical configuration (right-bottom)

Fig. 9. Sideways grasp of retail objects using soft gripper in cylindrical configuration (left-top, right-top, left-bottom) and spherical configuration (right-bottom)

grippers. Once the anomalous object is grasped, the mobile base moves towards position 3, where the second rack will be present. The system then follows the same procedure as before and grasps the second object with the second gripper. Then the omnidirectional mobile base navigates the diagonal path between positions 3 and 4, once it reaches position 4 it rotates about its own axis and finally drops the items on the table at the drop location. The system was able to complete the task autonomously.

V. Conclusion

In this work, we have presented an early prototype of a complete system for autonomous management of shelves in supermarkets and other kinds of retail settings. As a platform,

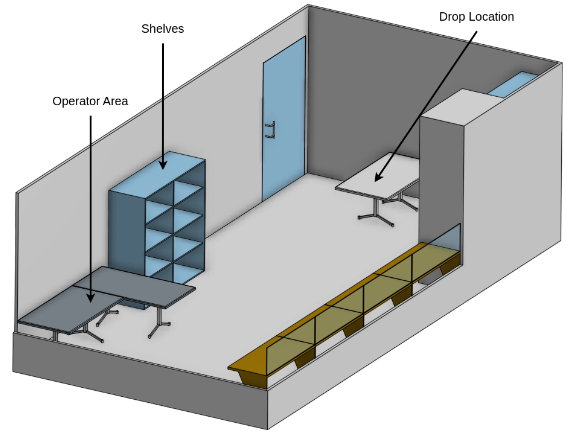

Fig. 10. Mock retail setup schematic

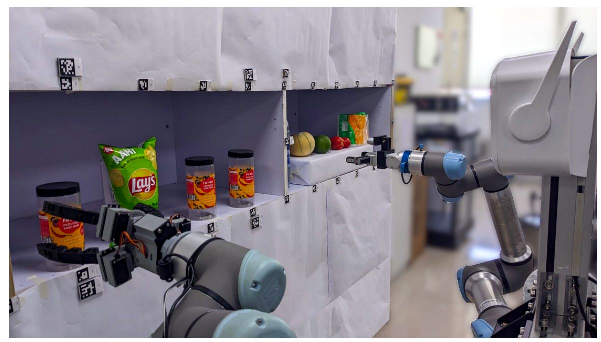

Fig. 11. The mobile manipulator robot operating in the mock retail setup

it consists of an omnidirectional base, dual manipulator arms, and a novel grasping solution in the form of a soft robotic gripper. Describing each module, we have demonstrated how such a system can be used for alleviating labor constraints in retail stores through automation. However, these early results are in a controlled laboratory environment. In order for such a system to be commercially successful, each module needs to be scaled up and made more robust, particularly in terms of the software components.

For instance, the perception module needs to be able to segment individual shelves, and segregate items within those shelves without the use of external markers, even when items are placed close together. Our ongoing efforts include exploring methods for more fine-grained object segmentation and anomaly detection in noisy visual frames.

The grasping module shows promising results in handling objects of various shapes, sizes and levels of fragility. Ongoing work for grasping primarily involves the decision-making process for selecting between the soft gripper and the rigid two-fingered gripper. Also, if the soft gripper is selected, another decision must be made on the configuration of the fingers in the palm based on the object's shape. We are exploring the use of reinforcement learning to perform this dynamic finger reconfiguration along with grasp planning. Haptic sensing will also be integrated in the soft fingers to perform closed-loop grasping, robust against slippage.

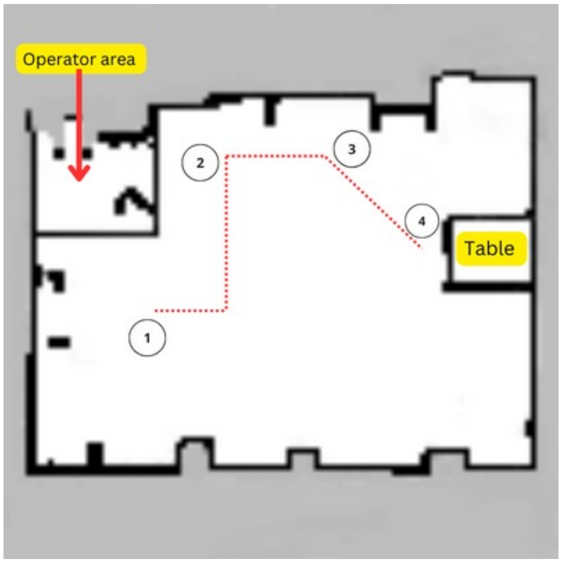

Fig. 12. Map of the mock retail setup

Finally, grasping individual items from packed shelves is an ongoing challenge, also being explored through reinforcement learning involving dual-arm synergies.

In terms of planning, the operation of the robot in this work triggers each module with execution happening in a sequential manner. A real retail setting will involve highly complex task and motion planning for an individual robot, and multi-robot coordination if a group of robots were to be deployed in a store, along with scheduling and fleet management. These challenges constitute future work for such systems.

While the system presented in this work is not currently mature enough to be deployed in real stores, it shows how the necessary hardware and software components can be integrated for this task, and identifies the enhancements necessary in each module to achieve the ultimate goal of retail store automation.

REFERENCES

[1] M. Horrigan, M. Heggeness, K. Bahn, and M. R. Strain, "Is there a labor shortage?" *Business Economics*, vol. 57, no. 1, pp. 6–22, 2022.

[2] V. Sengar, A. Kapoor, N. George, V. Vatsal, J. Gubbi, A. Pal, *et al.*, "Challenges in applying robotics to retail store management," *arXiv preprint arXiv:2208.09020*, 2022.

[3] J. Gee, B. Bogolea, and M. A. Shah, "Inventory-tracking robotic system," U.S. Patent USD819 712S1, May 05, 2018.

[4] W. Knight, "Robot makes sure stores don't run out of doritos," MIT Technology Review, 2015. [Online]. Available: https://www.technologyreview.com/2015/11/11/165116/robot-makes-sure-stores-dont-run-out-of-doritos/

[5] M. McCarthy, "Chilean robot-as-a-service company zippedi raises 6.9 million to digitize stores," Bloomberg Linea, 2021. [Online]. Available: https://www.bloomberglinea.com/english/chilean-robot-as-a-service-company-zippedi-raises-69-million-to-digitize-stores/

[6] B. Pistrui, D. Kostyal, and Z. Matyusz, "Dynamic acceleration: Service robots in retail," *Cogent Business & Management*, vol. 10, no. 3, p. 2289204, 2023.

[7] T. Kelly, "Japanese robot to clock in at a convenience store in test of retail automation," Reuters, 2020. [Online]. Available: https://www.reuters.com/article/idUSKCN24G11Q/

[8] A. Tonioni and L. D. Stefano, "Product recognition in store shelves as a sub-graph isomorphism problem," in *International Conference on Image Analysis and Processing*. Springer, 2017, pp. 682–693.

[9] E. Frontoni, M. Contigiani, and G. Ribighini, "A heuristic approach to evaluate occurrences of products for the planogram maintenance," in *2014 IEEE/ASME 10th International Conference on Mechatronic and Embedded Systems and Applications (MESA)*. IEEE, 2014, pp. 1–6.

[10] Y. Wei, S. Tran, S. Xu, B. Kang, and M. Springer, "Deep learning for retail product recognition: Challenges and techniques," *Computational intelligence and neuroscience*, vol. 2020, 2020.

[11] K. Kleeberger, R. Bormann, W. Kraus, and M. F. Huber, "A survey on learning-based robotic grasping," *Current Robotics Reports*, vol. 1, pp. 239–249, 2020.

[12] R. Sakai, S. Katsumata, T. Miki, T. Yano, W. Wei, Y. Okadome, N. Chihara, N. Kimura, Y. Nakai, I. Matsuo, *et al.*, "A mobile dual-arm manipulation robot system for stocking and disposing of items in a convenience store by using universal vacuum grippers for grasping items," *Advanced Robotics*, vol. 34, no. 3-4, pp. 219–234, 2020.

[13] N. Kuppuswamy, A. Alspach, A. Uttamchandani, S. Creasey, T. Ikeda, and R. Tedrake, "Soft-bubble grippers for robust and perceptive manipulation," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 9917–9924.

[14] B. E. Ilon, "Rad fuer ein laufstabiles, selbstfahrendes fahrzeug," Germany Patent DE2 354 404A1, 1974, German Patent No. DE2354404A1.

[15] R. Lima, V. Vakharia, U. Rai, H. Mehta, V. Vatsal, and K. Das, "Model-mediated delay compensation with goal prediction for robot teleoperation over internet," in *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2023, pp. 1037–1043.

[16] P. Beeson and B. Ames, "Trac-ik: An open-source library for improved solving of generic inverse kinematics," in *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2015, pp. 928–935.

[17] A. Kapoor, V. Sengar, N. George, V. Vatsal, J. Gubbi, B. P, and A. Pal, "Concept-based anomaly detection in retail stores for automatic correction using mobile robots," in *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2023, pp. 163–170.

[18] J. Peng, C. Xiao, and Y. Li, "Rp2k: A large-scale retail product dataset for fine-grained image classification," *arXiv preprint arXiv:2006.12634*, 2020.

[19] 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.

[20] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[21] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 03 1960. [Online]. Available: https://doi.org/10.1115/1.3662552

[22] S. Thrun, "Probabilistic robotics," *Communications of the ACM*, vol. 45, no. 3, pp. 52–57, 2002.

[23] D. A. Forsyth and J. Ponce, *Computer vision: a modern approach*. prentice hall professional technical reference, 2002.

[24] H. Taheri, B. Qiao, and N. Ghaeminezhad, "Kinematic model of a four mecanum wheeled mobile robot," *International journal of computer applications*, vol. 113, no. 3, pp. 6–9, 2015.

[25] D. Morrison, P. Corke, and J. Leitner, "Closing the loop for robotic grasping: A real-time, generative grasp synthesis approach," *arXiv preprint arXiv:1804.05172*, 2018.

[26] V. Vatsal and N. George, "Augmenting vision-based grasp plans for soft robotic grippers using reinforcement learning," in *2022 IEEE 18th International Conference on Automation Science and Engineering (CASE)*, 2022, pp. 1904–1909.

[27] N. George and V. Vatsal, "Modeling of soft robotic grippers for reinforcement learning-based grasp planning in simulation," in *2023 Ninth Indian Control Conference (ICC)*, 2023, pp. 287–292.