LAR@Home 2023 Team Description Paper

Tiago Ribeiro, Fernando Gonçalves, Marco Luís, Eduardo Fernandes, Rui Silva, António Dias, Nuno Oliveira, Diogo Martins, Syed Fawad, Gil Lopes, A. Fernando Ribeiro

Laboratório de Automação e Robótica (LAR)
DEI -Universidade do Minho,
Campus de Azurém, 4800-058 Guimarães, Portugal
fernando@dei.uminho.pt
https://lar.dei.uminho.pt/home

Abstract. This paper describes the development of the new RoboCup@Home robot team LAR@Home from the University of Minho, Portugal, for the competition that will take place in Bordeaux, France, in 2023. Main research efforts in this new robot focus essentially on its mechanical and electronic structure, with an anthropomorphic approach, in its location, navigation, and segmentation of objects, and in the ability to interact as naturally as possible with the human being and the surrounding environment.
The 25 years of experience in other RoboCup leagues and the analysis of other teams robots, helped the development of this new robotic platform, especially regarding the locomotion, and the vision system, allowing a more reliable and robust platform here described.

1 Introduction

Minho Team is a very active RoboCup team from the University of Minho, Portugal, which competes in the Middle Size League (MinhoTeam) and @Home League (LAR@Home). The team already participated in RoboCup@Home in 2010, with the MARY robot, however, the project stood still for 7 years and was reborn in 2017 with a new robot called CHARMIE, which stands for Collaborative Home Assistant Robot by Minho Industrial Electronics (**Fig. 1**).
This Team Description Paper is part of the qualifying package for RoboCup 2023, to be held in Bordeaux, France and includes the status of LAR@Home. The main research efforts for this new robot focus on both improvements from the MARY robot, and on adding new technologies and ideas that did not exist in 2010. Since the team members are Master and PhD Students from Industrial Electronics Department, the team focus was essentially on the mechanical and electronic structure, with an anthropomorphic approach; also on the localization, navigation, and object segmentation; and the ability to interact as naturally as possible with the human being and surrounding environment.

The team has contributed to RoboCup since 1999 in the MSL League, starting in the @Home league in 2010. It was the first team to introduce an omnidirectional three-wheel platform on the @Home league and organised and held the @Home Educational

2022 in Guimarães – Portugal along with the European RoboCup Junior competition. This has attracted new people to the league and also new students to the team.

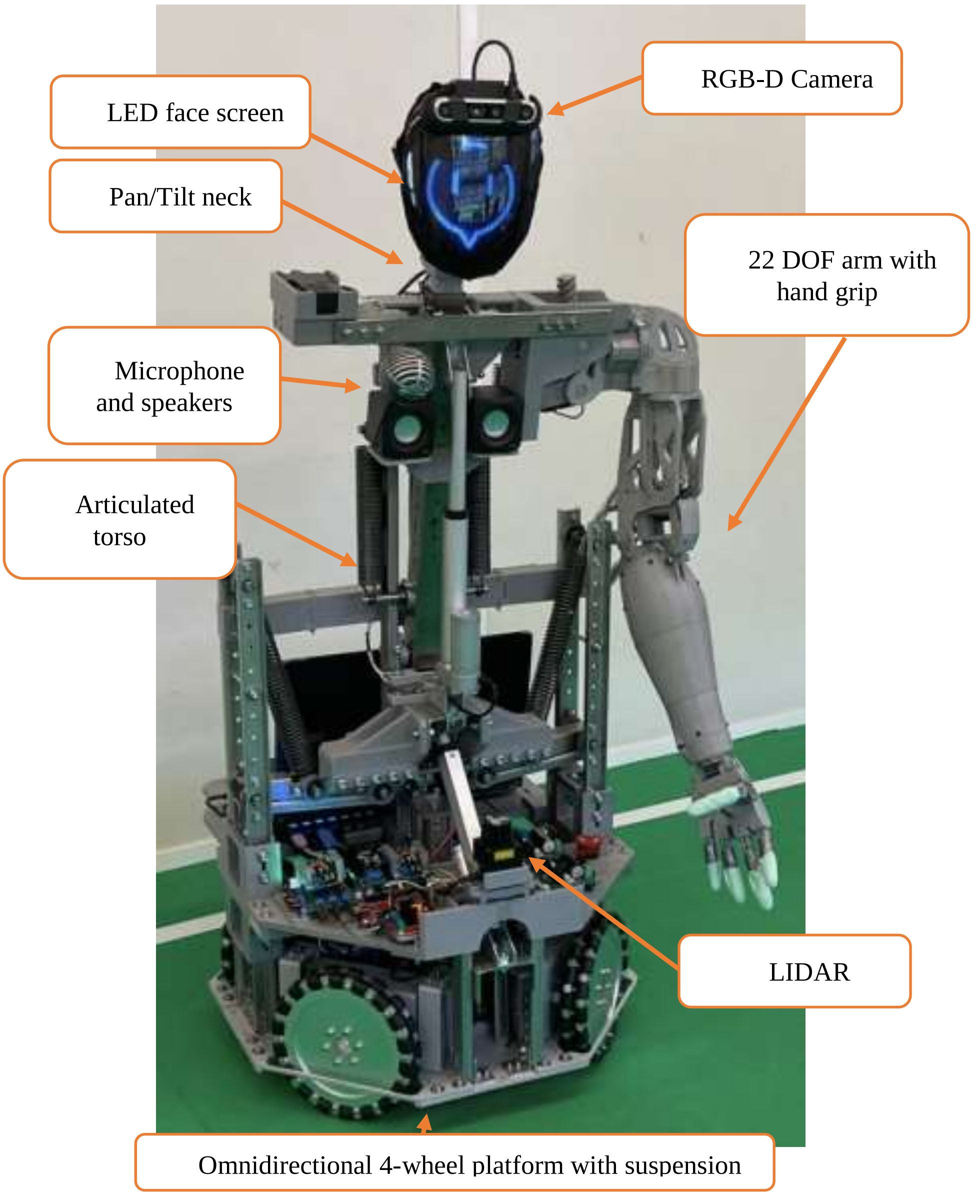


Fig. 1. CHARMIE's actual configuration and structure with a temporary complete left arm.

2 Mechanical Description

The CHARMIE mobile manipulator follows a human-inspired design to improve human-robot interaction [1]. Wheeled locomotion was adopted since it is more efficient for flat indoor locomotion [2]. The base makes use of four omni wheels placed at 90º to improve manoeuvrability, each with an associated independent suspension system to both guarantee continuous contact of all wheels with the floor, as well as to improve the robot's stability.

To increase the robot's manipulation workspace, an articulated torso with two degrees of freedom was developed. This torso allows the robot to squat in a linear motion and flex its trunk forward. Static balancing, and the use of linear actuators which are self-locking, were implemented into both motions to reduce energy consumption [3].

The current iteration of the robot is using a temporary left-handed arm, based on the InMoov project [4] divided into three main parts: (i) hand and forearm; (ii) bicep; and (iii) shoulder. The bicep and shoulder were altered from InMoov's project to adapt to the team's needs. The whole arm is printable on a 12 x 12 x 12 cm 3D, and its PLA parts weigh around 1.414 kg, with the actuators weighing around 0.766 kg. The whole weight of the arm is approximately 2.2 kg. From shoulder to hand, the arm measures 75 cm and can lift a maximum load of around 400 grams. It has 22 DOF moved by 10 actuators.

The next iteration of CHARMIE's upper limbs is currently under development. This new version will include two arms with six degrees of freedom each: 1) shoulder flexion\extension; 2) shoulder abduction\adduction; 3) shoulder internal\external rotation; 4) elbow flexion\extension; 5) elbow supination\pronation; and 6) wrist flexion\extension. In this future iteration, claw-like end-effectors will be used with a single degree of freedom for opening and closing the manipulator.

Two serial degrees of freedom are installed into the robot's neck, controlled by revolute motors, to allow its flexion and rotation.

Taking all listed systems into consideration, CHARMIE has a total of 25 degrees of freedom — three corresponding to the forwards\backwards, lateral, and yaw movements of the locomotion, 18 actuated by 16 revolute motors and two linear actuators, and 4 which are passive, whose configuration is defined by the compression springs for the wheel suspension. This results in the robot having a total weight of 60 kg and a height which varies between 128 cm and 142 cm. Besides the up/down movement, the robot also flex his trunk forward, reaching a minimum height of 92 cm. His trunk and arm can go under structures such as tables as low as 80 cm, allowing the robot to reach objects from the ground that are 60 cm farther from the robot platform.

3 Sensorial System

The section describes the sensorial system, which intends to achieve the following features:

- Semantic 2D SLAM;
- Safe navigation (obstacle detection and avoidance);

- Human-Robot Interaction (user and pose/gesture detection);
- Object detection and subsequent manipulation;
- Vocal interaction with the robot.

3.1 Semantic 2D SLAM

In order to solve the Simultaneous Mapping and Localization problem, CHARMIE is enabled with a 2D Semantic SLAM System. The output allows the creation of goals for the robot that are understandable to both humans as well as the CHARMIE robot, to ease communication and interaction. The system is composed of two main modules:

- a standard 2D LiDAR SLAM solver (i.e. Cartographer by Google [5]) that is responsible for delivering an output of both the map of the environment, through an Occupancy Grid representation and also the robot trajectory during its exploration through the unknown space.
- an Instance Segmentation Neural Network (i.e. YOLACT [6]) that performs object detection and segmentation of a select dataset of objects encountered in domestic environments.

The segmented objects are then added to the occupancy grid in the form of a semantic layer that serves as a way of establishing social goals but also to classify rooms and spaces presented in the domestic environment.

A couple of additional features allow the semantic map to achieve a clear and more exact representation. Firstly, a Monte Carlo Markov Chain is used to predict and then draw the estimated sizes of walls, doorways and resulting rooms, based on the occupancy grid delivered by the Cartographer. Secondly, a post-processing layer is added to the semantic layer to filter out non-coherent depth information from objects as well as fill in in-depth information that the segmentation system is unable to retrieve.

3.2 Robot Navigation

After parameterising all the static and dynamic obstacles in a nearby radius and defining a target location, CHARMIE uses dynamic non-linear systems as a distributed control architecture that generates navigation. Task constraints are component forces that are cast together into the vector field of this dynamical system. Different constraints are represented by repulsive and attractive forces acting on the heading direction. The attractive force attracts the system to the desired heading direction value, whereas the repulsive forces prevent the system from moving in an undesired direction. As the robot moves, the directions to the target and obstacles in the world variates, and consequently, the attractor and repelers also vary.

This solution sorts out the challenge of keeping the robot's front always facing the moving direction. This is a requirement for two reasons: the first is it is easier for humans to percept how the robot will be moving, and the second allows the robot to have all navigation sensors pointing forwards.

3.3 Human-Robot Interaction (user and gesture detection)

One of the biggest concerns regarding this project consisted of providing communication tools to ease interaction with users. To initialise a communication process, the robot must first recognise users and their pose/gestures. The functionalities regarding user detection demonstrate solutions using the RGB-D camera, both with and without the depth image. Initially, this team developed a solution using multi-stage deep learning pipeline using only an RGB camera [7].

When updating to 3D user recognition technology, the developed algorithm consisted of two independent CNNs, an Inception-Resnet V1 to deal with the RGB images and an Inception V3 to deal with depth maps [8]. The depth solution CHARMIE recognises a low number of users using various deep convolutional neural networks, similar to the one with just the RGB that integrates both the RGB and the depth images.

After recognising the users, CHARMIE analyses their poses. Merging an RGB-D camera and generating a 3D view of a person, together with the Mediapipe framework which is an open-source software provided by Google, an algorithm was created that detects and calculates the pose estimation of a person.

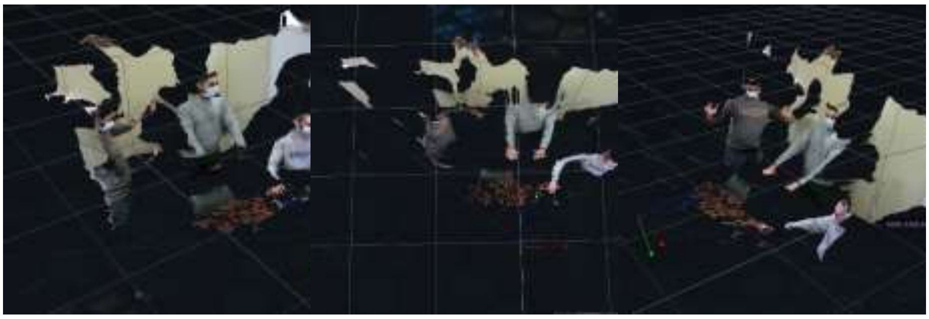
Mediapipe uses a holistic model pipeline that integrates separate models for pose, face and hands. With these separate models, the human pose is detected with BlazePose's pose detector, generating a total of 543 landmarks (33 pose landmarks, 468 face landmarks, and 21 hand landmarks per hand).

Making use of these landmarks, an algorithm was created to detect a person's human pose and it calculates the angles of the arm joints through trigonometric functions, sending the values to the robotic arm. These values are sent in real-time, allowing the robot to mimic the human operator and handle various objects, making it possible to realize specific tasks such as person tracking and follow-me behaviour.

3.4 Safe navigation (obstacle detection and avoidance)

To safely navigate a previously mapped environment, CHARMIE must detect dynamic and static obstacles that were not present in the environment map and navigate accordingly to overcome these new entities.

Regarding obstacle detection, the robot uses the same sensors as for the mapping function, the 2D LiDAR and the RGB-D camera. The 2D LiDAR is used to detect small obstacles at the bottom platform height. It starts by checking if there is any obstacle inside a 1.50 m radius and calculates both the obstacle position relative to the robot and its size. The RGB-D camera starts with the same principle regarding the 1.50 m radius and calculates its size and position relative to the robot, creating a virtual obstacle from the floor to the robot's height. In **Fig. 2**, an example of the same image from three different angles is displayed, showing an example of data fusion between the RGB camera and the Depth camera. The obstacles derived from the sensors are combined in a temporary virtual obstacle map that is constantly updated. By applying this method in consecutive frames, the direction of movement of dynamic obstacles is also added to the virtual obstacle map.



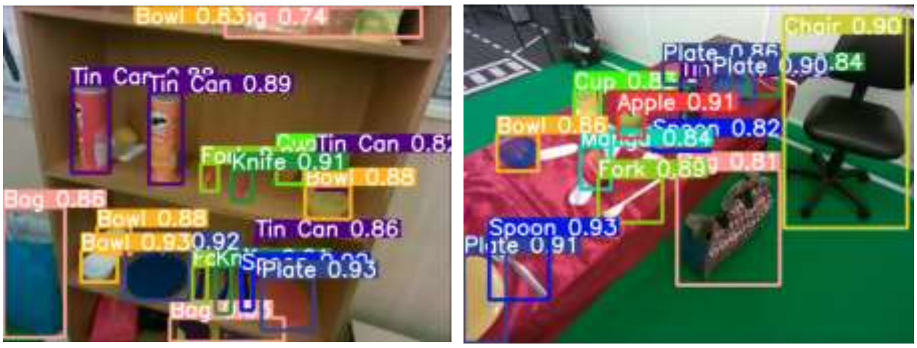
(**a**) (**b**) (**c**)

Fig. 2. Three different perspectives of three human users interacting with CHARMIE. The robot fuses the RGB and Depth cameras' information to create a 3D view of the environment: a) Left side view; b) centred view; c) right side view.

3.5 Object detection and subsequent manipulation

For learning and recognizing objects and household items, CHARMIE uses the YOLOv5-small architecture. To introduce new objects into the database, a video of the item is recorded and later used to retrain the network. **Fig. 3** shows some of the household items that the robot has already learnt to detect, like bottles, cans and bags of chips. The dataset consists of 20 classes, such as, trash can, garbage bags, bags, fruit (apple, pear, banana and mango), cans, forks, knives, spoons, plates, glasses, bags, mugs, bottles, cabinets/shelves, chairs, tables and people.

To grasp the detected objects, the robot's arm is used. Some objects with unusual shapes, such as a bag of chips, whose shape changes have different programmed collection algorithms. Nevertheless, for most of the items, the picking system is similar. After detecting the desired object, the robot calculates the inverse kinematics to place the robot arm in the best position to collect the object. Follows a hand-closing according to the object's physical properties. From the object's initial detection, the robot moves its platform and lifting mechanism to best fit the item's position. The arm is moved right next to the object to manipulate and move the hand right next to it to grab it. In this case, it starts by using the thumb as a back wall and then it closes the fingers one by one from the index finger to the little finger. This sequence of movements allows the robot to grab the can, similar to humans use to grab it.



(**a**) (**b**)

Fig. 3. Two different outputs from the YOLO detection algorithm detect and locate various pre-trained household objects simultaneously in real-time, from the 20 classes in the dataset: a) in a shelf; b) on a table.

3.6 Vocal interaction with the robot

The user can interact with the robot by different means. Vocal interaction is achieved by an Automatic Speech Recognition (ASR) open-source software named Whisper. This ASR is an offline solution, computationally very light and robust, with multiple language support (over 80 languages), enabling the CHARMIE robot to recognise different spoken languages. For the opposite direction, the text-to-speech (TTS) synthesizer package called Festival is used. This is also open-source and produces a good tone and clear voice in English sentences. The robot listens to commands from the ASR system, replying with vocal sentences to the user using the TTS system.

4 Conclusions

This paper presents a brief introduction to the CHARMIE robot. It is actually performing the necessary tasks that allow the team to participate in the RoboCup@Home league. Since the team restarted in 2017 building a new robot from scratch, not only tried to understand the limitation of the previous robot MARY but also the limitations of other team's robots that participate on RoboCup@Home. The introduction of a mechanical suspension, was thought after the difficulties of normal platforms to overcome uneven floors or floor mats. The same can be said by the introduction of 22 cm diameter omnidirectional wheels. A forward-bending torso is another advantageous introduction to this new platform, as this degree of freedom along with the tilting neck, allows CHARMIE to see very close to its platform on the floor to avoid collisions or over-running. It also allows easier litter floor-picking by lowering and forward bending the torso, approximating the arm and hand grip to the floor and obtaining a better image from the RGB-D camera that sits on top of the forehead.

In terms of software, the actual robot implemented a set of different solutions for each need as explained in this TDP paper. From SLAM to object detection and grasping, user and pose recognition and vocal communication, this new robot sets a new paradigm for the whole team that now has a very much improved platform to work with in a more reliable, robust, complete and straightforward way. A detailed explanation of the robot's hardware and software can be found in the extended paper from 2021 [9]. Our public repository for the CHARMIE project is available on: https://github.com/LARobotics

References

1. Fink, J. (2012). Anthropomorphism and Human Likeness in the Design of Robots and Human-Robot Interaction. In Social Robotics: 4th International Conference, ICSR 2012, Chengdu, China (Vol. 7621, pp. 199–208). https://doi.org/10.1007/978-3-642-34103-8_20
2. Bruzzone, L., & Quaglia, G. (2012). Review article: locomotion systems for ground mobile robots in unstructured environments. Mechanical Sciences, 3(2), 49–62. https://doi.org/10.5194/ms-3-49-2012
3. Gonçalves, F., Ribeiro, T., Ribeiro, A.F., Lopes, G., Flores, P. (2022). Dynamic Modeling of a Human-Inspired Robot Based on a Newton-Euler Approach. In: Kecskeméthy, A., Parenti-Castelli, V. (eds) ROMANSY 24 - Robot Design, Dynamics and Control. ROMANSY 2022. CISM International Centre for Mechanical Sciences, vol 606. Springer, Cham. https://doi.org/10.1007/978-3-031-06409-8_8
4. Langevin, G. InMoov—Open Source 3D Printed Life-Size Robot. Available online: https://inmoov.fr/ (accessed on 1August 2021).
5. W. Hess, D. Kohler, H. Rapp, and D. Andor, Real-Time Loop Closure in 2D LIDAR SLAM, in Robotics and Automation (ICRA), 2016 IEEE International Conference on. IEEE, 2016. pp. 1271–1278.
6. D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, “Yolact: Real-time instance segmentation,” arXiv.org, 24-Oct-2019. [Online]. Available: https://arxiv.org/abs/1904.02689v2. [Accessed: 25-Nov-2022].
7. N. Pereira, T. Ribeiro, G. Lopes and A. F. Ribeiro, "Real-Time Multi-Stage Deep Learning Pipeline for Facial Recognition by Service Robots", RoboCup Asia-Pacific 2021 Symposium Aichi Japan, Nov. 2021.
8. S. Baixo, T. Ribeiro, G. Lopes and A. F. Ribeiro, "3D Face Recognition using Inception Networks for Service Robots," 2022 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC), 2022, pp. 47-52, doi: 10.1109/ICARSC55462.2022.9784804.
9. Ribeiro, T., Gonçalves, F., Garcia, I. S., Lopes, G., & Ribeiro, A. F. (2021). CHARMIE: A Collaborative Healthcare and Home Service and Assistant Robot for Elderly Care. Applied Sciences, 11(16), 7248. MDPI AG. Retrieved from http://dx.doi.org/10.3390/app11167248