IEEE Robotics & Automation Magazine (RAM) paper, presented at ICRA 2026, Vienna, Austria. Cite as RAM paper.

Serving Innovation: Seamless Service by Advancing Food Runners with Mobile Manipulation

Sankalp Yamsani[1], Kevin Gim[2], Tyler Smithline[2], Richard Qiu[2], Roman Mineyev[3], Kenta Hirashima[2], Sungmin Kang[2], Kyungseo Park[4], Yoon-Koo Kang[5], Seulbi An[5], Sunghwan Ahn[5], Joohyung Kim[1]

***Abstract*— The Mobile Object Manipulation Operator (MOMO) is an innovative and reconfigurable robotic system that transforms traditional serving robots into mobile manipulators. Leveraging the form factor and mobility of serving robots, MOMO integrates up to three pluggable devices, including six-DoF manipulators of varying sizes or a three-DoF sensor head. Its design incorporates two independent shoulder lifts to enhance vertical reach. The adaptability of the system tailors its capabilities to tasks beyond simple object transportation. As opposed to current food delivery robots, MOMO showcases its ability to remove obstructions from the floor and deliver items to recipients without human intervention.**

I. Introduction

RISING labor costs, the lingering effects of the COVID-19 pandemic, and staffing shortages have accelerated the adoption of restaurant delivery robots—a trend expected to continue over the next decade. Industry analysts project that the market for such autonomous systems will reach at least $72 billion by the end of 2032, an increase of 22% from the levels of 2023 [1]. One robot system that has gained traction in this landscape is the food runner, designed to autonomously transport orders from the kitchen to customers.

However, most current food runners are limited in functionality: they lack manipulation capabilities and still require human intervention at key points —retrieval and placement— of the delivery process. To address this gap in automation, we present the Mobile Object Manipulation Operator (MOMO)–a highly modular, reconfigurable robotic platform designed to extend the functionality of traditional food service robots by adding manipulation, perception, and sensing capabilities. Unlike most existing delivery robots, MOMO introduces a task-driven hardware modularity, allowing the same base to be rapidly reconfigured for diverse service tasks within seconds.

Built on a mobile base traditionally used in food serving robots, MOMO features a novel torso equipped with three independent plug-and-play ports. These ports support seamless integration of robotic modules, such as a six-degree-of-freedom (DoF) manipulator [2] to extend manipulation and workspace capabilities, or a three-DoF sensor module [3] to enhance perception. Two of these ports are mounted on decoupled, motorized shoulder lifts, providing an additional DoF that significantly expands vertical reach while preserving the form factor of a traditional food runner. This design uniquely enables MOMO to transition from passive transport to complex manipulation tasks. In Fig. 1 we show a few of the configurations that can be set up within seconds.

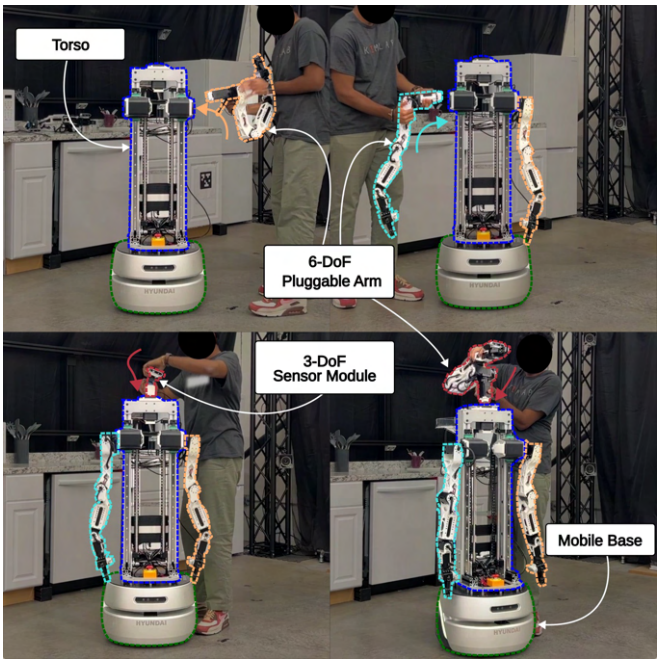

Fig. 1: Example configurations of the MOMO utilizing six-DoF manipulators and a three-DoF sensor module.

This article presents a comprehensive overview of MOMO, a modular mobile manipulator designed for flexible, task-driven reconfigurability through a plug-and-play hardware architecture. The hardware and software architecture is presented in sections III and IV. Section V showcases MOMO's real-world applications, demonstrating how its modular design extends the capabilities of existing food runner robots via expansive workspace and simultaneous manipulation by multiple robotic arms.

II. Related Mobile Manipulators

Researchers have been developing mobile manipulators to achieve generality in recent years. These systems offer mobility and manipulability, increasing their adaptability in

[1] Department of Electrical and Computer Engineering, University of Illinois Urbana-Champaign, Champaign, IL, USA. {yamsani2, joohyung}@illinois.edu
[2] Department of Mechanical Science & Engineering, University of Illinois Urbana-Champaign, Champaign, IL, USA.
[3] Siebel School of Computing and Data Science, University of Illinois Urbana-Champaign, Champaign, IL, USA.
[4] Department of Robotics and Mechatronics Engineering, Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu, South Korea
[5] Robot Research & Development Team in HD Hyundai Robotics, Daegu, South Korea

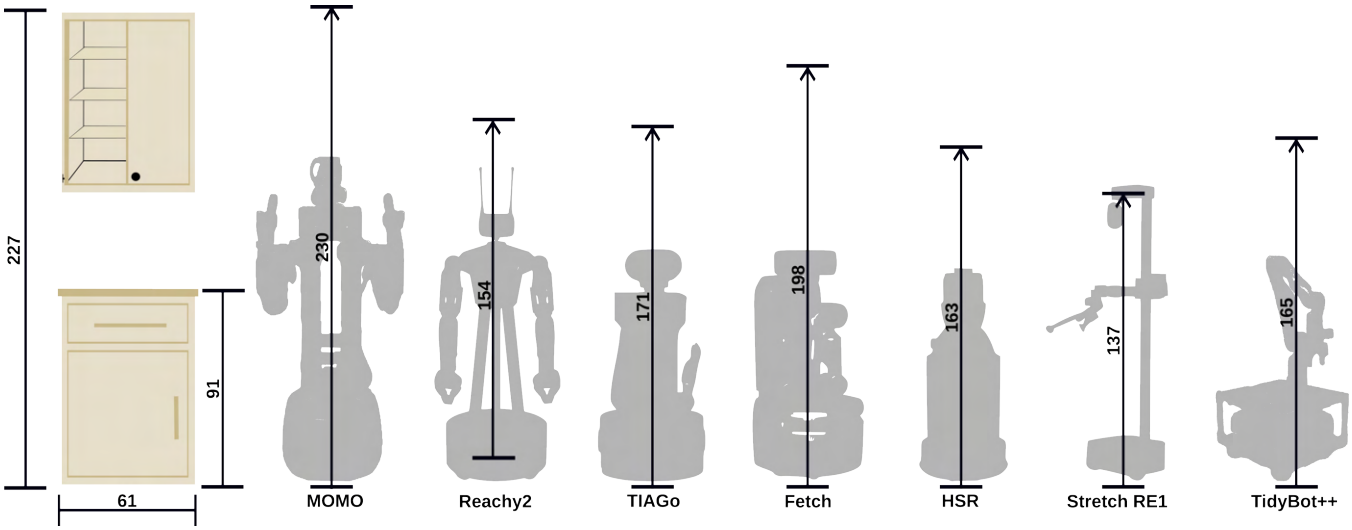

Fig. 2: Comparison of common mobile manipulators and their reachability compared to a common kitchen cabinet.

TABLE I: Comparison of Wheeled Mobile Manipulator Systems

Mobile Manipulator	Dimensions [Wcm x Lcm x Hcm]	Weight [kg]	Payload [kg]	Vertical Span [cm]	DoF
Average Food Runner [4]	54 x 52 x 127	46	42 (On Trays)	—	2
KIMLAB: MOMO (No Attachments)	48 x 53 x 133	62	15 (On Trays)	—	4
KIMLAB: MOMO (Dual Arm/Head)	48 x 53 x 141	74	3 (Per Arm)	0 - 192	21
KIMLAB: MOMO (Three Arm)	48 x 53 x 152	79	3/2.5 (Per Short/Long Arm)	0 - 230	25
Pollen Robotics: Reachy2 [1] [2]	50 x 50 x 136 - 166	50	3 (Per Arm)	13 - 167	22
PAL Robotics: TIAGo [3] [4]	54 x 54 x 110 - 145	70	3 (No Gripper)	0 - 171	14
Fetch Robotics: Fetch [3] [4]	51 x 56 x 110 - 149	113	6 (No Gripper)	0 - 198	13
Toyota: HSR[3] [5]	43 x 43 x 101 - 103	37	1.2	0 - 163	11
Hello Robotics: Stretch RE1[3] [4]	34 x 33 x 141	23	1.5 (No Gripper)	0 - 137	9
TidyBot++[6] [5]	50 x 54 x 102	34	3.5	0 - 165	11

environments such as being able to navigate through the ever-changing restaurant, but also having the dexterity to interact with and alter their surroundings. The capabilities of these systems habe been tested by performing pick-and-place tasks [6], [7] as well as more complicated tasks such as cleaning a kitchen [8], folding clothes [5], and even cooking [9], [10].

In these previous works, several innovative designs have been presented, but they all come with major trade-offs between form factor, reachability, and compactness. Several modular robots have emerged with reconfigurability in mind to overcome these trade-offs. Modular robots are designed with the flexibility to adjust structure and functionality to meet a wide array of operational demands. Instead of requiring entirely new systems for different applications, users can modify key components, making these robots highly scalable and adaptable. Many mobile manipulators [11], [12] embody this principle of reconfigurability through the use of linear vertical and horizontal lifts to adapt to various tasks. MOMO extends this concept by using pluggable components on a serving robot base to enable mobile manipulation. As seen in Table I, its modular configurations facilitate trade-offs between manipulation capability and form factor, enabling the system to adapt to spatial and task-specific requirements. Table I also highlights MOMO's advantage with its weight-to-payload ratio, enabling it to lift heavier objects without sacrificing performance.

As shown in Fig. 2, we compare the vertical reach of common mobile manipulators used in industry and academia to a standard kitchen cabinet. These systems utilize wheeled mobile bases with footprints similar in size to the traditional serving robot, approximately 53 cm x 52 cm. The design of these systems focuses on general-purpose tasks in a human-centric application, such as assisting in homes. While most designs of mobile manipulators can access objects placed at countertop height, MOMO can reach objects placed on upper cabinets. MOMO takes a unique design approach to achieve its competitive vertical span. MOMO employs distinct shoulder lifts for each arm, in contrast to a conventional single torso lift, which confines both arm bases to the same vertical plane in dual-arm settings. Separate shoulder lifts enable multi-level manipulation, such as retrieving an item from the floor while simultaneously interacting with objects on a

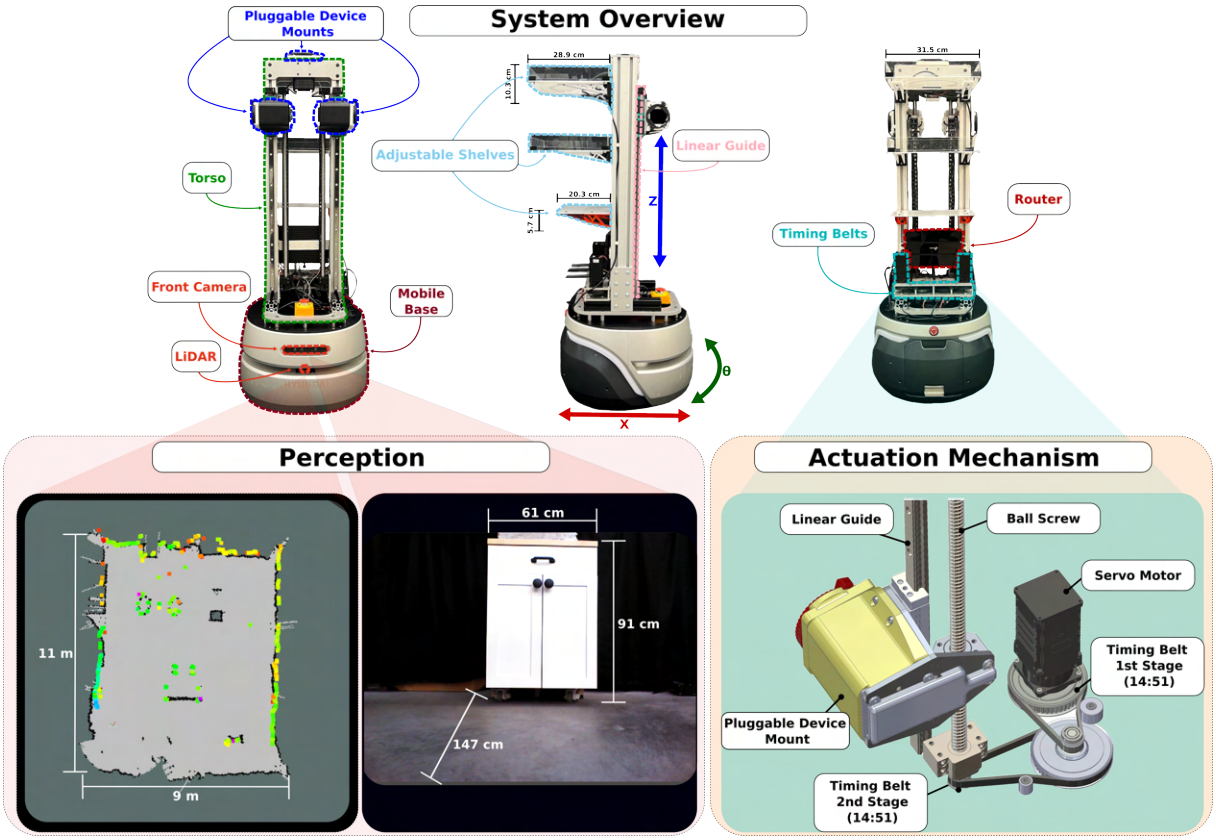

Fig. 3: Overview of the proposed mobile manipulator. The perception module is depicted in two figures: the left figure illustrates the LiDAR's field of view, and the right figure shows the front camera's field of view.

countertop, which is difficult to achieve in single-lift designs.

III. System Overview

In this work, we propose modifications to a traditional serving robot's design to enhance its adaptability in environments. We achieve this through the optional use of pluggable systems. To this end, our design considerations include the following:

1) **Form Factor:** The system should resemble a traditional food runner, prioritizing compact and agile form to navigate cluttered spaces and narrow passages, such as doorways, while maintaining the capability to carry objects.
2) **Adaptability:** Integration with pluggable components should ensure the system can adapt its form for versatile tasks, allowing the system to navigate and interact with dynamic environments while performing manipulation tasks.
3) **Individuality:** Each component should operate independently, yet collaborate during complex tasks.

[1]https://www.pollen-robotics.com/wp-content/uploads/2024/10/Reachy2-Dual-arms-with-mobile-base-Datasheet.pdf
[2]https://github.com/pollen-robotics/Reachy2Teleoperation/
[3]https://robotsguide.com/
[4]https://github.com/robot-descriptions/awesome-robot-descriptions
[5]https://github.com/ToyotaResearchInstitute/hsr description
[6]https://tidybot2.github.io/docs/

We further discuss the design of the mobile base and torso components.

A. Mobile Base

We used a commercialized mobile base, Hyundai B1, as the foundation for MOMO's mobility, considering its dimensions, payload capacity, and navigation. The mobile base has a differential drive system, which allows smooth navigation in its surroundings at a maximum speed of 1.0 m/s. The compact dimensions of the mobile base, measuring 480 mm x 523 mm x 330 mm, facilitate its seamless travel through passageways. MOMO houses two PCs: the Mobile Base PC orchestrates the control of peripheral devices, including motors and various sensors such as a LiDAR, depth camera, and IMU, essential for the system to navigate its environment. A comprehensive summary of the available sensors is presented in Table II. The second PC is the Torso PC, which bridges the gap between the pluggable devices and the mobile base. The communication between the two computers is facilitated through a wired connection.The mobile base's battery powers both computers, providing up to eight hours of continuous operation.

B. Torso

The torso is mounted to the mobile base using screws and stand-offs at the top, streamlining modularity, the assembly process, and maintenance accessibility for both the torso

TABLE II: System Specifications

Submodule	Specification	Details
Mobile Base	Weight	38 kg
	Dimension	480 x 523 x 330 mm
	Max Speed	1.0 m/s
	Payload	50.0 kg
	LiDAR	SICK TiM571 LiDAR
	IMU	BNO055 IMU
	Front Camera	Orbbec Astra Camera
	Peripheral PC	Nvidia Jetson TX2
	Torso PC	Yanling Mini PC
Torso	Weight	23.8 kg
	Dimension	423 x 387 x 996 mm
	Actuator	PH54-200-S500-R (200W)
PAPRAS	Length	Long Arm: 894 mm Short Arm: 794 mm
	Weight	Long Arm: 5.776 kg Short Arm: 5.600 kg
	Actuators	PH54-200-S500-R (200W) PH54-100-S500-R (100W) PH42-020-S300-R (20W)
	Eye-On-Hand Camera	Intel RealSense™D435
	Gripper	RH-P12-RN (Parallel Jaw)
Sensor Module	Weight	0.603 kg
	Dimension	90 x 114.8 x 170 mm
	Actuator	XM-430-W350R
	Camera	Intel RealSense™D435i

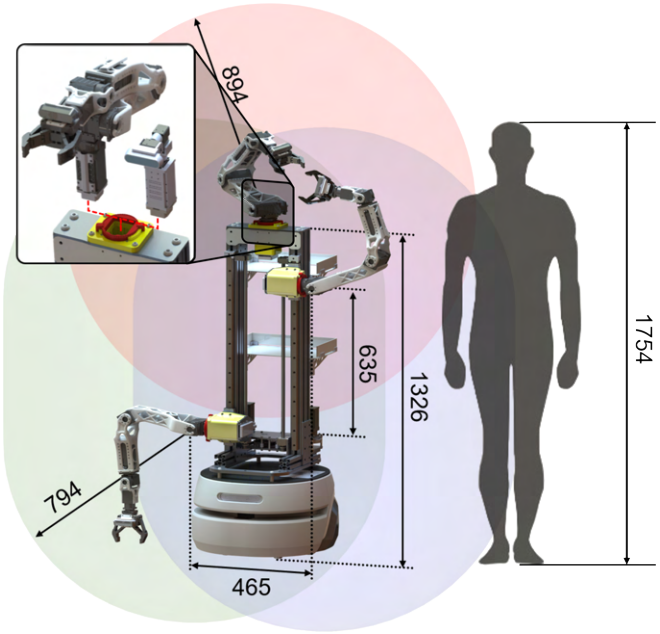

Fig. 4: Overall CAD drawing of MOMO System. The workspace of the head mount (Red), Right shoulder (Green), and Left shoulder (Blue) with PAPRAS are illustrated with shaded areas. All dimensional units are in millimeters. As shown, the mounts can take in pluggable devices. These mounts are housed on the left, right, and top of MOMO's torso. The combined height of the torso and mobile base is 1326 mm, the vertical lift has a range of 635 mm, and the shoulder width, measured between the axes of the second joint of the two manipulators is 465 mm.

and the mobile base. Standing at a height of 1.33 m, MOMO is slightly taller than an average serving robot. The torso is constructed from aluminum extrusions and brackets, resulting in a torso frame that weighs 23.8 kg. Three height-adjustable shelves are attached to the rear of the torso, as shown in Fig. 3. These shelves are designed to handle a payload of five kg to help facilitate object transport. To transform MOMO from a traditional serving robot into a mobile manipulator, we propose integrating the Plug-and-Play Robotic Arm System (PAPRAS) [2]. By equipping MOMO with pluggable devices, the system is capable of interacting with its environment, expanding its utility in real-world applications. Three PAPRAS mounts are located at the neck and on each side of the torso, as seen in Fig. 3. These mounts are supplied with 24 V DC power directly from the mobile base's battery. PAPRAS can collaborate in overlapping workspaces for more complex tasks. This design significantly boosts MOMO's overall manipulation capabilities. Beyond manipulation, the MOMO system can enhance perception by integrating a 3-DoF camera-based sensor module, using the same plug-in port as the arm, as shown in Fig. 4. This addition enhances the sensing capabilities of the mobile robot and provides the ability to transmit nonverbal, non-display communicative signals through human-like head motions [3], enhancing MOMO's versatility and effectiveness in various environments and human-robot interaction applications. The pluggable systems can be inserted easily and detected by the system within seconds as seen in the supplementary video.

The system's mobility effectively mitigates horizontal range constraints for manipulation tasks. Therefore, the focus of our design is to expand the vertical reach of the system to broaden the range of applications the system can effectively handle. Directly increasing the size of the manipulator to expand its workspace may not be the most effective solution, as it typically requires greater torque and results in larger, heavier components. These changes can consequently limit the overall functionality of the mobile manipulator. MOMO overcomes this issue through its unique linear shoulder mechanism. As illustrated in Fig. 3, the drive train for one side of the shoulder mechanism comprises the following key components: a ball screw, a linear guide, a timing belt pulley transmission system, and a servo motor. The linear lift utilizes the same servo motors as the six-DoF manipulators, simplifying software integration because all the actuators can be controlled within the same serial network. While the motor's high-gear reduction ratio offers ample torque, a direct connection to the ball screw would result in slow linear speed. To optimize for speed, a two-stage timing pulley system is employed, increasing the output rotational speed with a 1:13.5 gear ratio. This modification allows the linear shoulder mechanism to achieve a speed of 65.39 mm/s, traveling its full range in about 10 seconds. To ensure homing accuracy and safety, optical stop switches are connected to the lifts'

servo motors, with the switches located at the lift's zero pose. Cable chains are employed for wiring organization and safety. A summary of all the available actuators, sensors, and specifications is presented in Table II.

C. Linear Lift Characterization

To evaluate the positioning accuracy of MOMO's pluggable arm and assess the effect of the ballscrew-driven linear lift, we conduct a series of repeatability and tracking experiments. Following the ISO 9283 standard [13], we select five representative points that span the arm's reachable workspace and define a feasible trajectory passing through these points. The trajectory is executed ten times under two conditions: (1) no payload and (2) with a 1.5 kg payload, which exceeds the typical object weight encountered in a restaurant setting (less than one kg [14]). The end effector's motion is tracked using a high-precision motion capture system, and the error between the commanded and actual trajectories is calculated. In the no-payload condition, the root mean squared error (RMSE) of the trajectory is six mm. With a 1.5 kg payload, the RMSE increases slightly to 10 mm. The highest deviations are observed when the end effector is near the edge of the workspace (see Fig. 5). These errors, while within an expected range, could potentially impact the reliability of pick-and-place tasks in a restaurant setting. To mitigate the effects, MOMO's manipulator is equipped with eye-on-hand cameras that provide local visual feedback relative to the mobile base, helping to account for the observed errors. Additionally, MOMO can use planning strategies to reposition its mobile base closer to the target object, which typically reduces the errors associated with picking up objects at the edge of the workspace.

To isolate the error introduced by the ballscrew mechanism, we perform a separate experiment in which a sinusoidal input drives the lift throughout its full range of motion, using the pluggable arm as a payload. As shown in Fig. 5, the resulting positioning error remained consistently below one mm, confirming that the ballscrew introduces negligible error in typical operation.

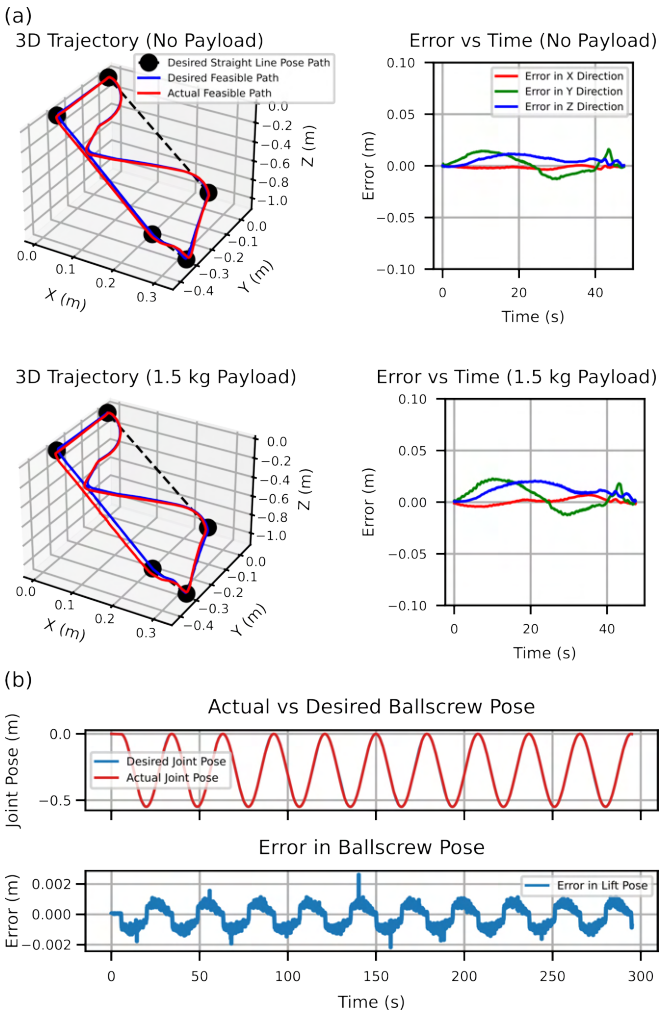

Fig. 5: (a) The 3D tracking error of the arm and prismatic joint with and without a payload. (b) The isolated error of the prismatic joint with the pluggable arm as a payload.

D. Dynamic Stability and Center of Mass Aware Motion Control

With a design that focuses on upward exploration, the system's center of mass (CoM) is shifted upwards. To ensure dynamic stability, a CoM aware motion controller that uses the Zero Moment Point (ZMP) criterion based on the linear inverted pendulum (LIP) model [15] is developed. By analyzing the system's URDF and the mass properties of each link, the controller calculates the CoM based on joint positions in real time. It then uses this information to determine the maximum safe acceleration and deceleration, ensuring the CoM stays within the support polygon defined by the mobile base's footprint. This approach is then evaluated against a first-order smoothing filter (naive controller).

To evaluate the performance of this controller, we conducted a set of dynamic braking experiments. In each trial, MOMO—configured with three manipulators—accelerates to a target speed, cruises briefly, and then decelerates as quickly as possible. We tested five different joint configurations that shift the CoM in both vertical and horizontal directions: a neutral resting pose, a low-CoM pose with arms tucked in, a high-CoM pose with arms raised, and two manipulation scenarios—one holding a broom and one with three arms extended forward while carrying 0.5 kg weights. In the experiments, the goal velocity is set to be 0.5 m/s.

The resulting velocity profiles are shown in Fig. 6 for both the CoM aware and naive controller. Table III summarizes the CoM aware controller's minimum and maximum acceleration, stopping distance from peak velocity, and whether the system tipped over. While the CoM aware controller requires longer stopping distances, it successfully prevented tipping in all cases. Thanks to MOMO's onboard sensors and long-range perception, it can anticipate braking requirements well in advance, helping offset the need for a longer deceleration phase. While we use a LIP model to simplify controller design, more accurate physical modeling or data-driven system identification could further enhance performance.

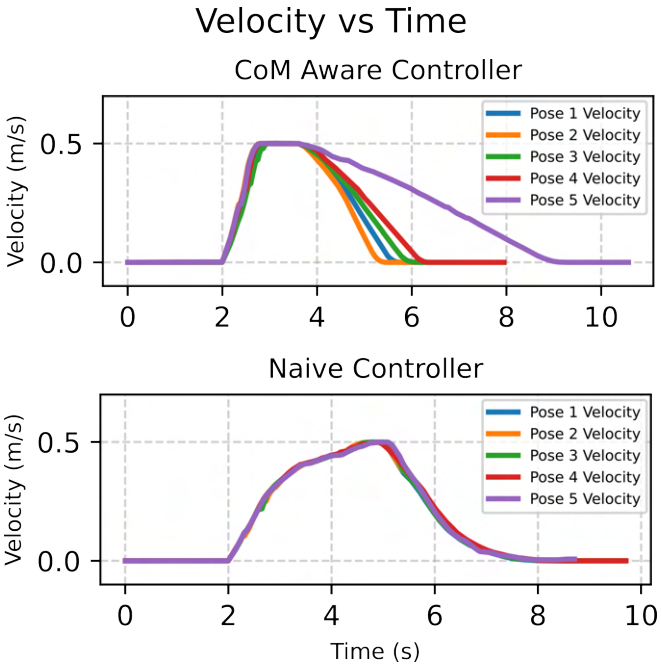

Fig. 6: The velocity profiles under different joint configurations illustrate how the CoM aware controller adjusts acceleration and deceleration based on shifts in the center of mass, preventing the system from tipping.

TABLE III: CoM Aware Controller Effects

Pose	Min / Max Accel. m/s^2	Stopping Distance (m)	Tipped CoM Aware (T/F)	Tipped Naive (T/F)
1	-1.26/3.56	0.55	F	F
2	-1.55/4.39	0.46	F	F
3	-1.04/3.26	0.62	F	F
4	-0.88/3.99	0.71	F	F
5	-0.35/4.23	1.42	F	T

IV. Software Control

In this section, we explore the software components required to transition MOMO's capabilities from traditional serving to complex tasks. The software section is divided into two sections: the offboard and onboard control. The onboard control consists of the Mobile Base PC and the Torso PC, while the offboard control consists of wirelessly connected devices that interface with MOMO. The overall software architecture is illustrated in Fig. 7. To help with communication and coordination, the entire system operates with a locally distributed network. The Robot Operating System (ROS1) [16] is used to facilitate the communication between devices with the Mobile Base PC set as the ROS1 master.

A. Mobile Base PC

The Mobile Base PC controls the movement of the mobile base's actuators and governs the various sensors integrated into the system, both on the torso and the mobile base itself. These sensors play a crucial role in gathering information about the robot's surroundings and the current environment. A summary of the available sensors for the prototype is found in Table II. This PC is also responsible for publishing the robot's odometry, which is used by the localization, mapping, and navigation modules.

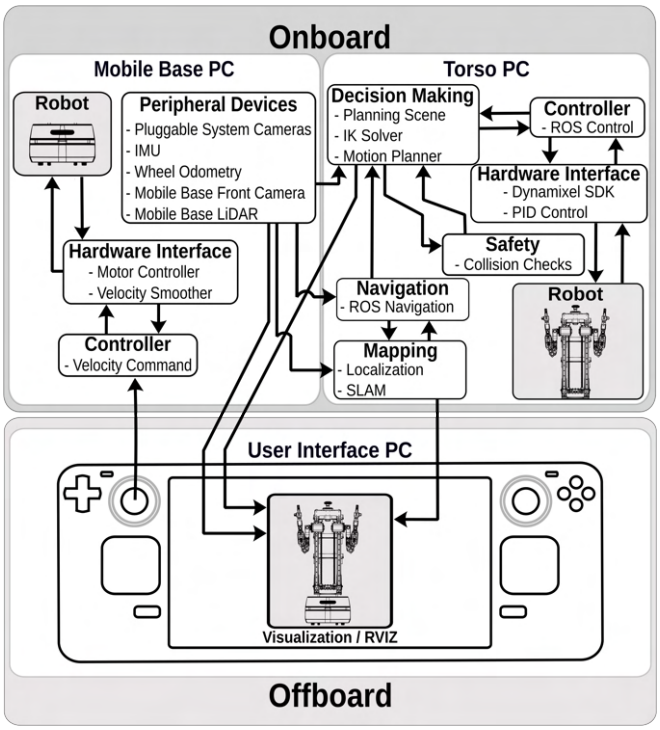

Fig. 7: MOMO Software Architecture: The offboard contains external tools to interface with MOMO, while the onboard contains the software that manages the sensors and actuators. The offboard PC communicates with the onboard PC through WIFI. The Mobile Base PC handles all the peripherals and perception, while the Torso PC handles the motion planning of the mobile base and pluggable devices. The Mobile Base PC and Torso PC communicate through Ethernet.

B. Torso PC

The Torso PC communicates with the linear lift mechanism and the modular components plugged into the pluggable mount. To generate feasible motion plans, we leverage the MoveIt framework [17]. The Torso PC receives information provided by the sensors connected to the Mobile Base PC regarding the manipulable objects and the overall movement of the mobile base. This data is then relayed to the planning scene for obstacle avoidance. When specific actions like grasping or linear lift movement are required, the Inverse Kinematics (IK) Solver comes into play. In our implementation, we utilize OpenRAVE IKFast as our IK solver [18]. Subsequently, a trajectory is generated using the OMPL motion planning library [19], which is then transmitted to the joint trajectory controller.

The joint trajectory controller feeds the hardware interface with dynamically feasible commands to the actuators. Furthermore, the controller ensures a real-time update for the

planning scene, maintaining an up-to-date state for future calculations.

In addition to controlling the pluggable devices, the Torso PC controls MOMO's navigation module by utilizing Simultaneous Localization and Mapping (SLAM) techniques powered by its LiDAR sensors. SLAM allows MOMO to autonomously map unknown environments and localize itself within saved maps using Adaptive Monte-Carlo Localization (AMCL) [20]. With this mapping and localization capability, MOMO navigates the environment using a combination of global path planning and local planning to avoid unexpected obstacles not included in initial maps. The range of sensing capabilities from the mobile base camera and LiDAR can be seen in Fig. 3.

C. Offboard PC

The offboard component of the software architecture is responsible for the user interface and input for MOMO. We primarily use the Steam Deck[7] to directly control the mobile base by sending velocity twist messages through the joystick of the controller. The buttons on the controller allow the user to control the manipulators and execute predefined motions. The onboard screen provides access to the current visualization of MOMO's current state. Through the visualization and user interface on the screen, the user can set the end-effector positions of the manipulators to have MOMO interact with the environment.

V. Demonstrations and Experiments

The main feature of MOMO is its ability to reconfigure itself from transportation capabilities to mobile manipulation capabilities. The system's ability to reconfigure itself with up to three arms or a head allows the system to conform to the physical requirements of a task. To illustrate this transformation, we conduct various demonstrations that showcase MOMO's advanced functionalities through application-oriented implementations. This approach aims to highlight the flexibility of the system and its potential in a variety of real-world contexts.

A. Modularity

To demonstrate MOMO's modularity and versatility, we initially configure it with a single manipulator attached to the neck joint. By integrating a manipulator as MOMO's head, the system maintains the form factor of a conventional food runner robot with a slight modification. This simple yet powerful modification enables MOMO to gather, transport, and deliver packages without any human assistance, marking a leap in functionality.

The demonstrated advanced serving capability is illustrated in Fig. 8, task one. In this demonstration, MOMO starts in the kitchen area and utilizes an eye-in-hand camera to scan the kitchen counter for an object to be transported. Once the object is identified, MOMO grasps it and securely places the object on its back tray. MOMO then navigates to the intended location. Upon reaching the destination, the arm retrieves the object from the tray and either places it down or hands it off to an intended user. In a similar scenario, MOMO could be equipped with two manipulators and the sensor module head. With the head, MOMO is capable of performing human-like nodding motions to help convey intent to the user. A comprehensive demonstration of these functionalities, including multiple configurations, is available in the supplementary video material.

Although MOMO extends the capabilities of a traditional serving robot with manipulators, it needs to adapt to the human-centric world. Therefore, MOMO will not only need to reach the height of common countertops, but also extend its reach to overhead cabinets as well. In scenarios where MOMO needs high vertical reachability, the system allows for a configuration of three manipulators. In this configuration, MOMO's workspace can reach from the floor to high kitchen cabinets through interactions between the three manipulators. To showcase MOMO's modularity and ability to expand its workspace, the system is tasked with grasping an obstruction off the floor and placing it in a kitchen cabinet. First, MOMO scans the floor for the object. Once the object is detected, MOMO uses one of its shoulder manipulators to pick up the object. The object is then handed off from the floor to the manipulator mounted on the neck. MOMO then navigates to the cabinet and localizes it. Once localized, MOMO coordinates the movement of its three arms by using the two shoulder manipulators to open the cabinet door and place the object on the cabinet shelf. The second row of Fig. 8 illustrates the sequence of these steps. MOMO demonstrates its ability to transfer and leverage the use of all three arms. These demonstrations show the importance of being able to explore different configurations and functionalities in a unified robotic platform. The ability to effortlessly transition between various setups showcases the versatility of the system.

B. Mobility

To be a mobile manipulator, the system should be able to navigate an environment. To demonstrate MOMO's autonomous navigation capability, we conducted experiments in a lab environment with common household items like cabinets and bookshelves to best replicate a real-life kitchen or living room environment. The room dimensions are approximately 11 m x 9 m and the room configuration can be seen in Fig. 9. MOMO starts at the kitchen counter and navigates to one of five locations where an obstruction is placed in the environment. At runtime, MOMO determines a global plan illustrated by the different paths in Fig. 9 and navigates to the vicinity of the desired object. Once successfully navigated, it utilizes the camera on the mobile base to determine the precise location of the object with the help of an April tag. MOMO determines where to move so the object is within the reachable workspace. This decision is made by sampling if there is a feasible inverse kinematic solution to grasp the object by positioning the mobile base at a sampled desired position. Once the mobile base moves to the desired position, MOMO utilizes the eye-on-hand camera of the pluggable arm to re-localize the object and pick it up. The system then

[7]https://store.steampowered.com/steamdeck/

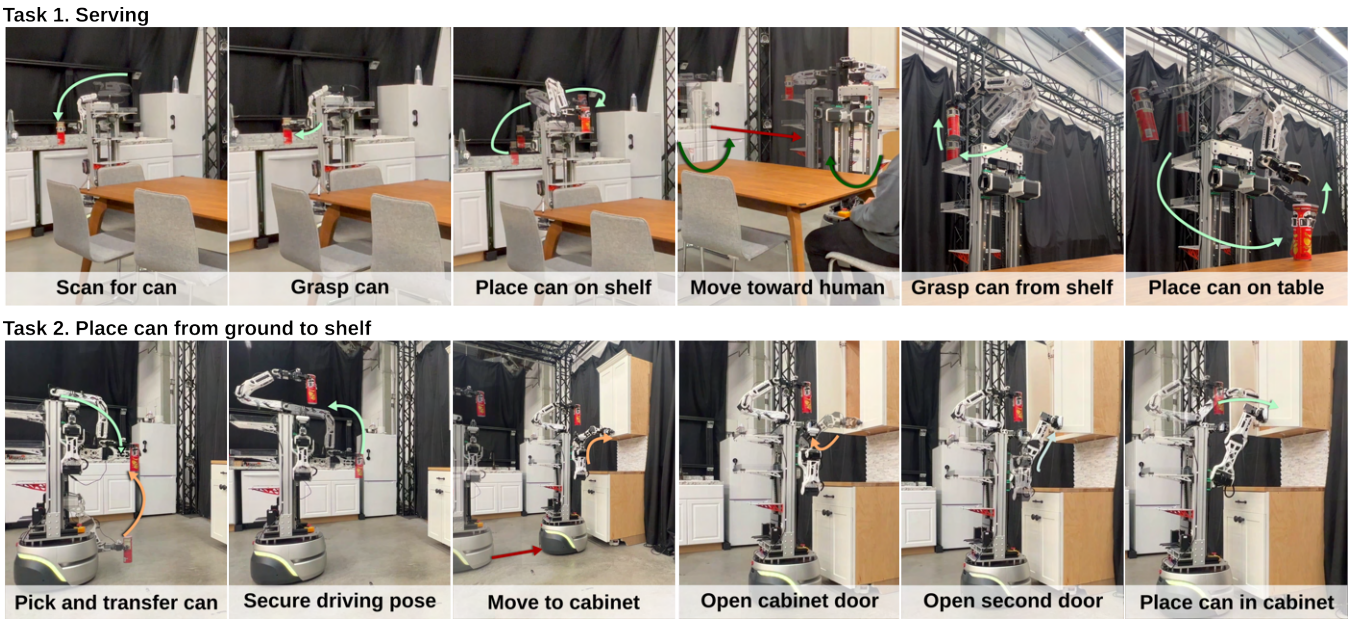

Fig. 8: Demonstration of the modularity of the proposed system. (Task one) MOMO utilizes a single arm on the neck mount to show its ability to extend the capabilities of a traditional serving robot by grasping, transporting, and placing an object without any human assistance. (Task two) Demonstrates how in the three-arm configuration, MOMO's workspace is extended by grasping an obstruction off of the floor, handing it over to the arm on the neck mount, and then placing the obstruction on a shelf in a cabinet after opening the cabinet door.

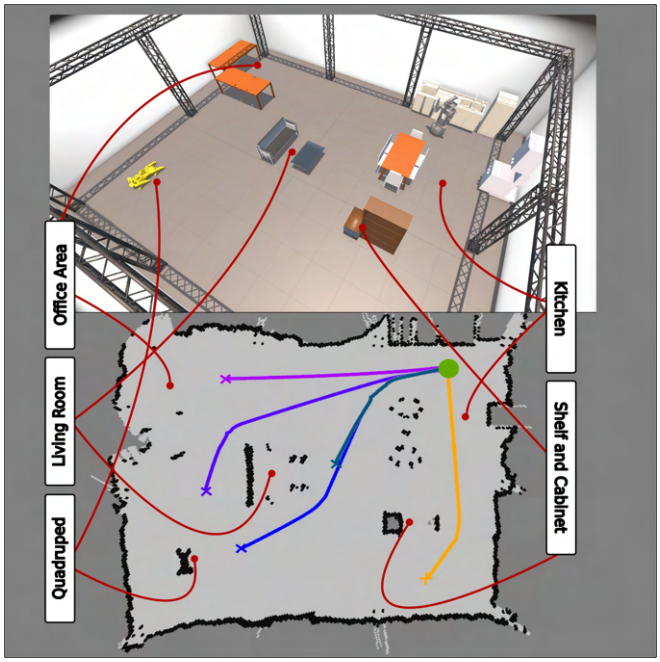

Fig. 9: The mapped environment with common household furniture. For all five of the trials, MOMO started from the same position and found paths to different goal locations indicated by the cross. The different poses one through five are displayed by different colored paths (blue, magenta, turquoise, orange, and purple respectfully).

TABLE IV: Mobile Navigation Timings

Pose	Entry Time (s)	Grasping Time (s)	Exit Time (s)	Place Time (s)	Total Time (s)
1	28	137	29	55	249
2	23	137	23	55	238
3	21	142	15	55	233
4	25	139	39	56	259
5	25	136	25	57	243

grasps the object and hands it off to the manipulator located on its neck. After the object is handed off, MOMO navigates to the cabinet and places the transported object inside the cabinet.

Table IV provides a breakdown of the time taken for MOMO to reach each position, grasp the object, navigate to the kitchen cabinet, and place the object in the cabinet. Almost all of the poses, except for pose four, had an equal amount of time between the entry time and exit time. Pose four's exit time was greater than the entry time because as the robot was trying to move back to the cabinet, there was a narrow passage that caused the planner to act more cautiously to avoid hitting the furniture. This task demonstrates the system's ability to navigate a simulated home environment, perform common tasks like putting away objects, and operate effectively in real-world settings.

C. Workspace

MOMO's expansive workspace is demonstrated using a household bookshelf scenario with three cups placed on differ-

ent shelves. MOMO begins by localizing the bookshelf using the camera on its mobile base, simultaneously identifying any cups on the lowest shelf. The manipulator mounted on its head then scans the upper shelves from top to bottom, pinpointing the locations of additional cups. After identifying all the cups, MOMO calculates the optimal base position to grasp three cups simultaneously, assigning each manipulator a cup based on its respective location and height. Once the best position is determined, MOMO navigates to the desired spot and executes the grasping sequence. All three arms move concurrently, avoiding inter-arm collisions using the planning framework. After successfully grasping the cups, MOMO safely transports them to a designated table, demonstrating its ability to manipulate multiple objects in a multi-level environment.

Table V shows the total time and individual breakdown for localization, grasping, and placing the objects across five trials. In Table V the placements of the three cups relative to the first cup are recorded, and a bounding box encompassing all target positions is computed. This bounding region represents the effective workspace MOMO is able to cover with all three arms simultaneously. Notably, the system successfully interacts with all objects within this area in parallel, demonstrating its capacity to coordinate multiple manipulators across an expansive manipulation zone. The current planning framework has limitations with multi-arm movements, slowing down the operation of all the manipulators. The motion planner's requirement for all arm motions to end simultaneously, coupled with the motion planner's tendency to select first-available solutions, results in issues. Additionally, the planner's reliance on random sampling causes problems navigating constrained environments, as seen with the bookshelves. Future work will focus on developing a custom framework to improve the motion planning and utilizing the full speed of the system. The robot's capability to simultaneously grasp three objects not only demonstrates the independent functionality of each arm but also highlights their synergy in accomplishing complex tasks.

TABLE V: Bookshelf Task: Time and Workspace Metrics

Task	Total Time (s)	Cup Positions (X, Y, Z) (m)	Vertical Workspace (m^2)
1	213 (21/76/28/88)	(0.0, 0.0, 0.0) (-0.03, -0.26, 0.44) (-0.03, -0.39, 0.0)	0.17
2	287 (39/132/28/88)	(0.0, 0.0, 0.0) (-0.04, -0.24, 1.38) (-0.04, -0.37, 0.93)	0.51
3	281 (18/147/28/88)	(0.0, 0.0, 0.0) (-0.02, -0.31, 1.84) (0.01, -0.6, 0.00)	1.11
4	293 (31/146/28/88)	(0.0, 0.0, 0.0) (-0.03, -0.36, 1.85) (-0.02, -0.51, 0.95)	0.94
5	296 (32/148/28/88)	(0.0, 0.0, 0.0) (0.0, 0.51, 1.40) (0.0, 0.51, 1.40)	0.71

† (Scan/Grasp/Navigate/Place) time is detailed below Total Time.
‡ Vertical Workspace is the workspace area, Y*Z.

VI. Conclusion

Unlike their human counterparts, food delivery robots are currently limited to transporting food. Expanding their capabilities with manipulation will expand their capabilities beyond just food delivery. To this end, we present the design and implementation of Mobile Object Manipulation Operator, which takes inspiration from a traditional serving robot and incorporates the ability to utilize up to three independent pluggable devices.

The modular design enables flexible, task-driven exploration of hardware configurations across a single system. We presented the software control of the system, which consists of the Mobile Base PC, the Torso PC, and an external controller used for the teleoperation of MOMO. Specifically, the Mobile Base PC handles the movement of the mobile base and any peripherals, while the Torso PC manages the planning scene, planning pipeline, and actuation of the pluggable robotic systems.

To demonstrate the increased capabilities of MOMO, we provided demonstrations that highlighted MOMO's modular and mobility capabilities of reconfiguring its expansive workspace depending on the task at hand.

Although MOMO has an expansive workspace and unique design, the system has a higher center of mass compared to other mobile manipulators. We developed a software-based solution to show how we can mitigate these effects, but the physical limitation remains a challenge. In the future, incorporating a moment compensator could assist the system's balance. Further work will include improvements by enhancing the software modules, such as perception and planning, to increase the system's autonomy. By using a more customized planning framework, better motion plans would enable the full speed of the system while still ensuring collision-free motions.

Acknowledgments

The material presented in this work is partially supported by HD Hyundai Robotics.

References

[1] SNS Insider. (2024) Restaurant delivery robot market size & share, report 2032. SNS Insider. [Online]. Available: https://www.snsinsider.com/reports/restaurant-delivery-robot-market-3493

[2] J. Kim, D. C. Mathur, K. Shin, and S. Taylor, "Papras: Plug-and-play robotic arm system," 2023. [Online]. Available: https://arxiv.org/abs/2302.09655

[3] C. Moon, S. Yamsani, and J. Kim, "Development of a 3-dof interactive modular robot with human-like head motions," in *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2023, pp. 141–146.

[4] ERIA Company. (2025) Discover our food delivery robots. [Online]. Available: https://www.eria.company/blog/robot-as-a-service-2/discover-our-food-delivery-robots-17

[5] J. Wu, W. Chong, R. Holmberg, A. Prasad, Y. Gao, O. Khatib, S. Song, S. Rusinkiewicz, and J. Bohg, "Tidybot++: An open-source holonomic mobile manipulator for robot learning," in *Conference on Robot Learning*, 2024.

[6] C. C. Kemp, A. Edsinger, H. M. Clever, and B. Matulevich, "The design of stretch: A compact, lightweight mobile manipulator for indoor human environments," in *2022 International Conference on Robotics and Automation (ICRA)*, 2022, pp. 3150–3157.

[7] M. Costanzo, G. De Maria, G. Lettera, and C. Natale, "Can robots refill a supermarket shelf?: Motion planning and grasp control," *IEEE Robotics & Automation Magazine*, vol. 28, no. 2, pp. 61–73, 2021.

[8] T. Lew, S. Singh, M. Prats, J. Bingham, J. Weisz, B. Holson, X. Zhang, V. Sindhwani, Y. Lu, F. Xia *et al.*, "Robotic table wiping via reinforcement learning and whole-body trajectory optimization," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 7184–7190.

[9] Z. Fu, T. Z. Zhao, and C. Finn, "Mobile aloha: Learning bimanual mobile manipulation with low-cost whole-body teleoperation," in *Conference on Robot Learning (CoRL)*, 2024.

[10] K. Kawaharazuka, Y. Obinata, N. Kanazawa, K. Okada, and M. Inaba, "Robotic state recognition with image-to-text retrieval task of pre-trained vision-language model and black-box optimization," in *2024 IEEE-RAS 23rd International Conference on Humanoid Robots (Humanoids)*, 2024, pp. 934–940.

[11] M. Car, B. A. Ferreira, J. Vuletic, and M. Orsag, "Structured ecological cultivation with autonomous robots in agriculture: Toward a fully autonomous robotic indoor farming system," *IEEE ROBOTICS & Automation Magazine*, vol. 30, no. 4, pp. 77–87, 2023.

[12] G. Gorjup, C.-M. Chang, G. Gao, L. Gerez, A. Dwivedi, R. Yu, P. Jarvis, and M. Liarokapis, "The aroa platform: An autonomous robotic assistant with a reconfigurable torso system and dexterous manipulation capabilities," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 4103–4110.

[13] ISO, "Manipulating industrial robots — performance criteria and related test methods," International Organization for Standardization, Standard ISO 9283:1998, 1998. [Online]. Available: https://www.iso.org/standard/22244.html

[14] L. E. Urban, J. L. Weber, M. B. Heyman, R. L. Schichtl, S. Verstraete, N. S. Lowery, S. K. Das, M. M. Schleicher, G. Rogers, C. Economos, W. A. Masters, and S. B. Roberts, "Energy contents of frequently ordered restaurant meals and comparison with human energy requirements and u.s. department of agriculture database information: A multisite randomized study," *J. Acad. Nutr. Diet.*, vol. 116, no. 4, pp. 590–8.e6, Apr. 2016.

[15] S. Sugano, Q. Huang, and I. Kato, "Stability criteria in controlling mobile robotic systems," in *Proceedings of 1993 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'93)*, vol. 2. IEEE, 1993, pp. 832–838.

[16] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, A. Y. Ng *et al.*, "Ros: an open-source robot operating system," in *ICRA workshop on open source software*, vol. 3, no. 3.2. Kobe, 2009, p. 5.

[17] S. Chitta, I. Sucan, and S. Cousins, "Moveit! [ros topics]," *IEEE ROBOTICS & Automation Magazine*, vol. 19, no. 1, pp. 18–19, 2012.

[18] R. Diankov, "Automated construction of robotic manipulation programs," Ph.D. dissertation, Carnegie Mellon University, Robotics Institute, August 2010. [Online]. Available: http://www.programmingvision.com/rosen_diankov_thesis.pdf

[19] I. A. Şucan, M. Moll, and L. E. Kavraki, "The Open Motion Planning Library," *IEEE Robotics & Automation Magazine*, vol. 19, no. 4, pp. 72–82, December 2012, https://ompl.kavrakilab.org.

[20] E. Marder-Eppstein, "Ros navigation [software]." [Online]. Available: http://wiki.ros.org/navigation