UTBots@Home 2024 Team Description Paper

1st Gustavo Fardo Armênio[1]
gustavofardoarmenio@alunos.utfpr.edu.br

2nd Bruno Bock Silva[2]
brunobock@alunos.utfpr.edu.br

3rd David Segalle[1]
davidsegalle@alunos.utfpr.edu.br

4th Victor Boechat Errera[1]
verrera@alunos.utfpr.edu.br

5th João Henrique Cordeiro Pires[2]
joaopires@alunos.utfpr.edu.br

6th Enzo Holzmann Gaio[1]
enzoholzmanngaio@gmail.com

7th Larissa Bialski Lima da Silva[2]
larissabialski8@gmail.com

8th Bruno Emanuel Zenatti
1
brunozenatti@alunos.utfpr.edu.br

9th Christoffer Nogueira Spring[1]
christoffer@alunos.utfpr.edu.br

10th Gabriel Tsuneo Oshiro Shinzato[2]
gabrielshinzato@alunos.utfpr.edu.br

11th Gabriel Spadafora[1]
gspadafora@alunos.utfpr.edu.br

12th João Pedro dos Reis Mendes[1]
mendes.2021@alunos.utfpr.edu.br

13th Luiz Guilherme Silvino Greca[2]
luizgreca@alunos.utfpr.edu.br

14th Luiza Ribeiro Romero Insfran Ocampos[2]
ocampos@alunos.utfpr.edu.br

15th Matheus dos Santos Pedrozo de Lima[1]
matheus.pedrozo@gmail.com

16th Marlon de Oliveira Vaz[2, 3]
marlon.vaz@ifpr.edu.br

17th André Cristiano Cordeiro[1]
ac.cordeyro@outlook.com

18th André Schneider de Oliveira[2]
andreoliveira@utfpr.edu.br

19th João Alberto Fabro[1]
fabro@utfpr.edu.br

***Abstract*—Following UTBots@Home's participation in previous years, this year's development features significant changes and upgrades in software, hardware, and mechanical systems. In the software domain, our most important changes involve natural language processing, manipulation algorithms – stemming from our research and development of robotic arms – and task orchestration methods. A new and improved NLP (Natural Language Processing) node is under development, using the RASA Natural Language Understanding(NLU) software. The orchestration system should allow for automatic switching of nodes, in order to allow the robot to execute navigation, voice processing, visual processing, or any other algorithms, in a seamless fashion. Our hardware and mechanical systems have seen notable improvements in navigation and manipulation domains, with a new robotic arm and mobile base being built from scratch to enhance our control over interactions with the world.**

Our team's approach aims to optimize resource expenses and assist future developments by maximizing the use of affordable and readily available hardware through hardware coupling and custom yet modular software solutions.

***Index Terms*—RoboCup Brazil, Robot Manipulation, Task Orchestration, Natural Language Processing, Robot Vision, Robot Voice, State Machine, Behavior Tree, Robotic Arm, Mobile Base, Emotional Interface, Object Detection, Object Position Estimation, Face Recognition, Robotic Manipulation.**

1 - DAINF(Academic Department of Informatics), UTFPR(Federal University of Technology – Paraná), Curitiba, PR, Brasil, 2 - DAELN (Academic Department of Electronics), UTFPR, Curitiba, PR, Brasil, 3 - IFPR (Federal Institute of Paraná), São José dos Pinhais, PR, Brasil

I. Overview

The UTBots@Home[1] team has been dedicated to the ongoing development of service robots designed to assist humans in household tasks since 2014. Driven by the challenges presented by the Robocup@Home competition [1], our efforts have advanced in both complexity and abstraction. Initially focused on basic control of robotic systems, we now integrate and orchestrate our hardware and software systems at a level of abstraction suited to the ontology of a household environment in line with the competition's characteristics.

Over the years, the competition has served as a platform for numerous academic research endeavors at UTFPR. Our research involves, but is not limited to, navigation with fuzzy logic [2] [3], computational intelligence frameworks [4], human-robot emotional interface [5], fast methods for dataset capture for object detection and manipulation [6], methods for estimating an object's representative 3D point in space [7], and new methods for people recognition [8].

In the domain of software, our current work is primarily focused on Natural Language Understanding, Computer Vision, and task orchestration techniques involving state machines and behavior trees.

Additionally, several upgrades to our hardware are in development. To achieve more control over our robot, the team is developing a new mobile base from scratch, utilizing motors

[1]HomePage of the Team: https://laser.dainf.ct.utfpr.edu.br/doku.php?id=utbots_at_home

and controllers from a hoverboard, inspired by the SHARK approach [9]. Furthermore, we are building a 5 DOF robotic arm with brushless motors and open-source position drivers, coupled with magnetic encoders. Considering the many years of developing new capabilities, an opportunity arose to develop software to synchronize with a complex task orchestration system.

This Team Description Paper (TDP) provides an overview of the team's current solutions and ongoing research for various challenges in the competition. Section 2 provides a description of the innovations the team has made for this year's participation. Section 3 focuses on the current solutions we apply for several components of our robot and the tasks it has to perform. Section 4 concludes and offers insights into future perspectives.

II. Innovation

A. Software

1) Robot Vision:

- **Object detection and identification:** The upgrade from YOLOv3 [10] to YOLOv8 [11] provided improvements in recognition quality, ease of training datasets and usage. All with a significant increase on accuracy and a minor hardware resource consumption. Object segmentation and tracking are also a part of our research in the improvement of 3D object pose estimation.

2) Voice and Natural Language Processing:

- **NLU**: A new system of Natural Language Understanding gives our robot more possibilities of gathering information from a dialog with an operator. RASA is an open-source tool built on a AI model for chatbots [12]. It allows the detection of intentions in a text and can give responses, trigger actions and inform entities of a class involved in a dialog – such as an object's name or a room's name. Research on language models is also on our radar, given the new challenges involving answering open questions about common knowledge.

3) Task Orchestration: Two orchestration systems have been developed in an attempt to compare them and proceed with the most promising one.

- **State Machine:** Software designers have used finite state machines from a long time now [13]. Its ease of implementation into a system coupled with the ease of designing new states for simple tasks made them an attractive choice for simpler and unreactive tasks.
- **Behavior Tree:** First developed for controlling non-playable characters in video games [14], behavior trees offer much more modularity than traditional state machine solutions. Additionally, it offers reactivity without an additional complexity increase [15]. For the advantages it presents there are also trade-offs, because developing such systems present challenges in implementation, not to mention designing trees to solve complex tasks.

  One of the advantages of behavior trees over state machines is that all previous nodes on the tree are ticked constantly, allowing for reactivity without complex systems transitions. This, however, creates the necessity for a different planning strategy which is more complex than that of state machine system. A simple system which solves a level one task is present in figure 1, with both State Machine (upper part) and Behavior Trees (lower part of the Figure). In our experience, BTs become more and more interesting when the complexity of the tasks increase.

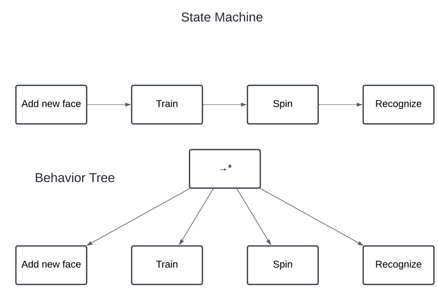

Fig. 1. On the top, a simple State Machine which solves the face recognition task. On the bottom, a Behavior Tree which solves the same task.

4) Navigation System: To address the new hardware of the mobile base, a new navigation system has been implemented. Given that this system is a more isolated component compared to others, and as ROS1 support is coming to an end, we have integrated the hoverboard driver [16] used for differential control of the motors with ROS2 Humble, this being our first package implemented using this version of ROS. The communication between ROS2 and ROS1 packages will be implemented using the ros_bridge package [17].

B. Mechanical and Electrical Subsystems

This year marks a major redesign of the robot since we began participating in the competition. The key elements are mostly related, although not limited to, a completely new mobile base under development:

- **New Mobile Base:** Inspired by the SHARK project [9], a cost effective decision was to use hoverboard brushless motors and hijack its controller board, flashed with a firmware that implements Field Oriented Control [18]. On top of that, we used aluminium extrusions coupled with metal plates to hold mechanical and electronics systems, and a custom suspension system for the wheels.
- **New Arm:** Designed to uphold up to 1.5kg, the new arm has 5 degrees of freedom (DOF) and is built of aluminum, 3D printing, outrunner brushless motors, cycloidal 25:1 gearboxes, magnetic encoders and Odrive [19] position control drivers.
- **Battery System:** The Lead Acid batteries were replaced by Li-ion 10s2p batteries, the ones used for hoverboards. The new batteries are smaller, lighter, and have a better energy density. Researches for a new Battery Management System are being made, considering the power and

voltage necessity of our hardware systems, in an effort to reduce unnecessary wastes of energy.

C. Robot Version Comparison

Figure 2 presents our robot changes from 2022 to 2023. Our retrofitted industrial robotic arm was changed for the a new robotic arm made with 3D printing and stepper motors and the robot structure was improved with aluminum plates and extrusions. Our 2024 robot is being built and will be ready soon for this year's competition, with a new mobile base made with brushless motors, a hoverboard controller and a new robotic arm with position-controlled brushless motors and magnetic encoders, all with a structure made from aluminum extrusions and plates.

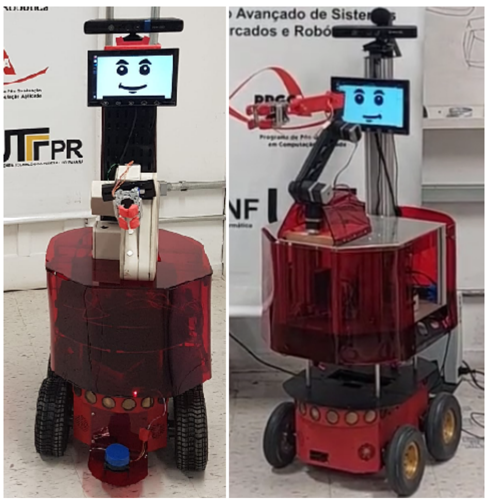

Fig. 2. On the left, our robot in 2022. On the right, our robot in 2023.

III. Current State of Solutions

A. Voice and Language Processing

Interaction with operators occur mostly through verbal communication. Our voice and language processing systems allow the capture of audio, transcription of speech into text, understanding of operator intentions and commands in a variable vocabulary and feedback communication with text-to-speech modules, all integrated through ROS1. First, our integration of *Silero Voice Audio Detection* captures audio and detects when human voices are heard. The human voice audio is then sent to a package that implements the *Whisper* Speech-To-Text real-time transcription. Applying the open-source Natural Language Understanding tool RASA [12], we can detect intentions inside variable phrases that can trigger verbal responses, such as answering questions, or certain commands. Finally, in the realm of speech synthesis, a ROS node harnesses the *Mimic3* speech synthesis engine, seamlessly integrated into the robot's system for natural sound production with human-like intonation, cadence, and emotions, elevating user experience and fostering deeper engagement with robotic systems.

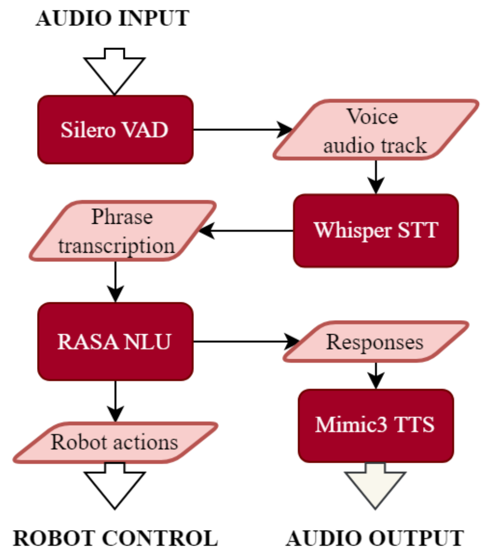

Fig. 3. Voice and language processing systems diagram

B. Emotion Interface

Motivated by a interest of making human-robot interactions in the domestic environment more human, our team has two main emotion interfaces integrated. Our package *display_emotions* [20] implements a facial interface with smooth emotions transitions based on Plutchik's Wheel of Emotions [21], selectable by ROS topics. An RGB LED heart, 3D printed, compatible with the former package, changes color according to the emotion selected.

C. Robot Vision

A Microsoft Kinect v1 [22] offers visual and depth information essential for generating point clouds and enabling various functionalities such as object recognition, person detection, tracking, pose estimation and face recognition.

1) Object Detection: We apply the YOLOv8 neural network [11], performing a transfer-learning training with the new competition objects each year. The object detector outputs a bounding box for a certain object or a person in a RGB image as shown in Figure 4, that are used in several tasks, as well as by other vision systems.

2) 3D Object Position Estimation: For manipulation tasks, estimating the object's 3D position is crucial. By using RGB-D images with an object detector, distance information for a detected object can be obtained. Matching the RGB and Depth images allows the extraction of pixel distances. Applying a statistical method to these distances provides a unique value, which is then transformed into a 3D Cartesian point through trigonometric calculations, yielding the object's 3D location relative to the robot. This method aims to reduce sensor errors from the Microsoft Kinect, enhancing the accuracy and precision of object manipulation. Different statistics were tested, showing a maximum error of 7.6 cm [7].

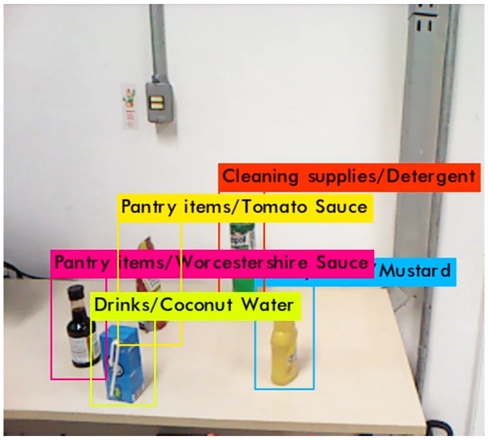

Fig. 4. Objects being recognized by YOLO.

3) Face Recognition: In order to solve personal recognition tasks, a K-nearest neighbours (KNN) algorithm was chosen. This algorithm is divided in 3 phases, image gathering, training and recognition.

During software development our goal was to keep recognition percentages high enough to be usable while reducing training times to fit into a 1(one) minute limit on the computer embedded in the robot, which is considered to be a plausible time within a household environment. For this reason one of the most studied topics was maximizing the number of training images, as well as extracting the most out of each image.

Figure 5 shows a user being recognized after a capture and training [8], that allows the introduction of up to 10 new users with the processing time being less than 60 seconds(1 minute). As future work, we intend to use machine learning models to integrate the newly presented faces to a base of "known" persons, so that the robot can recognize both the "already known" users and the "newly introduced" users seamlessly.

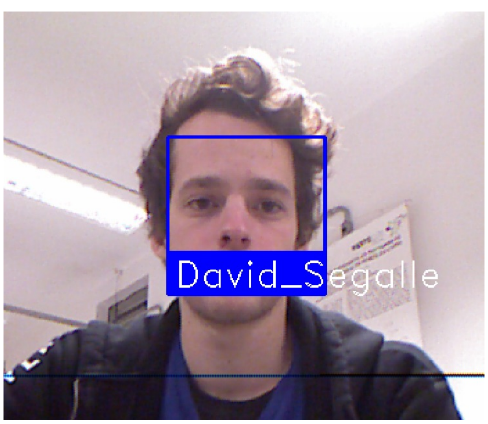

Fig. 5. Voice and language processing systems diagram

While the base system and algorithms remain the same, additions were made to allow for the training with multiple datasets of people, as well as performing the recognition on different models as needed. Moreover, in order to avoid common KNN issues such as overfitting the system has been made resistant against multiple detections of one person as well as recognizing unwanted users.

D. Manipulation

Following our strive for more control over our hardware, our approaches in manipulation focuses around cost-effective strategies for building robotic arms from scratch. As a proof of concept, we first built a 3D printed modified iteration of an open-source arm [23] that integrates larger size and high-reduction gearboxes with stepper motors in 4 DOF (Degrees of Freedom) and has Arduino-based control with RAMPS 1.4 and Pololu drivers, all strategically chosen for their affordability and widespread availability, balancing cost and efficiency.

Using the aforementioned robotic arm as a simpler, less refined solution for testing our manipulation algorithms, we also started the build of a more reliable and controllable robotic arm, with 5 DOF. It is a modular robotic manipulator designed for relatively heavy loads up to 1.5kg. Being composed of aluminum and 3D printed plastic, it supports ample customization. Outrunner brushless motors were chosen to drive most of the arm for their superior torque output even in still position. The gear reduction is provided by a custom designed 25:1 cycloidal gearbox, lowering the RPM increasing the torque by 25x. Each motor has a magnetic encoder for feedback, and the control is made with open-source position control drivers, the ODrive [19].

E. Navigation

Since the first participation of the team, we have been dependant on commercial robotic mobile bases for our robot's navigation. Our two mobile bases, a Pioneer 3-AT [24] and a Pionner LX [25], are complete solutions easy to use for carrying the heavy load of a complete robot, integrated odometry with other sensors and autonomous navigation algorithms with ROS. However, commercial mobile bases have a relevant disadvantage for the context of long-term and low-resource development: maintenance and support expenses.

For the first time, we have pivoted for a solution made from scratch, inspired by the SHARK project [9]. Building the robot's structure with aluminum extrusions, we couple, with a suspension, two brushless motors recycled from a hoverboard. The motors control is made with the original control board, flashed with a firmware that implements Field Oriented Control [18]. This firmware gives access to the control board from ROS topics and enables odometry and autonomous navigation. The navigation system that allows access of the new firmware from our ROS systems was implemented with ROS2 Humble, and connected to ROS1 with ros_bridge [17].

F. Integration and Task Orchestration

All our hardware and software systems are integrated with ROS1 and ROS2 across a number of computing devices: a Intel Nuc [26], a NVIDIA Jetson Nano [27], a Dell Inspiron 15 notebook and microcontrollers such as Arduino Mega, ESP32 and Raspberry Pi 3B+. The first three devices have a Ubuntu 20.04 OS and communicate in a ROS1 network with ROS2 modules connected via ros1_bridge [17]. All the remaining have a connection with ROS as well, with, for instance, serial communication.

Up to 2023, our task orchestration was mainly based on directly calling required nodes through topics. In our current research endeavour, we intend to analyse state machines and their simplicity of design and coding against behavior trees, which offer a simpler logic for achieving reactivity.

IV. Conclusion and Future Work

The UTBots@Home team has demonstrated the ability to successfully accomplish several core tasks in the Robocup@Home category and has made rapid progress in developing more complex functionalities and robot behavior control in recent years. Noteworthy developments include emotion simulation, task orchestration, voice and language processing, a new manipulation and base which are innovative contributions to the competition. Ongoing projects offer promising insights into the team's future evolution, such as the construction of a new robust mobile base and all new task orchestration system, both contributing to advancing tasks that remain challenging for most teams. Additionally, innovations in voice and language processing opens up new horizons for a more sophisticated human-robot interaction, with research being made in the realm of language models as well. The team strives to expand the robot's capabilities, exploring innovative frontiers to advance research across various fields of knowledge. An important a goal is to bring credibility to Latin America in the realm of service robotics and technology as a whole.

Acknowledgment

The authors would like to thank LASER (Advanced Laboratory of Embedded Systems and Robotics), UTFPR (Federal University of Technology – Paraná) and the PPGCA (Graduate Program in Advanced Computing) for their support and resources, which have contributed significantly to the successful completion of this endeavor.

Additionally we would also like to thank GranoZen and TRACTIAN for their financial support.

References

[1] Robocup@home official homepage. http://www.robocupathome.org. Accessed Nov. 2023.

[2] J. Cerbaro, D. Martinelli, A. S. Oliveira, and J. A. Fabro. Waiterbot: Comparison of fuzzy logic approaches for obstacle avoidance in dynamic unmapped environments using a laser scanning system (lidar). In *2020 Latin American Robotics Symposium (LARS), 2020 Brazilian Symposium on Robotics (SBR) and 2020 Workshop on Robotics in Education (WRE)*, 2020.

[3] A. L. Sordi Filho, L. P. Oliveira, A. S. Oliveira, J. A. Fabro, and M. A. Wehrmeister. Simultaneous navigation and mapping in an autonomous vehicle based on fuzzy logic. In Nadia Nedjah, Heitor Silvério Lopes, and Luiza de Macedo Mourelle, editors, *Studies in Computational Intelligence*, volume 664, pages 53–68. Springer International Publishing, 1 edition, 2017.

[4] T. Becker, A. S. Oliveira, J. A. Fabro, and R. L. Guimaraes. Lida bridge: A ros interface to the lida (learning intelligent distribution agent) framework. In Anis Koubaa, editor, *Studies in Computational Intelligence*, volume 625, pages 703–728. Springer International Publishing, 2016.

[5] D. Santos, P. Palar, A. S. Oliveira, J. A. Fabro, and T. Becker. Adding conscious aspects and simulated emotions through facial expressions in virtual robot navigation with baars-franklins cognitive architecture. In *2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE)*, João Pessoa-PB, 2018.

[6] J. Fabro, M. Vaz, and A. Oliveira. Design and development of an automated system for creation of image datasets intended to allow object identification and grasping by service robots. In *January 2020*, pages 1–6.

[7] G. F. Armênio, J. A. Fabro, R. R. Tognella, F. P. Conter, M. V. Oliveira, and E. S. Silva. Estimating the 3d center of an object with kinect sensor rgb-d images. In *Latin American Robotics Symposium (LARS), XX Simpósio Latino Americano de Robótica (LARS 2023)*, Salvador, Bahia, Brazil, 2023.

[8] D. Segalle, G. F. Armênio, G. F Lewin, E. S. Silva, and J. A. Fabro. Identifying new persons in the context of the robocup@home competition using knn. In *Latin American Robotics Symposium (LARS), XX Simpósio Latino Americano de Robótica (LARS 2023)*, Salvador, Bahia, Brazil, 2023.

[9] Nicolas Freitas Dias, Jardel dos Santos Dyonisio, Luis Milczarek Quadros, Gabriel Amaral Dorneles, Marina Zanotta Rocha, Murilo Castanheira Bicho, João Francisco de Souza Santos Lemos, Igor Pardo Maurell, Jardel Prado, Henrique Bezerra, Cassio Teixeira, Rodrigo da Silva Guerra, and Paulo Lilles Jorge Drews Junior. Shark: Stable hoverboard-driven autonomous robot kit. In *BRAHUR/BRASERO 2024 - Brazilian Workshops on Humanoid and Service Robotics 2024, Curitiba/PR, Brazil, April 25-27*, 2024.

[10] J. Redmon and A. Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

[11] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics YOLO, January 2023.

[12] Open source natural language processing (nlp). https://rasa.community/open-source-nlu-nlp/. Accessed Jul., 2023.

[13] Adamczyk P. The anthology of the finite state machine design patterns. https://hillside.net/plop/plop/plop2003/Papers/Adamczyk-State-Machine.pdf. Accessed Jul., 2024.

[14] Isla D. Halo 3-building a better battle. https://web.cs.wpi.edu/~rich/courses/imgd4000-b12/lectures/halo3.pdf. Accessed Jul., 2024.

[15] Collendanchise M. Ögren P. Behavior trees in robotics and ai. https://arxiv.org/abs/1709.00084. Accessed Jul., 2024.

[16] Hoverboard Robotics. hoverboard-driver. https://github.com/hoverboard-robotics/hoverboard-driver/tree/humble, 2021.

[17] ROS 2. ros1_bridge. https://github.com/ros2/ros1_bridge, 2021.

[18] E Feru. hoverboard-firmware-hack-foc. https://github.com/EFeru/hoverboard-firmware-hack-FOC, 2019.

[19] Oskar Weigl and Joseph Morris. Odrivehardware. https://github.com/odriverobotics/ODriveHardware, 2018.

[20] P. S. Palar. Ros package ros_display_emotions. https://github.com/kpiatan/ros_display_emotions. Accessed Jun., 2023.

[21] R. Plutchik and H. Kellerman. *Emotion: Theory, Research and Experience*. Academic Press, New York, 1986.

[22] Develop network: Kinect sensor. http://msdn.microsoft.com/en-us/library/hh438998.aspx. Accessed Jun., 2023.

[23] P. Zhihui. Dummy-robot: Super compact smart robotic-arm. https://github.com/peng-zhihui/Dummy-Robot. (Accessed Jul. 2023).

[24] Adept MobileRobots. *Pioneer 3-AT Operations Manual, Rev. 6.4*. Amherst, NH, USA, 2013.

[25] Adept MobileRobots. *Pioneer LX User's Guide, Rev. D*. Amherst, NH, USA, 2017.

[26] Intel. What is intel® nuc? https://www.intel.com.br/content/www/br/pt/products/details/nuc.html. Accessed Jul. 2023.

[27] NVIDIA. Jetson nano developer kit. https://developer.nvidia.com/embedded/jetson-nano-developer-kit. Accessed Jul., 2023.