BahiaRT 2022: Team Description Paper for RoboCup @Home League*

Tiago de Jesus da Silva[1] João Vítor Café[1] Thomas Gael[1] Daniel Bahiense[1] Davi Miguel Brito Barbosa[1] Josemar Rodrigues de Souza[1] Robson Marinho da Silva[1] Marco A C Simões[1] Jorge Campos[1] Ana Patricia Magalhães[1]

***Abstract*— This paper presents the Bahia Robotics Team (BahiaRT) and describes an autonomous service robot named Bot Intelligent Large capacity Low cost (Bill) and its functions, such as navigation, manipulation, people and object recognition, human-robot interaction and decision making as well as the Bill's hardware and software systems. Furthermore, the paper highlights research interests and scientific contributions of BahiaRT from the Center of Computer Architecture and Operating Systems (ACSO) at the Bahia State University (UNEB).**

Keywords: BILL, service robotics, object detection, assistive robotics, speech recognition, RoboCup@Home.

I. Introduction

Applications in robotics advanced in the last years and evidently RoboCup has contributed for this growth. BahiaRT is a scientific cooperation group of ACSO that, since its creation (2006), have participated in the RoboCup leagues, as well as regional tournament that follows the same rules. For BahiaRT is very relevant to participate of RoboCup competitions where it is possible to strengthen the Brazilian engagement in Artificial Intelligence (AI) and robotics, sharing advances of solutions. In RoboCup Competition, initially, the team competed in 2D Soccer Simulation League and in the Mixed Reality demonstration competition. In Mixed Reality, BahiaRT got the third place in RoboCup 2009 and the fourth place in 2010. BahiaRT also has developed the MR-Soccer Server, the main module of MR software infrastructure. In other league named 3D Soccer Simulation, BahiaRT ranked the fourth place in 2015 and 2016 and the sixth place in 2017. In RoboCup@Home, BahiaRT got the 13th in 2015 and 21th in 2016. Specifically, in this @Home league of Latin-American and Brazilian Robotics Competition (LARC), BahiaRT got the second place in 2015, 2016 and third in 2017.

Thus, Bill (Fig. 1) is the proposal of BahiaRT for RoboCup@Home league. It was born in 2014 as results of research projects in assistive robotics and its main functions are communicating with humans through natural language processing, recognizing objects, faces and navigating through unknown environments.

Section II describes the main advances and scientific contributions of BahiaRT to assistive robotics. Section III describes the architecture and main functionalities currently developed for Bill. In section IV some experiments and results are presented. Section V presents the conclusions and future work. Finally, at the end of this Team Description Paper (TDP) there is a brief description of each hardware and software component used in Bill's development.

Fig. 1. Bill and its concept.

II. BahiaRT's advances in innovative technology and research interests

The main research interests of BahiaRT involve robotics and artificial intelligence, specifically focusing on interaction and cooperation between human-robot, applications of artificial intelligence for services, standardization and system integration, in addition to the construction of intelligent robots with low cost components and open source. Developing autonomous robot as Bill requires knowledge of computer and mechatronics technology to develop algorithms,

* This work is partially funded by CNPq/PIBIC, FAPESB/IC and UNEB/PICIN

[1]Centro de Pesquisa e Arquitetura de Computadores, Sistemas Inteligentes e Robótica (ACSO), Universidade do Estado da Bahia (UNEB), Salvador, BA, Brazil. teambahiart@gmail.com

integrated systems and hardware solutions. Bill uses various innovative technologies in an integrated approach, such as Robot Operating System (ROS) and its packages, Arduino boards, kinect OpenNI library, TurtleBot arm, computational vision and speech algorithms, among others. The experiments and texts used to evaluate Bill's behaviors involve specifications of parameters, control tasks, details of the software and hardware, innovative technologies, and adotion of other challenger tasks in AI. Thus, the development of new methods and technologies to provide integrated solutions and the training of practitioners to develop complex task in robotics is an important contribution of BahiaRT over the years. However, despite of the importance of this type of research, there is still not several autonomous robots running in human daily lives. Furthermore, as well as in other sciences, autonomous robots can achieve different results in practice due to the occurrence of unpredictable events and environmental changes. Therefore, there is still a long road ahead, with much work to be done for spreading services robots in the residences around the world. To fill this gap, practitioners should provide feedback on robotic research, sharing experiences for reducing differences among countries and their laboratories. To allow replication of research and to contribute for development of robotic research, BahiaRT will provide access to the codes, hardware and software list in this paper and in Bill's website.

III. Bill's Architecture and Functionalities

The functions implemented in Bill are represented in the diagram in Figure 2 and their main features are described in the following subsections. The Bill's architecture is basically divided into levels: the high level where the functions related to the robot's abilities are; and the low level where the controllers and drivers that send signals and receive commands from the sensors are located. At the end of this article there is a list of the hardware and software used.

During the pandemic, the researchers did not have access to the laboratory and therefore to Bill. So, we focused on researches related to navigation, speech recognition and people and objects recognition using simulated environments. Now we are working on the integration of these functionalities to the physical robot.

A. Navigation

Navigation is the keystone for efficient execution and environment interaction for robots. The components used by Bill for navigation are: encoders output, odometry, gmapping (ROS), move_base (ROS), Adaptive Monte Carlo Localization (AMCL) (ROS), map_server (ROS) and 360° laser scanner. The encoder data is used by odometry module to estimate the movements of the robot in space. Further, the odometry data is used to trace trajectory to a desired target by the move_base. Once all data are being published, the simultaneous mapping and localization using the AMCL [1] is activated integrating the 360° laser scan data. Simultaneous Localization And Mapping (SLAM) approach is used to map the environment and provide self-localization in this

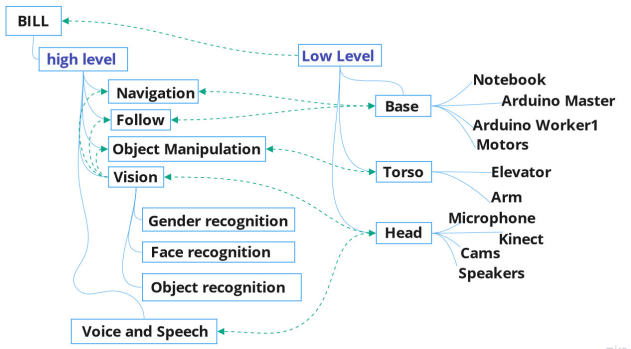

Fig. 2. Overview of the Bill's architecture and functions.

map. First the map is built using the incremental mapping package, Hector Mapping [2]. Then, the grid map generated is done by RPLIDAR 360 Laser Scanner sensor, which is able to get the 2D world data. The next step is creating the path planning based on the occupancy grid map that is updated based in Dynamic Voronoi ROS package approach [3]. Then the shortest path to the goal is computed by means of D* Lite algorithm [4] to ensure obstacle avoidance over incremental mapping. The motion planning in charge of getting the path planning and relating linear and angular motion is triggered, which applies the kinematics control law, and sends a message to low-level control.

B. People Detection and Tracking (Follow)

During Human Robot Interaction (HRI) people detection and tracking has a crucial role for service robots like Bill. The kinect OpenNI library provides position identification of key points on the human body, such as head, torso, knees, etc. This representation resembles a human skeleton, and it allows to obtain a persons' position relative to the camera on the robot. The library also assigns an ID for each person it identifies, allowing to track a specific person while he or she moves in front of the camera. That feature is used as an input for the navigation system, which will plan a path to follow a specific user in cluttered and dynamic environments. As we could not access Bill during the pandemic, we are currently working on people tracking implementation.

C. Object Manipulation

Object manipulation plays an important role to interact with a home environment. To meet that requirement, Bill has an arm based on the TurtleBot Arm, composed of 5 Degrees of Freedom (DoF) including a gripper, which will allow the robot to grab lightweight objects. During the pandemic we could not work on this functionality. So, we are currently working on the implementation of Bill's object manipulation.

D. Vision

This module is responsible for receiving, processing, and responding to external stimulus from image capture. It is divided into two sub-modules which are facial recognition and object recognition.

Facial recognition [5] is a biometric technique based on people's facial features. It consists of analyzing a characteristic facial pattern and use some algorithm to identify such an individual. Applied to robotics, facial recognition allows the robot to identify a specific person to perform actions based on that. In our project facial recognition is focused on recognizing people with and without masks. We understand that this is important and useful, for the current moment, all using a convolutional neural network model[6]. This model is called VGG16-Mixed-Finetuning, and considers 3 main characteristics: architecture, pre-processing, and initialization.

The model's architecture adopted is VGG16[7]. The input layer accepts 224x224 size RGB images. The image goes through a stack of convolutional layers, where filters are used with a very small 3×3 receptive field which is the smallest size to capture the notion of space. One of the configurations also uses 1×1 convolution filters, which can be seen as a linear transformation of the input channels. The convolution pass is fixed at 1 pixel, the convolution spatial fill at the input layer is such that the spatial resolution is preserved after the convolution. Spatial grouping is accomplished by five layers of maximum grouping, which follow some of the convolution layers (not all convolution layers are followed by maximum pooling). The maximum pool is performed in a 2×2 pixel window. It is built on the Caffe Deep Learning Framework using the data sets ImageNet[8] and IMDB-WIKI[9]. These data sets have photos that try to get as close as possible to real conditions, with different lighting, angles, and distances. Therefore, the algorithm can well recognize a face even in uncontrolled environments.

Pre-processing is a technique applied to extract the maximum possible characteristics of the faces that are in the photos before passing through the neural network. In our model, pre-processing is mixed, as it uses two techniques, those which are horizontal alignment and flat alignment.

Initialization defines how the Convolutional Neural Network (CNN) model started, if it was built from scratch or if it is a finetuning with a pre-training data set. The model we use is finetuning because it was pre-trained with the IMDB-WIKI dataset[9] before being trained with the ImageNet dataset[8]. Using a pre-trained model is one of the advantages of the CNN approach because it allows the faster achievement of better results since the model is already started with meaningful filters.

In our application we considered gender as classification problem, i.e. consider the definition of classes for men and women. So that, as soon as a face is detected, the algorithm identifies in which class the face fits better.

Object detection.[10] is a technique for identifying, classifying, and sometimes labeling inanimate objects. It can be used, for example, in autonomous cars for the detection of obstacles, traffic lights, among other objects. In robotics context, object detection allows the robot identify a particular object to manipulate it using a claw, or even identify the object it is seeing.

In this project we adopt the You Only Look Once (YOLO)[11] technology for object detection. It uses an open source convolutional neural network called Darknet[12] that divides the image into regions and provides bounding boxes and probabilities for each region. Detection accuracy varies according to how much the CNN was trained. So, this tool has an advantage when compared with systems based on classifiers that have a certain lack of consistency, as the detection accuracy varies a lot according to the environment. Darknet has a precise and fast detection structure, which depends on how the custom model was built, taking into account, for example, the set up of training data, and the number of image batches per object. This structure is more efficient than other solutions, such as systems based on classifiers, because is executed in CUDA[5], an Nvidia API focused on parallel computing, which allows the extraction of the maximum performance from the GPU. In addition, YOLO provides a set of pre-trained models to recognize objects available in YOLO repositories. Therefore, Darknet training involves selecting images that will be part of the dataset, labeling the images using a labeling tool called Yolo mark, train the network using a configuration file that varies according to the machine's processing capacity, use a file of pre-trained weights, and finally is possible to detect the objects taught to the neural network artificial.

E. Speech Recognition and Voice

The voice is the most used form of human-machine interaction to give commands to the robot, either through the line of command or natural language. For dialect recognition, Bill uses Vosk and Kaldi, which has greater flexibility for adaptation and personalization of speech, allowing to adapt from each dictionary and acoustic models to perform the conversion of each phoneme and contextual understanding.

Therefore, the grammar is capable of interpreting the commands and fulfill assigned tasks. To talk to people, Bill uses the ROS sound pack play which can translate a ROS topic [13] into sound. Within the synthesis process, the software Vosk allows developer to change various aspects of voice, such as tone, speech rate, among others, to ensure better understanding by the listener, enabling a better interaction experience.

IV. Experiments and results

Lots of experiments were made with Bill to measure and prove its abilities in different environments. To illustrate our experiments, this section describes the face recognition results and the speech abilities.

Seven people were used in face recognition tests, which were divided into two steps: the former, only one person in front of the robot and the second with two people. In the first step 50 rounds were done for each person, in which the person appeared in front of the robot for 300 cycles and the robot had to recognize it. Bill attained an average of 72.64% accuracy. In the second one, 7 rounds were done combining the people in different pairs for each round. As in the first step, the robot had to recognize people during 300 cycles.For every round, the robot had to recognize the people during 300 cycles. Bill scored the average result of 66.67% accuracy.

The speech recognition had understanding and answering a serie of questions. Four people in two different environments and for each person the questions were asked 100 times. The indoors environment tests were done with the operator 75 cm distant from the robot. Bill obtained an average of 61.25% accuracy in this requirement. The outdoors tests, using the same settings as previous, scored an average of 56.67% accuracy in this task.

V. Conclusions and future work

This paper presents the main abilities of Bill, a robotics solution proposed by BahiaRT for assisting humans in their daily chores.

In order to perform tasks, Bill integrates abilities related to voice recognition, computational vision, navigation and manipulation. However, due to the pandemic we did not have access to Bill for two years. So, over this time the researches focused on improve these abilities in a simulated environment. With the return to the physical lab we initially revised Bill's hardware and now we are working on integrating these skills into the physical robot. We hope to improve Bill's functionality by implementing the results obtained with the simulated environment.

References

[1] F. Dellaert, D. Fox, W. Burgard, and S. Thrun, "Monte carlo localization for mobile robots," in *Robotics and Automation, 1999. Proceedings. 1999 IEEE International Conference on*, vol. 2. IEEE, 1999, pp. 1322–1328.

[2] S. Kohlbrecher, O. Von Stryk, J. Meyer, and U. Klingauf, "A flexible and scalable slam system with full 3d motion estimation," in *Safety, Security, and Rescue Robotics (SSRR), 2011 IEEE International Symposium on*. IEEE, 2011, pp. 155–160.

[3] B. Lau, C. Sprunk, and W. Burgard, "Improved updating of euclidean distance maps and voronoi diagrams," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. IEEE, 2010, pp. 281–286.

[4] J. Neufeld, M. Sokolsky, J. Roberts, A. Milstein, S. Walsh, and M. Bowling, "Autonomous geocaching: Navigation and goal finding in outdoor domains," in *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1*. International Foundation for Autonomous Agents and Multiagent Systems, 2008, pp. 47–54.

[5] A. Tourani, "Cuda tutorial - how to start with cuda?" 12 2018.

[6] W. Samek, A. Binder, S. Lapuschkin, and K.-R. Muller, "Understanding and Comparing Deep Neural Networks for Age and Gender Classification," in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*. Venice: IEEE, Oct. 2017, pp. 1629–1638. [Online]. Available: http://ieeexplore.ieee.org/document/8265401/

[7] V. Khandelwal, ""the architecture and implementation of vgg16,"," https://pub.towardsai.net/the-architecture-and-implementation-of-vgg-16-b050e5a5920b, accessed: 2021-08-25.

[8] "ImageNet." [Online]. Available: http://www.image-net.org/

[9] "IMDB-WIKI - 500k+ face images with age and gender labels." [Online]. Available: https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/

[10] "Why object detection matters," https://www.mathworks.com/\discovery/\object-detection.html, accessed: 2020-08-23.

[11] "Yolo," https://https://pjreddie.com/darknet/yolo/\/yolo\/, accessed: 2020-08-23.

[12] "Darknet," https://pjreddie.com/darknet/\/darknet\/, accessed: 2020-08-23.

[13] "Sound_play," http://wiki.ros.org\/sound_play, accessed: 2017-01-18.

Bill Hardware and Software Description

To provide completely autonomous operation, Bill owns two main modules of control: (i) the High-level control, which includes algorithms to solve functionalities such as global task planning, navigation and tracking, recognition of objects and faces, user-interaction, among others; and (ii) a low-level to control sensors and actuators in the real world.

Bill Hardware Description

Bill has a motion base that presents higher mobility. It is a round base with 2 differential drive wheels and 2 free wheels -one in the front and other in the rear for maintaining balance. All the electronic parts were carefully checked to avoid short-circuits and increase power. The details of each hardware are described as follows:

- **Base:** Two Arduino Mega 2560; Two motors IG32P 24VDC 190 RPM Gear Motor with Encoder; One Notebook Samsung NP550XDA-KU1BR Intel Core i7-1165G7; One digital buzzer; One RPLIDAR 360°laser scaner; Three Sabertooths controlers; One LM35 linear temperature Sensor; Three batterys 11.1 volts and 2800 mAh; One digital push button;
- **Torso**: Mini actuator Firgelli Automations; One Emergency switch;
- **Arm**: five Dynamixel-ax-12A; One ArbotiX-M; Maximum load; 1kg.
- **Hea**d: One Dynamixel-ax-12A; One Microsoft Kinect sensor; Two Microsoft life Cam HD-3000; One Rode Videomic Pro.

Bill Software Description

The low level is composed of a proportional control running on arduino boards. The communication and high level system is composed of tools developed by our team and open source applications of the Robot Operating System (ROS). The software are:

- Navigation, localization and mapping: Hector mapping.
- Face recognition: OpenCV library.
- Speech recognition: Vosk and Kaldi library.
- Object recognition: YOLO.