Team description paper for Robocup 2024, PUMAS TEAM *

Jesus Savage[1] Oscar Fuentes[1] Rubén Sandoval[1] Oscar Durón[1] and Daniel Vanegas[1]

National Autonomous University of Mexico, Coyoacan Mexico, Mexico

Abstract. This paper describes the current research topics and main findings of Pumas Team as well as the efforts to implement all the developed software into the Toyota Human Support Robot (HSR). Pumas Team has participated in several national and international robotics competitions that have influenced our architecture for the development of better systems for our service robots. Current research topics include facial detection, the creation of RGB-D representations of the environment and action planning using state space representations. A novel YOLO training method is also used to recognize the required datasets. In our robotics architecture, the Virtual and Real robot system (VIRBOT), the operation of service robots is divided into several subsystems, each of which has a specific functionality that contributes to the final operation of the robot. By combining symbolic AI with digital signal processing techniques a good performance of a service robot is obtained. We consider that our robotics architecture, VIRBOT, will be transferred successfully into the Toyota Human Support Robot (HSR).

Keywords: Domestic Service Robot · HSR · Autonomous robotics system.

1 Introduction

Service robots are increasingly becoming a reality in many environments, such as houses, offices or hospitals. So the ability to safely interact with the environments and other agents, such as humans, pets or other robots is of the utmost importance. With this in mind, the final goal of service robots must be to make the lives of humans easier and more comfortable. Also, a robot can be an excellent companion, for example for elderly or lonely people, making their lives better and happier.

To achieve this, a service robot must be capable of understanding spoken and visual commands in a natural way from humans, navigate in known and unknown environments avoiding static and dynamic obstacles, recognize and manipulating objects, detect and identify people, among several other tasks that a person might request.

* UNAM, MEXICO University and Toyota Motor Company for HSR

The team Pumas has participated in national and international competitions, such as the Robocup 2018, in which our team obtained the second place in the category DSPL@Home. Then in the RoboCup 2019 we got 4th place. The same place obtained during the virtual version of the Robocup that took place in 2021 due to the COVID19 medical emergency. We hope the 2024 edition will mark the return to the real-world competition, where we hope to take where we left off and showcase some of the research being conducted in our lab since then.

Our service robot "Takeshi" is an HSR Toyota robot, has the following software configuration that it is based on the VIRBOT architecture [**virbot**], which provides a platform for the design and development of software for general purpose service robots, see figure 1. The VIRBOT architecture is implemented in our robots through several modules that perform well defined tasks [**muller**], with a high level of interaction between them. The principal framework used for interaction is ROS, where a module is represented by one or several ROS's nodes. Also, for modules using the Microsoft operating system, we use our own middle ware called Blackboard to link them with ROS nodes running on Linux. In the following sections are explained each of the layers of the VIRBOT system.

1.1 Inputs Layer

This layer process the data from the robot's internal and external sensors, they provide information of the internal state of the robot, as well as, the external world where the robot interacts.

Some of on board sensors are a laser scanner, a RGB-D camera, an array of microphones, one stereo camera and two cameras with wide-angle lens. Digital signal processing techniques are applied to the data provided by the internal and external sensors to obtain a symbolic representation of the data, as well as, to recognize and to process voice and visual data. Pattern recognition techniques are used to create models of the objects and the persons that interact with the robot. With the symbolic representation this module generates a series of beliefs, that represent the state of the environment where the robot interacts.

1.2 Planning Layer

The beliefs generated by the perception module are validated by this layer, it uses the Knowledge Management layer to validate them, thus a situation recognition is created. Given a situation recognition, a set of goals are activated to solve it. Action planning finds a sequence of physical operations to achieve the activated goals.

1.3 Knowledge Management Layer

This layer has different types of maps for the representation of the environment, they are created using SLAM techniques. Also in this layer there is a localization system, that uses the Kalman filter, to estimate the robot's position and orientation. A rule based system, CLIPS, developed by NASA, is used to represent

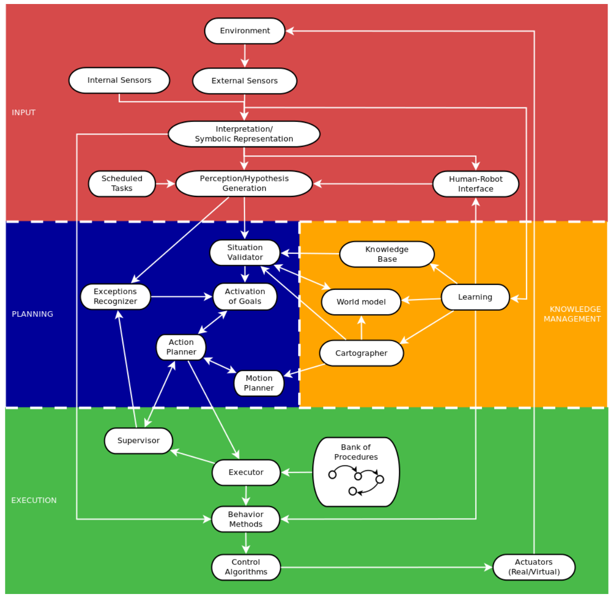

Fig. 1. Virbot System for mobile robots.

the robot's knowledge, in which each rule contains the encoded knowledge of an expert.

1.4 Execution Layer

This layer executes the actions and movements plans and it checks that they are executed accordingly. A set of hardwired procedures, represented by state machines, are used to partially solve specific problems, finding persons, object manipulation, etc. The action planner uses these bank of procedures and it joins some of them to generate a plan.

2 Technical challenge

How your team works under the following concepts:

2.1 Technical Challenges for 4S

Regarding Service Robots, describe what you understand for each aspect in the 4S philosophy (Speed, Smooth/Smart, Stable and Safe), the state-of-the-art, technical challenges, and your own approaches to the problem:

Speed In a way, speed is the price to pay for the attention we pay to safety. Regarding navigation and manipulation, speed of computation however is obtained via the gt1080GPU in an alienware laptop. The main calculation tool for our robot. This is of the outmost importance for getting results in realtime with algorithms such as Open Poses, YOLO and Facenet.

Smooth/Smart As mentioned above we are using the Virbot architecture, in which the knowledge management and the planning layers interact to obtain and update plans, given the knowledge database available to the robot at any given moment. We strongly believe the Human Robot Interface to be extremely important for a smooth experience. Accordingly we have invested plenty of time refining this interaction, with use of vocal and non vocal commands, any unexperienced user should be able to use our service robot.

Stable Stability has not been much of an issue so far. However we have some policies in place to assure this to continue like this. Again, the Virbot architecture allows for a modular development and growth of the code. All the sections are "transparent" so modifying one should not affect the other nodes.

Safe For any service robot, safety is the most important of the 4S's... Human users must be capable of interacting with the robot, no matter the circumstances, without any risk. Specially given the fact that HSR is meant to assist the elderly and/or delicate patients. The safety of the robot is also important, and conservative navigation approaches are preferred in our implementation. Besides the HSR's bumpers we relay on a obstacle detector capable of reading the point clouds from the Xtion and determine the presence of unmapped objects, that allow a trajectory re planning and dynamic obstacle avoidance, additionally we use a potential fields approach for local planning.

2.2 System Design for "Keep moving", "Move carefully" and "Be clever"

Navigation in a previously mapped location is done using several components. First of all, ROS's AMCL package is used for navigating the environment and planning routes (via A*). Additionally, as mentioned above, an obstacle detector is working in the cloud point to avoid unmapped obstacles. This allows for re planning the trajectory and safe navigation. On top of this we have implemented a reactive behavior, based on potential fields approach, that allow for smaller objects avoidance, as well as short routes that aren't replanned. The potential fields act as an extra safety layer as well as adding speed fir small adjustments that don't require route re planning. A final layer for localization recovery is done using Hidden Markov Models and the hokuyo readings.

3 Software development policy

3.1 Tools

All our software is shared via git (gitlab). All members of our team have equal rights and permissions to access the codes database. We intend out software to be open and feedback is always welcome in our lab. We share most of our codes with Justina, another service robot in our lab. Each robot has it's own git and software versions, exchange is not only allowed but incentivized. Since the different robot configurations have proved useful in creating less hardware dependent codes.

3.2 Open source software used by your team.

Our team has always tried to combine open software packages with digital signal processing techniques developed by our lab members. The open software packages we currently relay on are:

Package	**Use**
Ros's AMCL	(Navigation of previously obtained maps)
Hector Map	SLAM
OPEN POSES	Human Pose detection
YOLO (darknet)	Objects recognition
FACENET	Face Recognition
OPENCV	Computer Vision
Open AI Vosk	Speech Recognition

Table 1. Open source software used by Pumas Team

4 Additional Innovative technologies

The policy in the Bio-Robotics laboratory at UNAM has always been to develop our algorithms when possible. Another distinctive characteristic derives from the fact that we are part of the Digital Signals Processing department at UNAM, and as such we favor classic techniques, and statistical approaches that yield similar results as state of the art more complex algorithms with a lower computational cost. This is especially true when compared with neural networks and other deep learning techniques.

In figure 3 we perform a statistical analysis, to correct the point cloud given the position and orientation of the sensor. Once this correction is performed the segmentation task is trivial to solve. A **PCA** (Principal Component Analysis) is performed on each separate object to estimate the rotation of the object, to align the gripper in the appropriate axis.

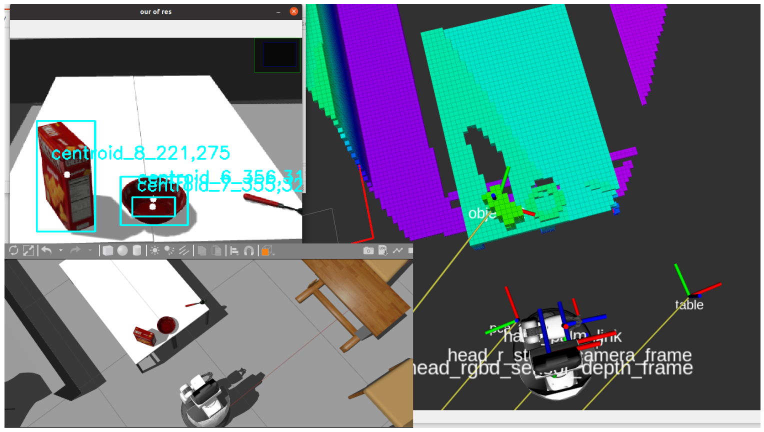

Fig. 2. Segmented Objects Note the rotation estimated with PCA is close enough to the real orientation of the object.

Another interesting front in our group research emerged during the COVID-19 pandemic lockdown. Since the university was closed, our team worked intensively on simulators like gazebo, and successfully crossed the reality gap when training 3D models of the objects that were later used in the real world objects. A more detailed version of the method can be seen in section 5.5

Objects from the YCB object data set were modeled using 3D software (Blender). This models were later used for fine tuning the YOLO v-5 used during the mexican national At Home League **TMR**.

5 Object Detection with Convolutional Neural Networks

Convolutional Neural Networks are the state of the art in object detection [**RazavianASC14**]. These types of neural networks are trained with large image datasets, such as ImageNet[**ImageNet**] or COCO[**Coco**], which include more than 330,000 tagged images. However, the training of this type of networks with own objects, or objects that are not in a public dataset, result in some problems: large amounts of labeled images are required by each class or object and the images must be labeled by specialized software, because labeling each object by hand could take a week of work. For that reason, We propose a system capable of generating an labeled image dataset automatically from video files, which is subsequently used for the training of a Convolutional Neural Network (YOLOv3[**yolov3**]). In addition, a system is proposed that using the detection of objects with the Convolutional Neural Network in a service robot.

The first system is made up of two main modules developed in ROS: the first module segments an object from each frame from a video file. The second module creates the database of synthetic images, that is, artificial scenes containing the

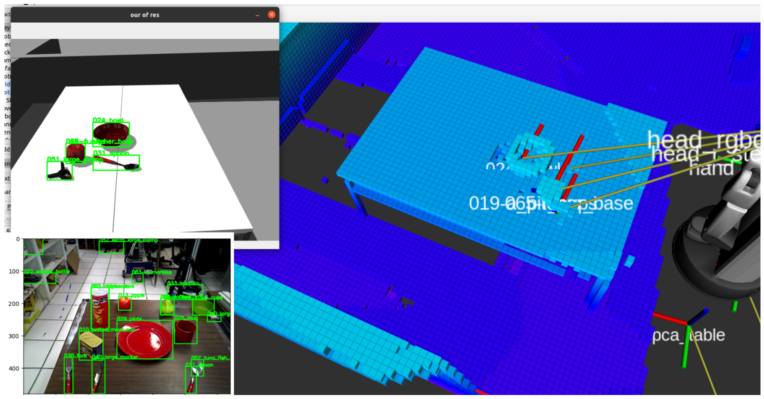

Fig. 3. Synthetic objects were used for training the model with good enough results in real life. (Reality Gap).

segmented and labeled objects of the previous module, finally, data augmentation techniques are applied to generate a more robust set of data. We generate 20,000 labeled images with 30 objects approximately in each one. Once the tagged synthetic image dataset has been generated, the training parameters for the Convolution Neural Network are configured.

In the second system, a methodology is proposed so that a service robot can manipulate an object autonomously. The system input is the images acquired by an RGB-D camera. Then, the objects are detected in a two-dimensional space using the training of the Convolutional Neural Network. Subsequently, the centroid calculation of each object in a three-dimensional space is made and the best way to manipulate the object is evaluated. Finally, in a second detection, it is confirmed that the object has been manipulated. In figure 4, We tested the proposed system with the images acquired by the Xtion sensor of own HSR robot for the HSR Challenge-4 presented last August.

The HSR robot and the Justina robots are the main projects in service robots developed in the Bio-Robotics laboratory of the UNAM Faculty of Engineering [**LabBiorobotica**]. Therefore, one of the main objectives of this system is to apply to these service robots.

6 Relevant publications

- *A SLAM system based on Hidden Markov Models SLAM, December 2021Informatics and Automation 21(1):181-212 DOI: 10.15622/ia.2022.21.7 Fuentes Oscar Savage Jesus*
- *Map representation using Hidden Markov Models for mobile robot localization*
  January 2018MATEC Web of Conferences 161(4):03011

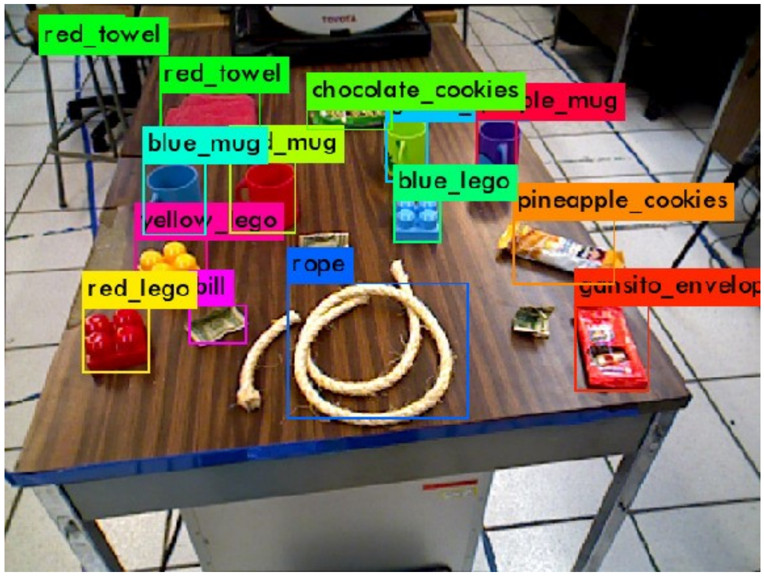

Fig. 4. Detected Own Objects with a 3-hour training in a Nvidia Quadro P4000. We used this objects in the HSR Challenge-4 at August 2019.

DOI: 10.1051/matecconf/201816103011
LicenseCC BY 4.0 Savage Jesus Fuentes Oscar Negrete Marco

- Sparse-Map: automatic topological map creation via unsupervised learning techniques
  August 2022Advanced Robotics 36(1):1-11
  DOI: 10.1080/01691864.2022.2114296 Jesús Hernández Savage Jesus Negrete Marco Fuentes Oscar

7 Link to Team Video, Team Website

Link to introduction video: `https://www.youtube.com/watch?v=_w2Ta3c3jQg`
Link to qualification video: `https://www.youtube.com/watch?v=_w2Ta3c3jQg`
Link to Team Website: `https://biorobotics.fi-p.unam.mx/`

Conclusions

It is clear, that during the 13 years in which our team Pumas has been participated in the RoboCup, 2 years in the Rockin [**Rockin**] in the category @Home and World Robot Summit 2018, the performance and research developed, in the service robot area, in our laboratory has been improved considerably. Our service robot architecture, the VIRBOT, has been evolving according to the requirement that these robotics competitions asked each year. In terms of software, we have change the way of conceiving the tests of the competition: from static state machines to inferred action planning generated by a rule based system. As for future work, the computer vision algorithms will be improved by using Hidden Markov Models (HMM) to have a better recognition of objects and persons. Also, it will be explored fault tolerant systems to help the robot to recover from failures.