# RoboFEI@Home Team Description Paper for RoboCup@Home 2020

Rodrigo C. Techi, Bruno F. V. Perez, Fagner Pimentel,
João Gabriel Rosa Amorim, Marina Y. Gonbata, Bruno Gottsfritz Silva,
Thiago S. B. Meyer, William Y. Yaguiu, Airton Stocco de Almeida,
Guilherme Nicolau Marostica, Amanda Freitas Maneira Zaccaro,
Rodrigo Mendes de Souza, Thiago Borges Coronado,
George Sena Campos Martins, Eduardo Miranda Pires,
Flavio Tonidandel, Reinaldo Bianchi, Plinio Thomaz Aquino-Jr.
robofei@fei.edu.br*

FEI University Center, Sao Paulo, Brazil
http://robofei.aquinno.com/athome/

**Abstract.** HERA (Home Environment Robot Assistant) is a robot created by RoboFEI@Home crew. It was designed to perform human-robot interaction and cooperation tasks. HERA counts with a series of sensors to aid in mapping and navigating the environment, as well as recognizing the human silhouette, individuals by their faces and objects and a gripper, designed and manufactured by the team, that allows the robot to interact with the environment. HERA was developed in 2015 as an initiative of the Human-Robot Interaction area in FEI robotics projects. The goal of the project is to support people in homes, such as assisting the tasks of people with low mobility. Throughout the text, the research goals and interests of the team will be presented, as well as the hardware and software stacks used in the development of the robot to solve the tasks of the @Home competition and some of the projects currently under development by the group. The article intends to cooperate with the sharing results on the specification of a service robot.

## 1 Introduction

Service robots consists on an amalgam of software and hardware systems, with the purpose of perform daily human-assisting tasks. To attain it, they have to be able to navigate through an environment, whether known or not. Also, the ability to detect both objects and people is essential. To enable interaction, recognize commands from humans is mandatory, whether by gesture or spoken commands. With the purpose of advance the state-of-the-art in service assistive robotics, the Robot Competition known as RoboCup@Home was created. Thus, to participate in the @Home competition, an autonomous mobile robot is required.

The RoboFEI@Home team created HERA, a robot designed to perform human-robot interaction and cooperation tasks. At its core, HERA features a

* 3972-B Humberto de Alencar Castelo Branco Ave, Sao Paulo, Brazil

Mecanum Wheel robotic platform, with PeopleBot's chest level extension to aid interaction with humans and the environment [1].

HERA also counts with a series of sensors to aid map and navigate the environment, as well as recognize the human silhouette, faces of individuals and objects. A gripper, designed and manufactured by the team, allows the robot to interact with the environment.

This Team Description Paper is organized as follows: section 2 presents RoboFEI@Home research focus and interests, as well as its history in other competitions. section 3 describes the software stack used in the robot to solve the tasks of RoboCup@Home Competition. In section 4, we present the main projects under development by the RoboFEI@Home team, describe how they contribute to the RoboCup community, present their applications in the real world, as well as current results. Finally, section 5 presents the conclusions and future works.

## 2 Research focus and interests

RoboFEI@Home focuses its research in the interaction between human and machine (computers, robots, autonomous cars or smart houses). Research in the field of human-machine interaction is crucial to the advancement of the state of the art in service robotics, in which machines can act together with humans in their daily lives.

The project also makes heavy use of and contributes to the development of methodologies, techniques, models and algorithms in the following topics: adaptive interfaces, brain-computer interfaces; planning; intelligent home and building automation; autonomous systems and the internet of things (IOT).

### 2.1 Team Achievements, Participations and Collaborations

RoboFEI@Home first participation in an @Home competition was during Brazilian Robotics Competition (CBR, from Portuguese *Competição Brasileira de Robótica*) 2015, in which the team ranked third. In 2016, the team made its debut in RoboCup, its first international competition, where it gained experience and maintained contact with other teams. The team also participated in CBR@Home 2016 competition in which took the first place. At CBR@Home 2017 and 2018 the team won first place again. RoboFEI@Home‘s fourth championship came at CBR@Home 2019, remaining so, the Brazilian champion in the @Home category. In the last Brazilian competition 8 teams participated demonstrating excellent league evolution in the region.

Also in 2019, the team participated in the RoboCup@Home competition in Australia, where it was possible to compete with the 10 teams ranked in Open Platform League.

The team participated in various scientific events, demonstrations, workshops, with the aim of publicizing the league locally, and demonstrating to people the importance and potential of social robotics.

In addition to participating, RoboFEI@Home team members are also involved in both CBR@Home and RoboCup@Home tournaments organization. Since 2018 the CBR@Home‘s Chair is a RoboFEI@Home team member, and at 2019, some members became part of both local and international Organizing Committees.

## 3 Description of the approach used to solve RoboCup@Home challenges

HERA uses ROS Melodic Morenia, released in May 2018 [2]. In the next sections, the vision, speech and navigation systems used in the robot will be presented.

### 3.1 Robot Vision

In the RoboFEI@Home project we utilize a single input device, a Microsoft Kinect 2, to achieve face recognition, people detection and object detection. The freenect 2 [3] was used, with the objective of accessing the main functions of the Kinect in several platforms.

The team developed a approach to object detection based on the use of a convolutional neural network and the single-shot detector technique [4]. With this approach, a dataset containing a very large number of examples for each object is necessary, but the neural network is capable of greater generalization while reducing processing time. In this work, we use MobileNet v2 [5] as our neural network architecture of choice. We trained MobileNet v2 with the aid of TensorFlow Object Detection API in a custom made dataset.

The big difference from this year to previous years is in the way the dataset is being generated. Realizing the lack of time for dataset generation in the early days of the competition, the team developed a software for generating synthetic examples of the objects, which yielded satisfactory results for object detection. This work developed for dataset generation will be presented in the projects section of this paper. Objects are detected by the aforementioned method using the video feed from the Kinect camera and are later placed in the 3D environment, with relation to the robot, by finding each object’s corresponding location in the point cloud, provided by the Kinect infrared sensor.

For people detection we are using OpenPose [6], the first real-time multi-person system to jointly detect human body, hand, facial, and foot keypoints on single images. Together we are using Wave! for gesture recognition. The robot performs a face recognition using Haar Cascade.

### 3.2 Voice Recognition

The team decided to use the Google Speech Recognition API. For this, a ROS package was developed which operates by a set of APIs. These are online tools that works directly in Ubuntu. In addition, a comparison with generic phrases is made through the Hamming distance for the recognition of phrases variations.

Our team, developed a usage of this API by researching methods to make the code easier to adapt to a certain environment, creating a new use of words choices in the speech.

### 3.3 Robot Navigation

When the robot is in an unknown location, it must know the environment where it is located, mapping and at the same time defining its position in the space. This technique is known as Simultaneous Localization and Mapping (SLAM). In the navigation, the robot has the capacity to choose the best possible route and avoid possible obstacles. For this to happen, it determinate parameters where there is the slightest path error and the robot is constantly correcting it.

## 4 RoboFEI@Home Projects

This section describes current research and projects under development for the robot.

### 4.1 Learning rules for social navigation

This project is to develop a social navigation model on environments with multiple contexts focused on naturalness, sociability and comfort. Ontological Information (knowledge) is expected to be used in a learning system of social norms and rules in this process as focus of research. The work seeks to accomplish collaboration at the research areas of artificial intelligence, navigation, and social interaction through social navigation.

Navigation problems will be explored on environments of multiple contexts, and varied complexity rating; artificial intelligence, combining advanced pattern recognition and developing intelligence with common sense; human social dynamics understanding and moral rules that can be totally integrated with human social life showing empathy and social behavior; as well as ethics and security precautions.

**Applicability in the real world:** The research of this project is based on the mapping of existing social navigation models to develop a generic and modular model using semantic knowledge through the robot's ontologies of the environment, people and their relationships, to learn social behaviors related to navigation. The advantage of this type of application is that the robot can use existing knowledge about the environment in which it is inserted (people, objects and places) to learn new knowledge (behavior). Using ontology, it will be possible, for example, to learn social norms and rules in simulated environments and to use this learning in real environments.

The project architecture is divided into 3 main layers: perception, learning and performance. The perception layer is responsible for the extraction of information (location, obstacles) from the sensor data, as well as the extraction of knowledge represented by ontology indicating the relationship of people, objects and places, and is also responsible for generating semantic maps that indicate where the robot can act. The acting layer represents the common navigation found in most robotic systems that use mobile social robot. The learning layer is the core of the proposed model, and the differential of the project in relation to other works found in the literature. While other works use social force model, proxemics, among other methods to define how a robot should act close to the human being, this work uses ontological information to learn how to behave in each social context. The model is capable of providing socially accepted behavior regardless of what type of actuator the robot has or the scenario it is inserted into, considering human comfort.

The learning process is divided into the following steps:

- Extraction of information from the environment;
- Extend ontology CORA to suit the needs of this project, such as characteristics of people and concepts of social navigator (comfort, naturalness and sociability);
- Beginning of the supervised training experiment. The experiment is carried out taking into account the degree of comfort analyzed in the people present in the context as a measure of reinforcement of the environment. This training process is applied in two phases. First, the training is performed in a simulated environment, where learning is performed. After that, it is performed in a real environment, where the model is refined to adjust itself to noises and inherent differences of the real world.
- Finally, the model is applied in a real environment. For each trained context the robot will behave in order to avoid discomfort in humans.

### 4.2 Dynamic power management

At each year‘s iteration, the robot‘s energy demand kept rising, reaching a bottleneck point, in with the batteries duration were not being enough to attend the team‘s needs. Aiming to increase the run-time without stopping to switch or recharge the batteries, the work is proposing a research on applying Dynamic Power Management (DPM) to the robot. According to [7] DPM consists on exploring idle system modules. When the idleness is detected, the power consumed by the idle module should be as least as possible.

**Re-usability for other research groups:** The Robocup@Home league demands of each robot the capability to navigate through different environments, and while doing so, it needs to be powered on at all times, wherever they are. At the moment, batteries are essential to make that happen, because it eliminates all wires that make the operation unfeasible. Taking this into consideration, a dynamic power management, which optimize the battery’s charge, is very beneficial to all league teams.

**Applicability in the real world:** As of today, batteries are almost everywhere, and each day their number grows. It powers up houses, cars, smartphones, health devices, and many people are already trying to extend how long these can run without the need to be replaced. Having such technology can help not only prolonging the run-time, but it can also harm less the environment, decreasing the number of discarded batteries.

### 4.3 Speech recognition

Speech is the main communication method used by humans when interacting, also when interacting with robots by using speech recognition. Through speech recognition and speech synthesis humans and robots may communicate and even perform collaborative tasks, however what has been said its not the only information one may extract from its signal. It is possible to recognize emotion in a user speech, this area is called SER (Speech Emotion Recognition) and can be used to create an adaptive interface for human-robot interaction.

**Re-usability for other research groups:** In the RoboCup@Home category, robots usually interacts with people in a home environment simulation and collaborate in tasks, and many times the user's feelings are ignored, but it is important to interact with humans and make them comfortable when talking to the robot so that the interaction may be customized for each and single user, creating new directions for the league tasks.

**Applicability in the real world:** For the real world, SER can be useful when using robots in homes, since understanding the user's feeling may help when adapting the robot into the environment. Robots may scare some people and for a better communication and adaptation, robots can adapt its interaction so that the human feels comfortable. Furthermore, robots using SER can also be used when taking care of people, since it could recognize if a person is sad, scared or even tired, for that an adaptive conceptual interface may be used, so that the robot may model tasks and users for a better usage and task completion.

### 4.4 Synthetic dataset tool

Although the current method works very well for performing object detection, it requires that a large amount of data collecting during the early days of the competition, so there is not enough time for data collection and network training. This way the robot won't be prepared to detect objects when tasks are officially initialized. In this context, this project considers the development of a software for synthetic images generation. This project aims to develop the approach that will be used by the team and it will streamline the process of training object detection models. The tool is in a trial version, once completed the code will be made public available in the team repository.

**Re-usability for other research groups:** In the @Home category robots interact with a home environment and perform tasks that involve moving objects from one location to another. This category aims to develop service robotics and assistance technology with great relevance for future home applications. A lot of research groups develop robots for this category, so this tool will be useful for their work as it will reduce the effort needed to get the necessary data to make their robot vision work and will complement their robotics research.

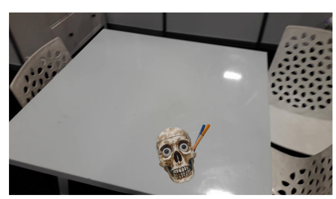
**Applicability in the real world:** This project is relevant to any application that involves deep learning and therefore requires a large amount of data. Object detection has a wide application in various scenarios of our world and with this tool, we hope to contribute to the development of research, facilitating the work of creating datasets.

**Results:** The project was divided into two main parts. The first is a method for automatically extract the background from the sample images of the objects. The second part is a method for using these images and the generated masks to create new images by pasting the objects into different background images collected for the dataset. In the tests it was concluded that, for a good object detection, about 15 photos per object in different perspectives of view are necessary. It is desirable that the images collected to compose the background of the new images be collected in the same environment where the robot will perform its tasks.

An efficient background removal method based on Deep Salient Object Detection [8] has been developed, which is an algorithm for detection of the most noticeable / important object in an image, returning the alpha channel of the image with the object. After mask generation, the object image is used for composing the new images with different backgrounds. In generation it is possible to set some properties as number of times an object will appear in the dataset, maximum rotation of the object in collage, maximum and minimum resizing of the object image and some other things. For the compositions some algorithms are used for standardization of the final image. The process of creating the dataset can be seen in Figure 1

(a) Picture of the object.

(b) Generated mask.



(c) Final image.

Fig. 1: Process of creating the dataset.

## 5 Conclusion

Since 2015, when RoboFEI@Home began competing, the team's research on service robots has improved substantially. Both the software and hardware has been evolving since then. At CBR@Home 2019, where the team won the fourth title, both, the new speech recognition method and the synthetic dataset generation tool for object detection, showed a good performance and were important to win the contest. Research being conducted by the team shows development in areas of interest for home assistance robotics, with some of the projects already showing positive results in competitions. As future works, the aim is to improve the performance in tasks of people and gestures recognition, using new technologies, as well as the development of a modular structure, to facilitate the transport and maintenance of the robot.

## References


1. P. T. Aquino Junior, B. F. V. Perez, D. R. Meneghetti, F. A. M. Pimentel, G. N. Marostica, J. G. R. Amorim, L. C. Neves, L. I. Gazignato, M. Y. Gonbata, R. M. Souza, R. C. Techi, T. S. B. Meyer, V. S. M. Schmiedl, and W. Y. Yaguiu, "Hera: Home environment robot assistant," in *Proceedings of the II Brazilian Humanoid Robot Workshop (BRAHUR) and III Brazilian Workshop on Service Robotics (BRASERO)* (C. U. FEI, ed.), vol. 1 of *1*, (Sao Bernardo do Campo, Sao Paulo, Brasil), pp. 68–73, IEEE South Brazil Robotics Automation Society Chapter, 4 2019.
2. "ROS.org | Powering the world's robots." `http://www.ros.org/`. Accessed on: May, 10h 2019.
3. L. Xiang, F. Echtler, C. Kerl, T. Wiedemeyer, Lars, hanyazou, R. Gordon, F. Facioni, laborer2008, R. Wareham, M. Goldhoorn, alberth, gaborpapp, S. Fuchs, jmtatsch, J. Blake, Federico, H. Jungkurth, Y. Mingze, vinouz, D. Coleman, B. Burns, R. Rawat, S. Mokhov, P. Reynolds, P. Viau, M. Fraissinet-Tachet, Ludique, J. Billingham, and Alistair, "libfreenect2: Release 0.2," Apr. 2016.
4. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*, pp. 21–37, Springer, 2016.
5. A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.
6. Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields," in *arXiv preprint arXiv:1812.08008*, 2018.
7. L. Benini and G. DeMicheli, *Dynamic power management: design techniques and CAD tools*. Springer Science & Business Media, 2012.
8. Q. Hou, M. Cheng, X. Hu, A. Borji, Z. Tu, and P. H. S. Torr, "Deeply supervised salient object detection with short connections," *CoRR*, vol. abs/1611.04849, 2016.

Fig. 2: A picture of our team's robot, HERA.

## Robot Technical Specifications

Hardware Description:

- Base: Mecanum Wheel Robot platform.
  - Sensors:
    - Hokuyo UTM-30LX-EW.
  - Actuators:
    - Omnidirectional wheels.
- Chest: PeopleBot extension.
  - Sensors:
    - Emergency switch.
    - Asus Xtion
  - Actuators:
    - 6 DOF manipulator + 1 DOF gripper.
- Head: Apple Ipad 2.
  - Sensors:
    - Microsoft Kinect 2;
    - Logitech c920 webcam;
    - 2 RODE VideoMic GO directional microphones;
- Control: Zotac Mini-PC i5 7500T CPU.

Software Description:

- OS: Ubuntu 18.04;
- Middleware: ROS Melodic Morenia;
- Localization/Navigation/Mapping: SLAM;
- Face detection: Haar cascades;
- Face recognition: LBP Algorithm;
- People detection and tracking: OpenPose
- Gestures/movement recognition: Wave! and NITE;
- Object recognition: MobileNet v2 + SSD on Synthetic Data;
- Object manipulation: Moveit!;
- Speech recognition: DeepSpeech (offline) or a package based on the Speech Recognition library (online);
- Speech synthesis: Flite (offline) or GTTs (online).