PMEC@Home 2024 Team Description Paper

José R. R. Teles   José G. R. Teles   Alexandre C. F. Filho
Ana L. C. M. Frayne   Carlos E. R. Duarte   Daniel R. da Silva
Eduardo D. Peixoto   Gabriel P. O. Ruotolo   Igor D. Aguiar
João G. da S. Chaveiro   João V. C. Souza   Lavínia R. Lima
Letícia L. Mendes   Luis F. F. de O. Freitas   Lygia M. Rebouças
Marcelo H. L. Ferreira

November 25, 2024

Abstract. This paper describes the service robot Miss Piggy of team Pequi Mecânico that intends on participating in the RoboCup@Home competition which will take place in 2025 in Salvador, Brazil. This competition has influenced the development of research in natural language processing, computer vision, robotic manipulation, simultaneous localization and mapping.

1 Introduction

1.1 The Pequi Mecânico

The Pequi Mecânico Robotics Center, abbreviated as Pequi Mecânico, is a non-profit student organization that brings together students from various undergraduate and postgraduate courses at the Federal University of Goiás. Its aim is to foster interest, research and development of robotics in the academic environment. It is a self-managed group of students from the university, with guidance from teachers and support from scientific institutions linked to the Federal University of Goiás, such as the School of Electrical, Mechanical and Computer Engineering (EMC) and the Institute of Informatics (INF). This, coupled with the support of the Center of Excellence in Artificial Intelligence (CEIA) and the Advanced Knowledge Center for Immersive Technologies (AKCIT) has been fundamental, providing resources and technical guidance that allow the team to overcome complex technical challenges.

Pequi Mecânico Robotic Team exists since 2011 and took part in Latin American and Brazilian Robotics Competition in various categories: IEEE Standard Educational Kit (SEK), IEEE Open, RoboCup Small Size Soccer (F180), IEEE Humanoid Robot Racing (HRR), IEEE Very Small Size Soccer (VSSS) and

RoboCup Soccer Simulation 2D. In 2019 we decided to compete for the first time in the Robocup@Home league.

Service robots are hardware and software systems that assist humans to perform daily tasks in complex environments. In order to achieve this, they have to be able to understand spoken or gesture commands from humans; to avoid static and dynamic obstacles while navigating in known and unknown environments; to recognize and manipulate objects and performing several other tasks that a person might request. Our robot's name is Miss Piggy (see Figure 1).

A view of the robot during development, Figure 1.

Before executing tasks, the Miss Piggy robot maps the environment and tracks objects using Lidar, odometry, and IMU data combined via an Extended Kalman Filter, generating a 2D occupancy map through SLAM. We are evaluating SLAM algorithms for hardware limitations and sensor uncertainties, with RTAB-Map, accessible via ROS \cite{roswikirtabmap}, proving effective for modular testing in virtual and real environments.
For localization, the EKF implementation in ROS \cite{moore2016generalized} integrates wheel odometry, visual odometry, and IMU data for relative positioning, supplemented by Lidar for absolute measurements. ORB-SLAM, adapted from RGB-D applications \cite{murORB2}, enhances visual odometry, while Adaptive Monte Carlo Localization (AMCL), as explored in \cite{thrun2005probabilistic}, addresses map-based localization challenges.

Fig. 1. Side view of the Miss Piggy Robot.

Pequi Mecânico's entry into the @Home category is relatively recent. Our current iteration of the team is built upon a post-pandemic environment where much of the knowledge within our robotics core in the category was lost. We are still in the process of reconstruction, building a culture of development across software, hardware, and structural aspects. Even with significant help from commercial solutions, as will be explored further, it is still a great challenge to create a solid foundation for healthy development and sustain a legacy for the future of Pequi Mecânico and the League.

2 Miss Piggy's Robotic Systems

2.1 Manipulator

The manipulator model that accompanies Miss Piggy is a ViperX 300s 6DoF (Trossen Robotics), consisting of 8 Dynamixel servos (both XM540 and XM430) coupled in a way that 6 degrees of freedom were obtained, with a maximum payload of 750g. The automation approach varies between classic control applications and reinforcement learning algorithms, depending on the desired activity.

2.2 Human Interaction

The primary interaction with the robot is through voice commands, leveraging enhanced Neural Networks optimized for embedded systems and deployed on Jetson Xavier using NVIDIA Riva SDK [1] for Automatic Speech Recognition and Speech Synthesis, delivering state-of-the-art performance. For Natural Language Understanding (NLU), a custom Llama 2 model processes conversational tasks, enabling contextual command execution and assisting the operator in a domestic environment. Additionally, commands can be input through a web-based Graphical User Interface (GUI) accessible via any mobile device. The GUI allows control and feedback via text and audio, featuring an animated face that reacts to user input.

Computer Vision Currently our perception stack consists of several Yolo11 based models, each one of them with a specific skill related, such as Object Detection and Pose Detection. The information provided by these models are included in different pipelines in order to achieve the tasks goals presented in the @Home category. These models were enhanced by TensorRT inference to achieve near real-time performance while mantaining good accuracy.

2.3 Navigation

In order to navigate through robocup@home environment, Miss Piggy uses *Navigation 2* [2]. Thus, the robot can move from wherever it is to a desired point avoiding obstacles in its map using three layers of costmap2d ros2 package[3]:

- obstacle_layer: uses data from LIDAR as laser_scan and point cloud provided from realsense's sensors as sources of observation;
- static_layer: map acquired from mapping stage (explained in next subsection);
- inflation_layer: propagating cost values out from occupied cells that decrease with distance as stated by environment.

To perform the best movement according to Miss Piggy's structure constraints (e.g. differential driver, acceleration limits, minimum velocity), we adopted a smac local planner. An implementation of this approach in Robot Operation System is available in [4]. We are using it to achieve optimization of global planner at during runtime and minimizing the trajectory execution time.

Mapping Before executing tasks, the Miss Piggy robot maps the environment and tracks objects using Lidar, odometry, and IMU data combined via an Extended Kalman Filter, generating a 2D occupancy map through SLAM. We are evaluating SLAM algorithms for hardware limitations and sensor uncertainties, with RTAB-Map, accessible via ROS [0], proving effective for modular testing in virtual and real environments.

For localization, the EKF implementation in ROS [0] integrates wheel odometry, visual odometry, and IMU data for relative positioning, supplemented by Lidar for absolute measurements. ORB-SLAM, adapted from RGB-D applications [0], enhances visual odometry, while Adaptive Monte Carlo Localization (AMCL), as explored in [0], addresses map-based localization challenges.

3 Current Research

3.1 Human Robot Interaction

Looking for amplify the user experience of the robot the PMEC@HOME Team is developing models for voice biometrics (to enable speech recognition of people) and wakeword detection openWakeWord Framework [0], enabling more refined robot interaction. Aswell as being capable to maintain sparses conversation over time with the same person, being able to recover past information from a speaker to personalize usage of the robot.

3.2 Computer Vision

Person recognition The skill of recognizing a person among the orders is one of the fundamental skills for the robot. The team is developing a way to extend the current capability of recognize people's face via DeepFace, to also be able to differentiate people from their whole body image, reducing the limitation of the previous approach that needed the person's face in the reach of the camera to perform the recognition.

3.3 Robot Behavior

The task execution is a challenge in the @Home category, each test in the competition consists of several steps that we currently implement via scripts that make use of the ROS interfaces (services, actions, node) and Nav2 Behavior Trees. One of our lines of research is approach the task planning and execution in a more generalistic fashion. Using our LLM to organize the steps that the robot should execute and interact with the well established behavior trees.

Role	**Item**
Operating System	Ubuntu 22.04
Middleware	ROS2 Humble
Navigation	ROS2 *navigation stack*
Localization	*EKF*
Mapping	*RTABMap*
Object Recognition	Yolo11
Face Detection	Yolo11
Human Detection	Yolo11
Face Recognition	DeepFace
Speech Synthesis	FastSpeech
Speech Recognition	Conformer-CTC
Natural Language Understanding	Mistral

Table 1. Components of Miss Piggy

4 Miss Piggy's Components

This section describes the organization of the Miss Piggy's electronic components, detailing the role of each component in the system's composition, as well as the methodology applied to ensure efficient communication between components and system performance.

The layout of the robot's electric components and their connections are illustrated in the figure below. Next, the specifications and application methodology for each element that makes up the aforementioned system will be outlined. Finally, the methodology used to implement communication between the selected computers and sensors

4.1 Jetson Xavier AGX

The Jetson AGX Xavier is a cutting-edge AI computing platform designed for high-performance and efficient processing. It features an 8-core ARM v8.2 64-bit CPU, a 512-core Volta GPU with Tensor Cores, and a dedicated Deep Learning Accelerator, making it capable of handling complex AI and deep learning tasks. With 32GB of LPDDR4x memory and a high bandwidth of 137GB/s, it ensures smooth and rapid data processing. Its compact form factor and energy-efficient design make it suitable for embedded applications. In Miss Piggy the Xavier is dedicated for the AI applications, such as inference in the LLMs, Speech-to-Text/Text-to-Speech, Vision, Reinforcement Models.

4.2 Clearpath Jackal UGV

The Clearpath Jackal UGV is a robust mobile base designed for versatile applications, ideal for the @Home competition. Its open-source ROS interface enables seamless integration for advanced robotics development. The Jackal's onboard computing efficiently handles sensor input, allowing the team to focus on autonomous functionalities while hardware efforts shifted to creating a custom base. By encapsulating communication in Docker containers, interface reliability and usability were improved. This setup facilitates sensor access via micro-ROS and ROS2, connecting the base to systems managing Miss Piggy's intelligent autonomous behavior. The Jackal's adaptability and performance are pivotal to the project's success.

4.3 ECS BOX

The ECS_BOX manages the power supply for the entire Miss Piggy robotic system, excluding the Jackal, which operates and distributes power using its own battery. All other components of the robot are powered through the Central Energy System, which regulates power and voltage precisely and modularly for each component. On its side, individual interfaces can be observed for each component, including a switch, a protective fuse, and an XT60 connector that can directly power each component. This entire setup is microcontrolled by an ESP32 integrated with electrical circuits and a micro-ROS framework that allows the energy feedback communicate directly with the ROS2 environment capable of managing the system's power and equipped with an alert buzzer, and on emergency situations shutdown components.

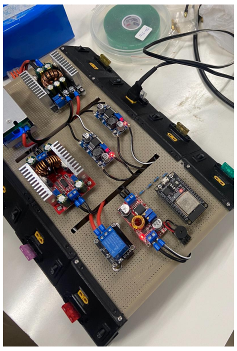

Fig. 2. Energy Center System BOX

Role	**Item**
Mobile Base	Clearpath Jackal UGV
Manipulator	Trossen ViperX 300s 6 DOF
Microphone	Rode Videomic Go
Displays	2x 15" LCD Displays
Main RGB-D Camera	Intel RealSense D455
Auxiliary RGB-D Camera	Intel RealSense D435i
Speaker	LG XBOOM XG5s
Human Detection	Yolo11
2D LIDAR	RPLidar A2
Embedded Systems	NVIDIA Jetson Xavier AGX
Microcontroller	ESP32
Gimbal Setup	2x Dynamixel Servos MX28AT

Table 2. Miss Piggy's Components table

5 Conclusions and future work

This is Pequi Mecânico's first attempt to compete in the category at this level. Our robot was designed based on other team's projects, information available at ROBOCUP @home wiki, and our own insights on the competition's challenges.

We have been looking towards to make the @Home experience more immersive, there are lines of research in the team coming up with new strategies based on point clouds obtained via a 3D Lidar using Geometric Deep Learning techniques to transpose the robot view of the world to VR/AR visualization.

The Deep Learning approaches we use will allow the robot to learn throughout the course while it receives commands and perform tasks. We already have several versions of Miss Piggy which perform different sets of tasks, and, until the competition, we will have many more. As a team, we are aware that we are still far from contributing significantly to the category scientifically. Our current focus is to establish a foundation from which we can better develop our ideas to solve the competition tasks more efficiently and to help evolve the league as a whole.

References

1. Nvidia riva sdk. `https://docs.nvidia.com/deeplearning/riva/user-guide/docs/index.html`. Accessed: 2024-07-23.
2. The marathon 2: A navigation system. `https://arxiv.org/abs/2003.00368/`. Accessed: 2024-07-23.
3. Ros2 costmap_2d. `https://github.com/ros-planning/navigation2/tree/main/nav2_costmap_2d`. Accessed: 2024-07-23.
4. Nav2 smac$_{planner.}$. Accessed: 2024-07-23.

Ros2wiki rtabmap. `https://github.com/introlab/rtabmap_ros`. Accessed: 2024-07-23.

Thomas Moore and Daniel Stouch. A generalized extended kalman filter implementation for the robot operating system. In *Intelligent autonomous systems 13*, pages 335–348. Springer, 2016.

Raúl Mur-Artal and Juan D. Tardós. ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras. *IEEE Transactions on Robotics*, 33(5):1255–1262, 2017.

Sebastian Thrun, Wolfram Burgard, and Dieter Fox. *Probabilistic robotics*. MIT press, 2005.

openwakeword. `https://github.com/dscripka/openWakeWord`. Accessed: 2024-07-22.

Timo Ojala, Matti Pietikainen, and David Harwood. Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In *Proceedings of 12th International Conference on Pattern Recognition*, volume 1, pages 582–585. IEEE, 1994.

Timo Ojala, Matti Pietikäinen, and Topi Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (7):971–987, 2002.

Ildoo Kim. Deep pose estimation implemented using tensorflow with custom architectures for fast inference. `https://github.com/ildoonet/tf-pose-estimation`.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.

Mykhaylo Andriluka, Leonid Pishchulin, Peter Gehler, and Bernt Schiele. 2d human pose estimation: New benchmark and state of the art analysis. In *Proceedings of the IEEE Conference on computer Vision and Pattern Recognition*, pages 3686–3693, 2014.

Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. In *arXiv preprint arXiv:1812.08008*, 2018.

Arthur Huletski, Dmitriy Kartashov, and Kirill Krinkin. Vinyslam: an indoor slam method for low-cost platforms based on the transferable belief model. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6770–6776. IEEE, 2017.

Giorgio Grisetti, Cyrill Stachniss, Wolfram Burgard, et al. Improved techniques for grid mapping with rao-blackwellized particle filters. *IEEE transactions on Robotics*, 23(1):34, 2007.

Christoph Rösmann, Wendelin Feiten, Thomas Wösch, Frank Hoffmann, and Torsten Bertram. Trajectory modification considering dynamic constraints of autonomous robots. In *ROBOTIK 2012; 7th German Conference on Robotics*, pages 1–6. VDE, 2012.

Roswiki amcl. `http://wiki.ros.org/amcl`. Accessed: 2024-07-23.

Roswiki gmapping. `http://wiki.ros.org/gmapping`. Accessed: 2019-06-17.

Roswiki teb_local_planner. `http://wiki.ros.org/teb_local_planner`. Accessed: 2024-07-23.

Roswiki navigation stack. `http://wiki.ros.org/navigation`. Accessed: 2019-06-17.

D. Adams. *The Hitchhiker's Guide to the Galaxy*. San Val, 1995.

Cloud speech-to-text - speech recognition | cloud speech-to-text | google cloud.

Open source conversational ai.

Ledell Wu, Adam Fisch, Sumit Chopra, Keith Adams, Antoine Bordes, and Jason Weston. Starspace: Embed all the things! *CoRR*, abs/1709.03856, 2017.

Zhiliang Peng, Wenhui Wang, Li Dong, Yaru Hao, Shaohan Huang, Shuming Ma, and Furu Wei. Kosmos-2: Grounding multimodal large language models to the world. *arXiv preprint arXiv:2306.14824*, 2023.

Anmol Gulati, James Qin, Chung-Cheng Chiu, Niki Parmar, Yu Zhang, Jiahui Yu, Wei Han, Shibo Wang, Zhengdong Zhang, Yonghui Wu, et al. Conformer: Convolution-augmented transformer for speech recognition. *arXiv preprint arXiv:2005.08100*, 2020.

Jungil Kong, Jaehyeon Kim, and Jaekyoung Bae. Hifi-gan: Generative adversarial networks for efficient and high fidelity speech synthesis. *Advances in Neural Information Processing Systems*, 33:17022–17033, 2020.

Adrian Łańcucki. Fastpitch: Parallel text-to-speech with pitch prediction. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6588–6592. IEEE, 2021.

micro ROS. micro-ros: Ros for microcontrollers. `https://micro.ros.org/`. Accessed: 2024-07-23.