UTBots 2024 Team Description Paper

João A. Fabro Marlon O. Vaz Bruno B. Silva
Gustavo F. Lewin João H. C. Pires Gustavo F. Armênio
David Segalle Enzo H. Gaio Victor B. Errera
Larissa B. L. da Silva Luiz G. S. Greca André S. de Oliveira

November 28, 2023

Abstract. UTBots at Home spearheads a multifaceted research initiative aimed at pushing the boundaries of robotic capabilities. Our diverse research pursuits span across various domains, seeking advancements in navigation, voice recognition and synthesis, emotion simulation, robot vision, object identification and localization and manipulation. Rooted in the need for comprehensive robotic functionality, our endeavors encompass a holistic approach, addressing distinct yet interconnected aspects of robotic technology.

Our project is based on 2 primary approaches. Firstly extracting the most out of cheap and easily available hardware through hardware coupling as well as custom yet modular software solutions. Secondly, we aim at creating cheap and modular solutions which can be used within and outside the confines of our laboratory and it's project.

Multiple facets of robotics are being researched by UTBots at Home, within them papers on face recognition, as well as object localization have been published this year on the Brazilian and Latin American Robotics Symposium. Other solutions were and are being developed, especially in the realm of mobile robot bases, arm manipulation and voice recognition.

1 Introduction

Since 2014, the UTBots@Home[1] team has been dedicated to the ongoing development of service robots designed to assist humans in household tasks. Our efforts are driven by the challenges presented by the Robocup@Home competition [1] and extend beyond its scope, in a variety of fields such as Human-Robot Interaction and Cooperation, Navigation and Mapping in Dynamic Environments, Computer Vision and Object Recognition under Natural Light Conditions, Object Manipulation, Adaptive Behaviors and Environmental Intelligence.

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

Over the years, the competition has served as a platform for numerous academic research endeavors at UTFPR. Master's and doctoral projects have played a pivotal role in enhancing the capabilities of the team's robots and pushing the boundaries of their respective research domains.

Current research and advances focuses on various topics, including but not limited to enhancing new object learning, striving for more precise 3D point estimation, leveraging natural language processing (NLP) to enhance human-robot interaction, the development of a new and robust robotic arm, and developing adaptive behavior for complex tasks.

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 approaches developed by the team. Section 4 focuses on the team's current research endeavors. Finally, section 5 concludes with a summary and offers insights into future perspectives.

2 Voice Recognition and Synthesis

For speech recognition within the Robot's system, a custom package named *whisper_cpp_ros* [2]was developed. This package integrates cutting-edge technologies, employing the *Silero Voice Activity Detection* (VAD) program and a C++ implementation of Whisper. The VAD node efficiently detects voice activity in audio streams, optimizing processing resources for relevant speech segments. Meanwhile, the integration of *whisper.cpp* enables automatic speech recognition and real-time transcription, bolstering the system's capabilities.

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. Leveraging ROS functionalities, this program interprets textual inputs and skillfully synthesizes them into lifelike speech. The sophisticated algorithms within *Mimic3* emulate human-like intonation, cadence, and emotions, elevating user experience and fostering deeper engagement with robotic systems.

3 Robot Vision

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

3.1 3D Object Position Estimation

For manipulation tasks, it is essential to estimate the object 3D position. RGB-D images, used alongside an object detector, can provide distance information of a certain detected object. YOLOv3 neural network [2] outputs a bounding box for a certain object in a RGB image. Matching the RGB and Depth images means

[2] Available at: https://github.com/UtBotsAtHome-UTFPR/whisper_cpp_ros

that the distances of the pixels of an object can be extracted. A statistic method applied on the multiple distances of the frame results in a unique distance value for the object and, transformed in a 3D cartesian point through trigonometric calculations, a 3D location in relation to the robot. This method attempts to minimize sensoring errors for the Microsoft Kinect and allow a more accurate and precise object manipulation. Different statistics were tested and the best one showed a maximum error of 7.6 cm. A paper [3] was written on the subject and got published into the Latin American Robotics Symposium as well as the Brazilian Robotics Symposium.

3.2 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 minute limit on portable, 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 1 shows a user being recognized after a capture and training. A paper was written on the subject and got published into the Latin American Robotics Symposium as well as the Brazilian Robotics Symposium[4].

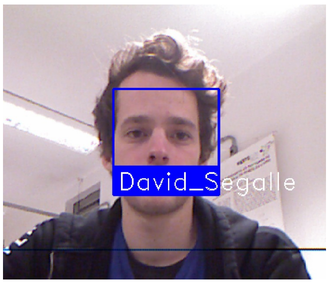

Fig. 1. Image showing a user being recognized and his position displayed in the image.

4 Manipulation

We've undertaken a diverse approach to manipulator research, focusing on two key objectives. Our primary concern involves tech recycling, targeting the issue of improper discarding of robotics and electronics parts. Salvaging a discarded Beckman Coulter ORCA [5] planar manipulator (Figure 2), we repurposed it into a functional all-purpose manipulator. Crafting a new Arduino-based Open Source control unit and integrating a simple off-the-shelf claw exemplifies our commitment to refurbishing technology.

Our secondary focus revolves around cost-effective strategies. Prioritizing simplicity, a modified iteration of an open-source arm [**?**] integrates larger size and high-reduction 3D printed gearboxes, balancing cost and efficiency. Employing trigonometry-based inverse kinematics, our adaptable code ensures easy replication for educational purposes. The latest manipulator integrates Arduino-based solutions with RAMPS 1.4 and Pololu drivers, strategically chosen for their affordability and widespread availability.

5 Other relevant contributions

5.1 Robot Navigation

In order to perform the navigation of the robot in complex environments, we have two mobile bases available, Apollo and Zeus. Both can be connected easily to the ROS network.

In robot navigation, odometry errors can lead to substantial uncertainty. It is important that the robot is able to correct its location based on the feedback from sensors in real time. SLAM (Simultaneous Localization and Mapping) [6] algorithms can achieve this. Currently, the LiDAR sensor readings and the robot's wheels odometry are fed to the standard ROS navigation stack and *move_ base* is utilized to send navigation tasks to the robot.

By constructing a map of the environment at the same time as it is updating the robot's position, the robot estimates it's position and updates wheel velocity. To accomplish this, SLAM has a number of tasks: extraction of reference points, data association, state estimation, status update, and reference point update. Parameters adjustments and tests under different SLAM configurations are crucial for improving navigation.

6 Current Research

6.1 Rapid object RGB-D dataset gathering

The classes detected and the accuracy of the detection depends on the number and quality of labeled images of the dataset used for training. To improve the dataset quality, an external apparatus allows for capturing of a large number of images of each object has been developed [7]. This system has two cameras (stereovision) and a depth sensor, which allows the capture of many rgb and point cloud images, each one viewed by a slightly different angle and distance.

Therefore, it is possible to generate a large dataset of every object, as well as a general 3D format of it. This information can be used to quickly add a new object to the detector and the 3D format can be of use to fine-tune the estimation of a 3D point for object manipulation.

6.2 Natural Language Processing and Understanding

In order to improve the human-robot interactions, it is valuable to increase the robot's interpretation of speech. In this context, the use of Natural Language Processing (NLP) comes in handy, with the capability of extracting intention and turning information into entities. The use of such functionalities overcomes the limitations of standard phrases and commands and makes the robot capable of processing synonyms and distinct phrase constructions for a given context. To implement such capabilities in the robot's interactions, research and development is being conducted with Rasa [8], an open source solution based on machine learning commonly used to implement chatbots.

6.3 New Robotic Arm

A new set of all-metal arms is currently under development. With less of a cost reduction oriented approach. Switching over to quaternion-based kinematics and arm positioning driven by AI. Modular exchangeable claws will also be implemented for greater versatility and better match our object recognition AI.

7 Conclusion and Future Work

The 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, person tracking, and 3D pose estimation with, which are innovative contributions to the competition. Ongoing projects offer promising insights into the team's future evolution, such as the contraption for RGB-D dataset generation of objects and a new robust arm, both contributing to advancing object manipulation tasks that remain challenging for most teams. Additionally, the use of NLP with Rasa opens up new horizons for a more sophisticated human-robot interaction. 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.

References

1. Robocup@home official homepage. `http://www.robocupathome.org`. Accessed Nov. 2023.
2. J. Redmon and A. Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.
3. 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.
4. 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.
5. Beckman coulter orca manipulator. `https://americanlaboratorytrading.com/lab-equipment-products/beckman-coulter-orca-control-module-with-robotic-arm-13916`. Accessed Jul., 2023.
6. S. S. Riisgaard and G. F. Blas. Slam for dummies. `http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-412j-cognitive-robotics-spring-2005/projects/1aslam_blas_repo.pdf`. Accessed Jun., 2023.

7. 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.
8. Open source natural language processing (nlp). https://rasa.community/open-source-nlu-nlp/. Accessed Jul., 2023.
9. Adept MobileRobots. *Pioneer LX User's Guide, Rev. D.* Amherst, NH, USA, 2017.
10. YDLIDAR. *X4 Datasheet, 01.13.000000*, 2018.
11. Intel. What is intel® nuc? https://www.intel.com.br/content/www/br/pt/products/details/nuc.html. Accessed Jul. 2023.
12. NVIDIA. Jetson nano developer kit. https://developer.nvidia.com/embedded/jetson-nano-developer-kit. Accessed Jul., 2023.
13. Develop network: Kinect sensor. http://msdn.microsoft.com/en-us/library/hh438998.aspx. Accessed Jun., 2023.

Apollo and Zeus Hardware Description

Specifications for the robots Apollo and Zeus are as follows:

- Apollo Base: this base weights only 12 Kg, with 12 Kg of payload, and maximum speed of 0.7 m/s. Its autonomy is of 2 hours, with 3, 12 volts, standard batteries. Its dimensions are: 50 cm wide, 50 cm deep and 28 cm high.
- Zeus Base: a Pioneer LX model [9], that weights 60 kg, can carry up to 60 kg of payload, travel at a maximum speed of 1.8 m/s, and has a autonomy of 13 hours of continuous operation.
  The dimensions of the Pioneer LX robot are:
  50 cm wide, 70 cm deep and 45 cm high.
- LIDAR: a YDLIDAR X4 sensor[10] for Apollo and a SICK LMS300 laser range finder for Zeus.
- Computing: an Intel Nuc [11] computer and a Jetson Nano [12] board.
- Arm: ORCA Beckman Coulter Robotic Arm, controlled by 2 Arduino MEGAs.
- Neck: a LCD screen, a microphone, a speaker and a Microsoft Kinect v1 [13].
- Power Source: 5 12V 9Ah Batteries
- Robot weight: 40kg for Apollo and 60kg for Zeus.

Fig. 2. Robot Apollo.

Robot's Software Description

For our robot we are using the following software:

- Platform: Ubuntu 20.04 for Intel Nuc and Jetson Nano.
- Navigation: ydlidar, move_base and navigation_stack packages.
- Face recognition: KNN integrated with our own implementation of data acquisition and online training of the model.
- Speech recognition: whisper.cpp, with our package whisper_cpp_ros.
- Speech generation: Mimic3 synthesis engine.
- Object recognition: YOLOv3.
- Body pose detection: Mediapipe Pose, with our package mediapipe_track.
- Emotion Interface: our package ros_display_emotions.

Robot software and hardware specification sheet