PROCEEDINGS OF SCIENCE

Robotics, AI and Machine Vision

Jesus Savage[*], **Angelica Nakayama, Carlos Sarmiento**
School of Engineering, National Autonomous University of Mexico (UNAM), Mexico
E-mail: robotssavage@gmail.com, annkym@fi-b.unam.mx,
ing.adriansarmiento@comunidad.unam.mx

This paper presents a compilation of the accumulated development in the Bio-Robotics laboratory of the Engineering Faculty at UNAM. The techniques used to build robotic service systems are briefly mentioned. In particular, a line of research has been developed using most of them interdependently in what is called in a hybrid system ViRBot (VIrtual and Real roBOt sysTem); a system to operate mobile robots.

ViRBot is an abstraction system integrated into four specialized layers; Input, planning, knowledge management, and execution to provide intelligence to a mechatronic agent to execute the service in domestic environments. This project has been developed incrementally with the participation of students and profesors. Currently, the ViRBot system inter-operates with ROS and together, they form an intelligent systems development platform for computer vision, digital signal processing, automatic planning, automatic control and human-robot interaction.

Artificial Intelligence for Science, Industry and Society, AISIS2019
October 21-25, 2019
Universidad Nacional Autónoma de México, Mexico City, México

---

[*] Speaker.

PoS(AISIS2019)043

1. Introduction

Nowadays robots are interacting more and more with people in everyday life, because of this, robots are becoming an important part of human society. Robots are being used on different fields and activities such as assistants, pets, toys, tour guides, personal service and are even sent to space to perform complex or dangerous tasks.

Specialists are not just betting on building robots that only perform specific tasks but also robots that are able to widely participate on human societies, are more agreeable and can fulfill humans will [1]. In general terms, human beings prefer to interact with robots in a similar way they do with other humans, that is why interactive robots are beginning to be very popular.

"Interactive robots" is used to describe robots in which social interaction plays a leading role on contrast with those that use conventional human-robot interaction (such as teleoperated robots). To achieve this kind of interaction, robots need to perform a certain degree of adaptability and flexibility, and sometimes is desirable the robot develops its own skills as time goes by.

These are the reasons why robots now need more complex cognitive skills in order to operate efficiently and safely on naturally human populated environments and to achieve higher human cooperation and communication levels.

Even though interactive robots are already being successfully used, there is still so much left to do, for instance, if we want robots to be accepted as “natural beings”to interact with, they need to show sophisticated social skills like context and social convention recognition. So now behavior design, appearance, cognitive and social skills are becoming a challenge that requires a series of interdisciplinary knowledge and abilities.

There is a need for reliable Human-Robot interaction systems as experienced by the proliferation of robotics’ competitions that exalt the robots social interaction with humans, such as the Robocup and Rocking in the league @Home. The goal of this league is to promote the development of real-world applications and human-machine interaction with autonomous robots, or as they put it: “The RoboCup@Home league aims to develop service and assistive robot technology with high relevance for future personal domestic applications”[2].

2. Robotics Architectures

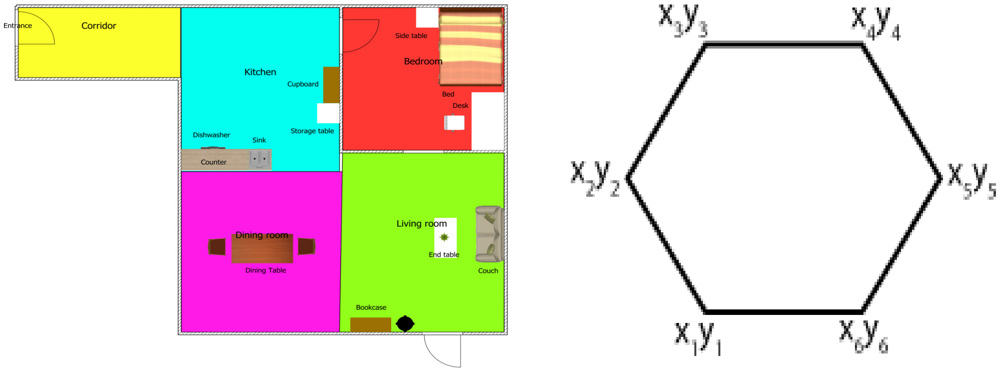
There are many different approaches when developing a robotic architecture; some of the most important are:

- Traditional
- Reactive

- Probabilistic
- Hybrid

2.1 Traditional Architectures

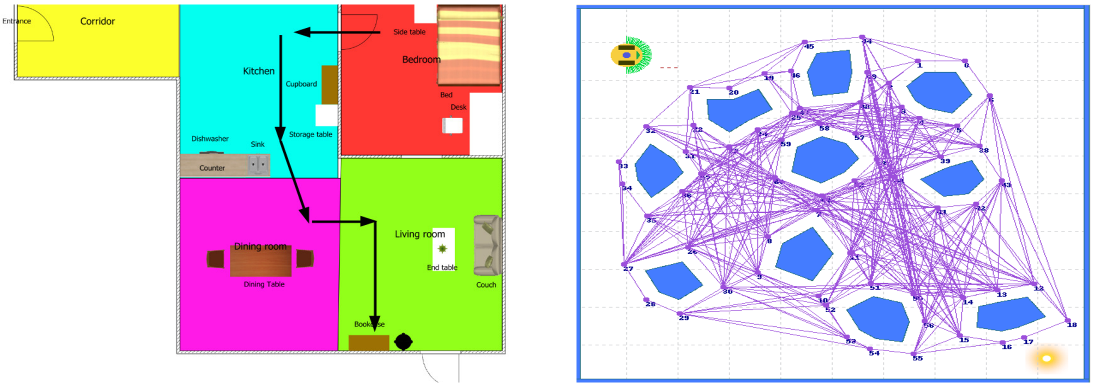
Traditional robotics architectures have distinctive features: They have a representation of the environment, with a symbolic representation of the objects in each room (Figure 1a). These are represented by polygons where they have their vertices x_i, y_i, ordered clockwise, these polygons separate the occupied space and the free space where the robot can navigate (Figure 1b).



(a) Environment [3]. (b) Polygons.

Figure 1: Environment representation

Movements and actions are planned using traditional artificial intelligence techniques such as search in topological networks (Figures 2a and 2b) where the basic problem is: Given a starting point (node), goal point (node) and a map of nodes and connections, find some path or find the "best" path (maybe shortest) and traverse it.



(a) Global Path [3]. (b) Local paths for each room.

Figure 2: Search in topological networks

These AI searching techniques can be summarized according to the goal they pursuit. Nevertheless, there are some cons of these architectures as they have a serial organization (Figure 3). If one module fails, the entire system fails; this type of systems are not suitable for dynamic environments or for robots that have errors in movement and sensing. The robot Shakey was the first example of this architecture, developed from approximately 1966 through 1972 at Stanford University.

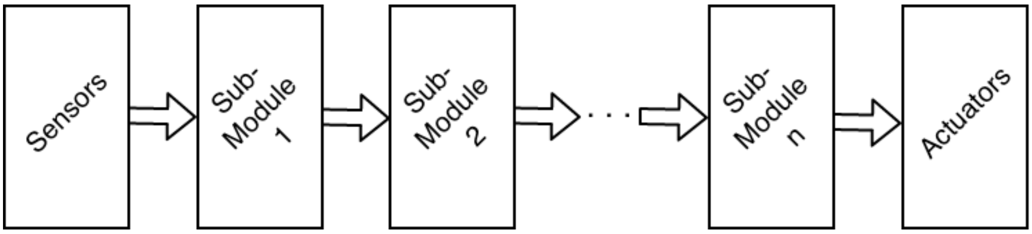


Figure 3: Serial organization.

2.2 Reactive Architectures

Reactive robotics architectures are based on the behavior of insects and do not need any representation of the environment and do not use action or movement planning. These architectures are suitable for dynamic environments with sensing errors and are based on behaviors running in parallel. These behaviors are represented using stimulus-response or SR diagrams (Figure 4).

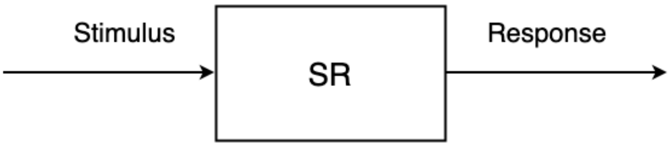


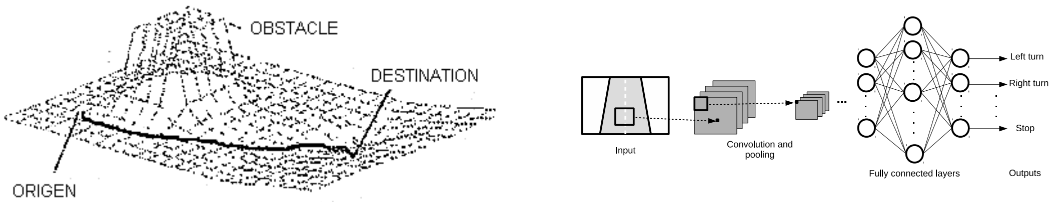
Figure 4: SR diagrams.

The output of each behavior must be instantaneous from the moment there is an entry. Behaviors are independent of each other and can be designed using zero order logic, state machines, potential fields, neural networks, etc.

2.2.1 Behaviours

On Zero Order Logic behaviour's design, the input sensors' values are checked and if they comply with a certain property, an output is generated, which lasts a certain time. These behaviors have no memory.

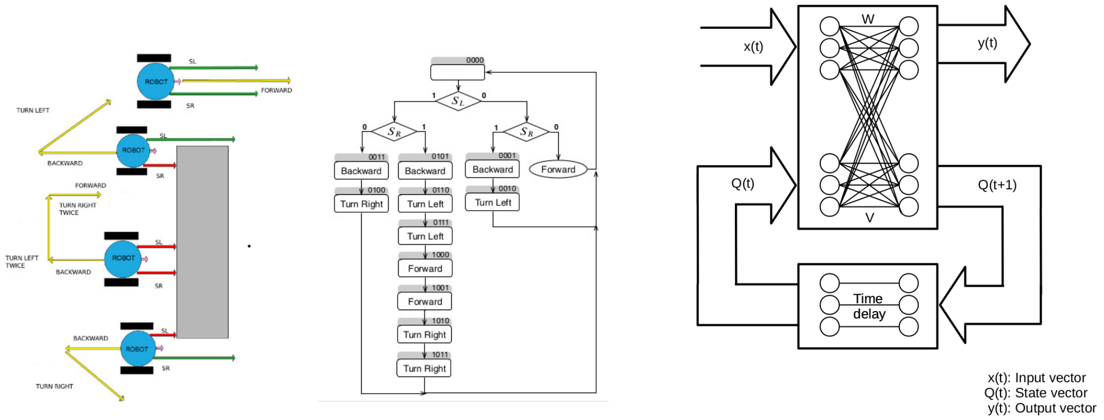
There are techniques without memory such as potential attractive and repulsive fields (Figure 5a) and neural networks (Figure 5b) or with memory like state machines algorithms (Figure 6a) or recurrent artificial neural networks (Figure 6b).



(a) Potential attractive and repulsive fields

(b) Reactive behaviors without memory using neural networks.

Figure 5: Techniques without memory



(a) Reactive behaviours using State Machine Algorithms.

(b) Reactive behaviours using recurrent artificial neural networks.

Figure 6: Techniques with memory

The SR can be combined in different structures by connecting them in parallel by adding the output of each of them or selecting one of the outputs using an arbiter (Figure 7).

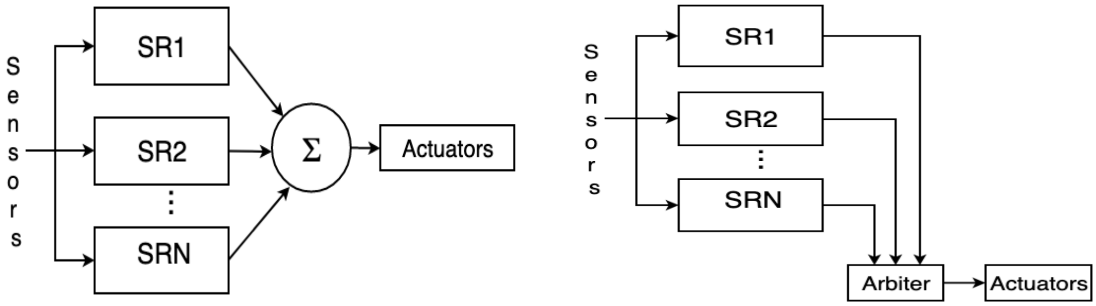


Figure 7: Combining SR structures

2.3 Probabilistic Architectures

Probabilistic robotics architectures are based on the concept that the robot's sensing of the environment and its movements are dependent on random variables, which can be manipulated using probabilistic concepts, for instance: Hidden Markov Models (HMM), Particle Filters, Markov Decision Processes, etc (Figure 8).

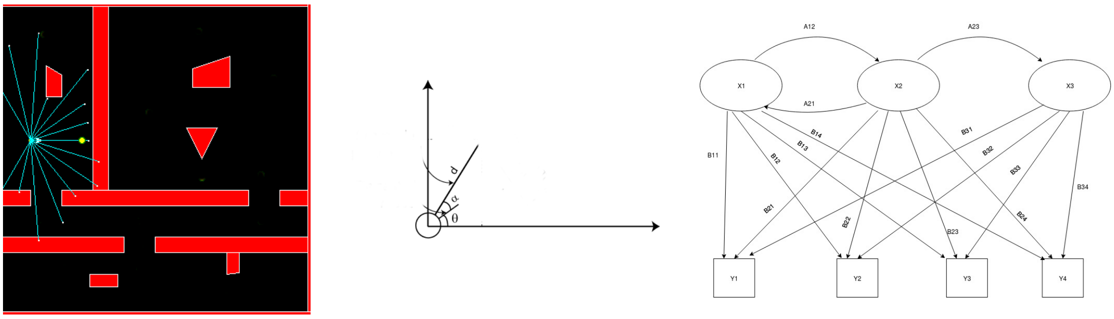


Figure 8: Using probabilistic architectures

2.4 Hybrid Architectures

On these kind of architectures, traditional, reactive and probabilistic architectures are combined to replace the deficiencies of each of them (Figure 9).

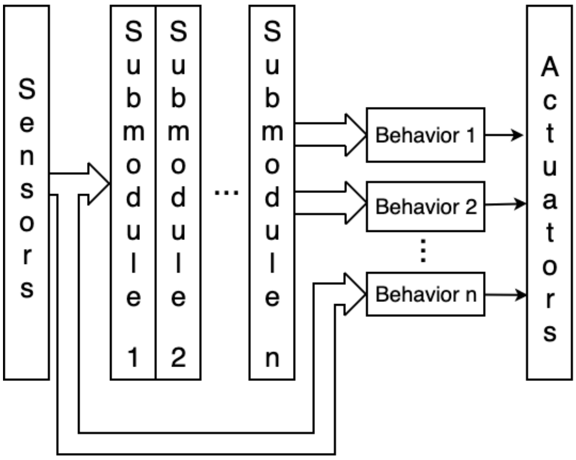


Figure 9: Hybrid robotics architectures.

3. Virtual Real Robot (ViRBot)

The VIrtual and Real roBOt sysTem (ViRBot) [4] is a hybrid robotics architecture whose goal is to operate autonomous robots that carry out daily service jobs in houses, offices and factories. ViRBot is being used by our robots: Justina (Figure 10a), a service robot created at our laboratory (bio-robotics lab UNAM) and Takeshi (Figure 10b), a Human Support Robot (HSR) from Toyota.

As service robots, Justina and Takeshi should be able to perform the following tasks:

- Autonomous navigation and unknown and dynamic obstacle avoidance.

(b) Robot Takeshi (Toyota HSR)

(a) Robot Justina.

Figure 10: Our Laboratory's Robots

- Object placement and object recognition without artificial marks.
- Person detection, recognition and tracking.
- Speech recognition and natural language understanding (NLU).
- Autonomous mapping.

ViRBot divides the operation of the service robot into four general layers: Input, Planning, Knowledge Management, and Execution (Figure 11). Each layer combines traditional, reactive, and probabilistic techniques to solve the tasks required from a service robot, such as safe and robust autonomous navigation in dynamic environments; obstacle avoidance; object detection, recognition, and manipulation; people detection, recognition, and tracking; and human–robot interaction via natural language [5]. By combining symbolic AI with digital signal processing techniques, a good performance in a service robot has been obtained.

In the following sub sections, we will describe most relevant VIRBOT modules categorized by layer.

3.1 Input Layer

This layer process the data from the robot's internal and external sensors (in a series of modules), they provide information of the internal state of the robot, as well as, the external world where the robot interacts. Some of our robots designs have lasers, sonars, infrared, microphones and stereo and RGB-D cameras.

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 applied to create models of objects, places

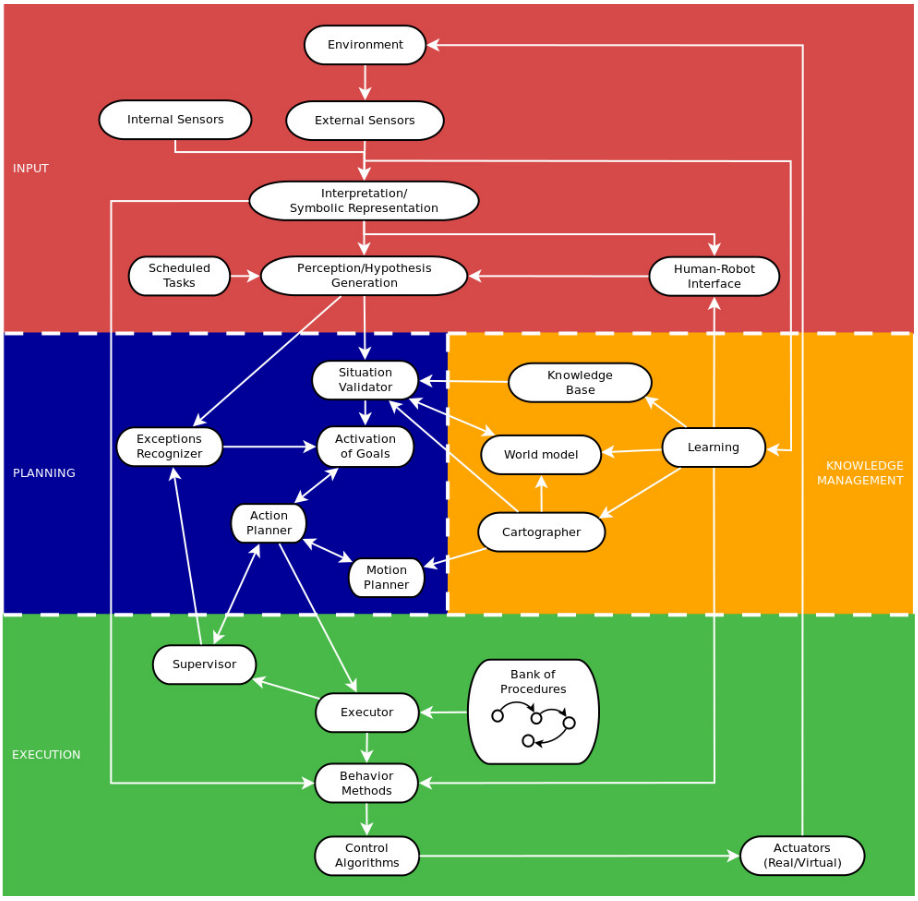


Figure 11: ViRBot architecture.

and persons that interact with the robot. To recognize objects, people and places using RGB-D cameras, vision systems that are robust to partial occlusions, scale and rotation changes are used.

The Human/Robot Interface subsystem in the ViRbot architecture is responsible of recognizing and processing voice and gesture commands, meaning that communication between the user and the robot can be through voice and manual gestures, the robot responds using synthetic voice and simple facial expressions and is able to perform speech recognition (Figure 12).

The Human/Robot Interface has tree modules: Natural Language Understanding, Speech Generation and Robot's Facial Expressions. The natural language understanding module finds a symbolic representation of spoken commands given to a robot [6]. It consists of a speech recognition system coupled with Conceptual Dependency [7] techniques (Conceptual Dependency is a theory developed by Schank for representing meaning). For instance, given these phrases:

- ***"Robot, give the newspaper to Dad"***
- ***"Please bring me the newspaper that is there"*** (Dad is giving the order to the robot)

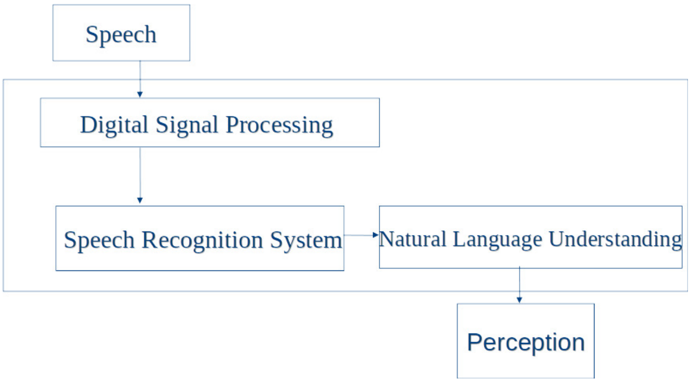


Figure 12: Speech Recognition System.

Both phrases are represented using the following Conceptual Dependency primitive:

(ATRANS (ACTOR Robot) (OBJECT newspaper) (TO Dad) (FROM newspaper-place))

With this symbolic representation a belief is generated, and a planner makes an action plan to achieve what the robot is being asked for. With the symbolic representation, the perception module generates a series of beliefs, that represent the state of the environment where the robot interacts.

3.2 Planning Layer

This layer is responsible of generating plans at a high level of abstraction and performing global reasoning. Beliefs generated by the perception module are validated in this module with information of the Knowledge Management layer. Once validated or recognized, a belief is considered knowledge and either stored or used to trigger the Action Planner, which will generate a plan of action or sequence of physical operations to achieve the desired goals. However, if something unexpected happens while executing a plan, the Goal Activator will be notified,interrupting the Action Planner and triggering the generation of a new plan (Figure 13).

3.3 Knowledge Management Layer

This layer involves all modules that store and provide access to the robot's knowledge. Such knowledge, which may not be symbolic, ranges from raw and probabilistic maps, to semantic knowledge of the language. The cartographer module has different types of maps for the representation of the environment (Figure 14), they are created using SLAM techniques.

In the world model module there is a representation of the objects, people and places where the robot interacts and the relationships that exist between them. There are 5 structures that represent objects, rooms, furniture, humans and the robot. All this information is updated by the input layer and by the actions of the robot.

Example: ***(human (name Mother) (room studio) (zone couch) (objects book) (pose 1.8 2.0 0.5) (locations main-bedroom)***)

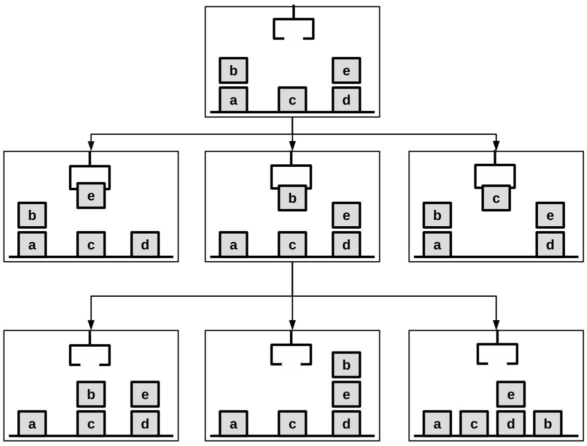
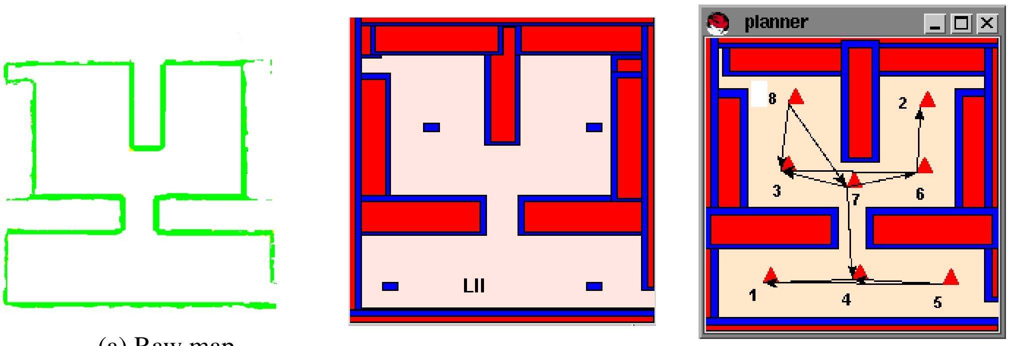


Figure 13: Action Planning.



(a) Raw map.

(b) Symbolic map.

(c) Topological map.

Figure 14: Types of maps

For the knowledge base module CLIPS [8] (a rule-based system from NASA) is used to represent robot knowledge by production rules, which correspond to the actions that the agent would do if certain conditions are met. For instance:

Shadow Rule{
If there are trees around the robot's path and it is a clear day then there will be a shadow in the path.
}

In the Learning module genetic algorithms and programming are used, the goal is to use an optimization algorithm such as genetic algorithms (GA), to find the best robot behaviors to avoid obstacles while they tried to reach a destination (Figure 15). On this layer also some probabilistic methods are used like Markov chains and Bayesian classifiers as well as clustering (K-means, Vector Quantization), Artificial Neural Networks and Reinforcement Learning.

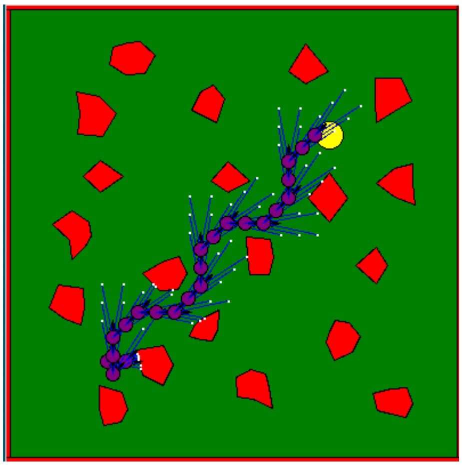


Figure 15: Genetic algorithms and programming.

The use of simulators for Robot Learning is very beneficial, for instance, with simulated images the robot is trained to navigate in new environments while it is in resting mode (Figure 16).

Figure 16: Gazebo 3D robot simulator integrated with a physics engine.

3.4 Execution Layer

This layer executes the actions and movements plans and it checks that they are executed accordingly. At its core, the Bank of Procedures encapsulates a set of hardwired functions, represented by state machines, and are used to partially solve specific problems, finding persons, object manipulation, etc. The executor uses these bank of procedures to execute a plan. Behavior methods are used to avoid obstacles not contemplated by the movements planner (Figure 17). The behavior methods are state machines, potential fields and neural networks.

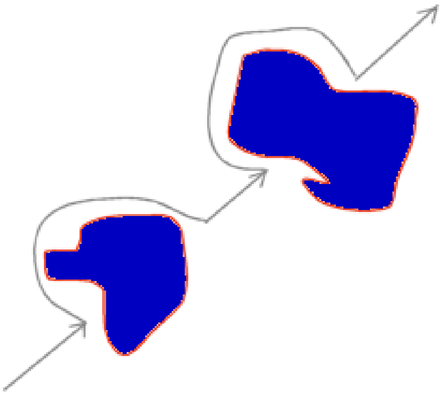


Figure 17: Obstacle avoidance.

Control algorithms, like PID, are used to control the operation of the virtual and real actuators (Figure 18).

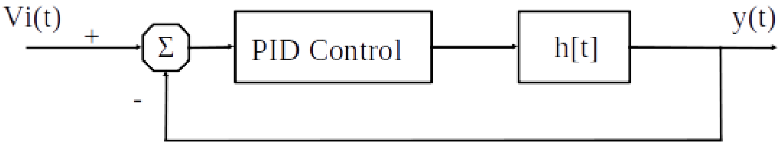


Figure 18: Control algorithms.

4. Contributions from Biorobotics laboratory members

Our laboratory members have contributed to ViRBot by developing some projects and adapting them to the architecture. Sorted by the related ViRBot layer we can mention the following ones:

- Input Layer
  - ***"Training New Objects in a Deep Neural Network Using YOLO"*** by Edgar Silva (Figure 19)
  - ***"Finding persons and their actions in the environment using deep neural networks"*** by Edgar Vazquez (Figure 20)
  - ***"Scene Classification and Understanding"*** by Hugo Leon (Figure 21)
- Knowledge Management Layer (Robot Learning using simulators)
  - ***"The Kalman filter and deep neural networks used to estimate the robot's position and orientation"*** by Diego Cordero (Figure 22).
  - ***"Gazebo 3D robot simulator integrated with a physics engine"*** by Oscar Fuentes (Figure 16).

Figure 19: Training New Objects in a Deep Neural Network Using YOLO

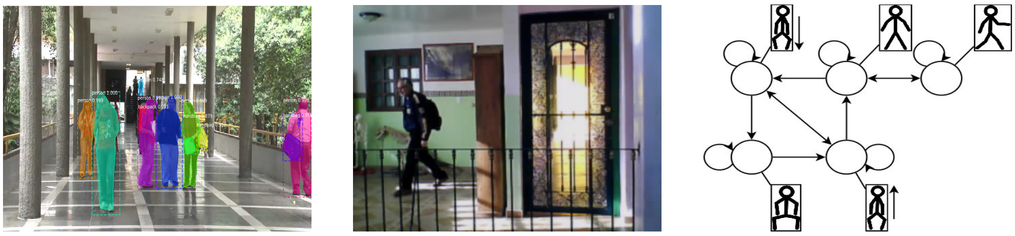


Figure 20: Finding persons and their actions in the environment using deep neural networks

(a) Indoor segmentation and support inference [9].

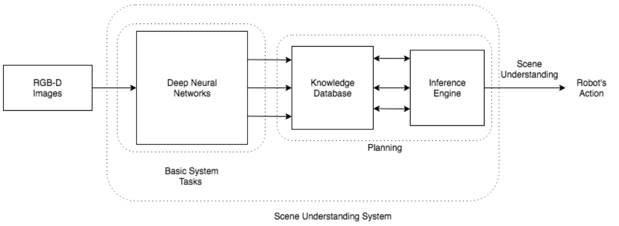


Figure 21: Scene Classification and Understanding.

- ***"The House Of inteRactions THOR"***, Allen Institute for Artificial Intelligence [10] simulator used to do reinforcement learning [11] and 3D perception for robot's navigation. Extended by Adrian Sarmiento (Figure 23)
- ***"Robot learn how to place objects under various obstacle layouts and illumination conditions using Neural Networks"*** by Angelica Nakayama (Figure 24).

Other projects are currently under design or development, for instance, "Image Synthesis with Generative Adversarial Nets (GAN)", with synthetic images the robot is trained to navigate in new environments while it is in resting mode. (Figure 25).

5. Tests and results

In our laboratory there have been various robots designed, developed and manufactured by Pumas Team members, some of the most famous are TX8, TPR8 and PACK-ITO. The most recent developed robot is Justina (Figure 10a) which is been under re-design since 2012, having won some important competitions over these years.

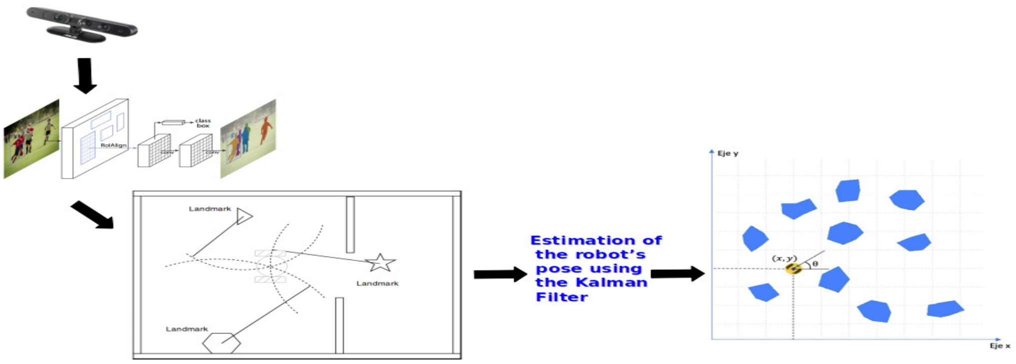


Figure 22: The Kalman filter and deep neural networks used to estimate the robot's position and orientation.

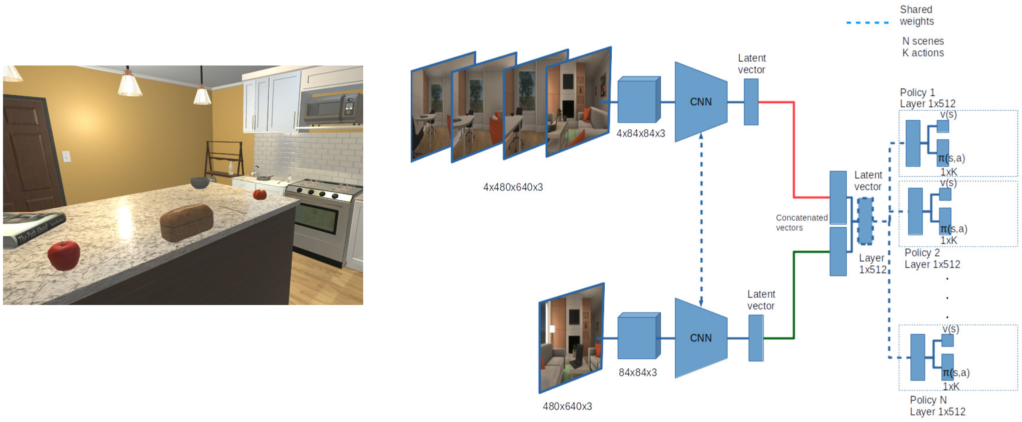


Figure 23: The House Of inteRactions (THOR) [10] [11]

We also have been working with a Toyota Human Support Robot that we call "Takeshi" and it has been at our lab since february 2017. All our robots have been participating on different competitions and events having better results every year as summarized on table 1.

6. Conclusion

ViRBot implements an architecture that allows the development of artificial intelligence systems in a flexible and modular way, facilitating the programming of new applications focused on the operation of mobile robots in service environments for human beings. This architecture allows the creation of high-level instructions that are independent of modifications in the operation of modules in lower layers, this allows a flexible and asynchronous development in student work teams. Currently, the integration of interoperability with ROS facilitates the increase of functionality in robots, making this architecture a contribution for the community interested in developing their proposed robotic agents.

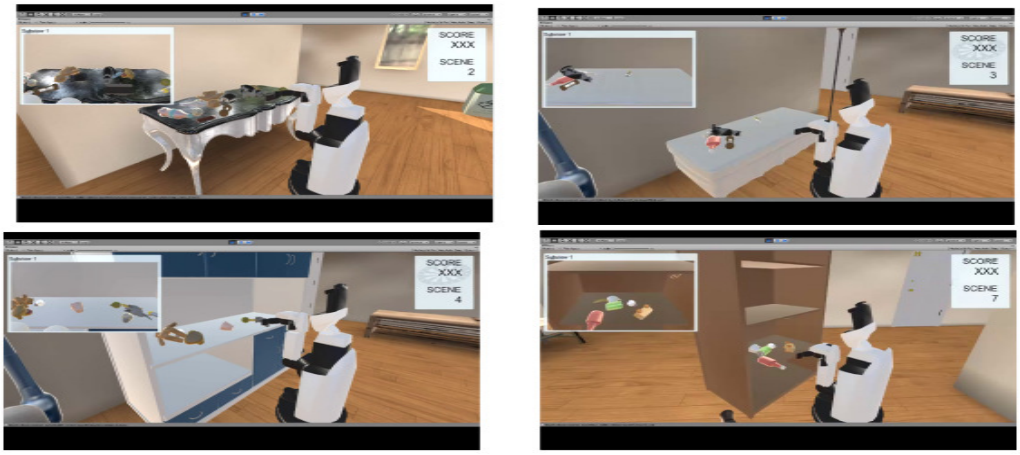
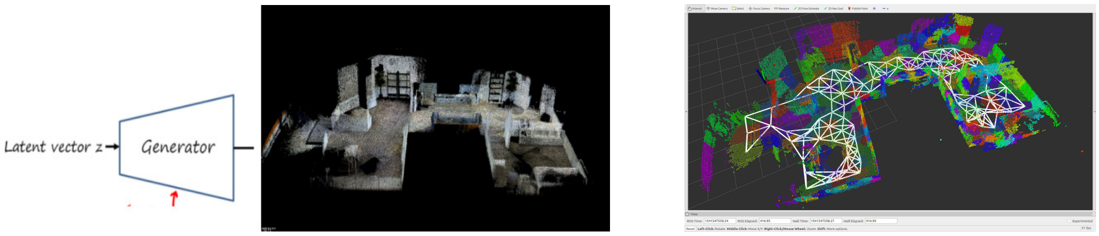


Figure 24: Learn how to place objects.



(a) The generator creates synthetic images of the environment.

(b) Training Path.

Figure 25: Image Synthesis with Generative Adversarial Nets (GAN)

References

[1] T. Fong, I. Nourbakhsh and K. Dautenhahn, *A survey of socially interactive robots*, *Robotics and Autonomous Systems* **42** (2003) 143.

[2] "Robocup at home league." `https://athome.robocup.org/`.

[3] M. . RoboCupAtHome, "Arena of robocup 2018." `https://github.com/RoboCupAtHome/Montreal2018`.

[4] J. Savage, A. LLarena, G. Carrera, S. Cuellar, D. Esparza, Y. Minami et al., *Virbot: A system for the operation of mobile robots*, in *RoboCup 2007: Robot Soccer World Cup XI*, U. Visser, F. Ribeiro, T. Ohashi and F. Dellaert, eds., (Berlin, Heidelberg), pp. 512–519, Springer Berlin Heidelberg, 2008.

[5] J. Savage, D. A. Rosenblueth, M. Matamoros, M. Negrete, L. Contreras, J. Cruz et al., *Semantic reasoning in service robots using expert systems*, *Robotics and Autonomous Systems* **114** (2019) 77.

[6] J. Savage, A. Weitzenfeld, F. Ayala and S. Cuellar, *The use of scripts based on conceptual dependency primitives for the operation of service mobile robots*, in *RoboCup 2008: Robot Soccer World Cup XII*, L. Iocchi, H. Matsubara, A. Weitzenfeld and C. Zhou, eds., (Berlin, Heidelberg), pp. 284–295, Springer Berlin Heidelberg, 2009.

[7] R. Schank, *Conceptual Information Processing*. North-Holland Publishing Company, 1975.

[8] L. B., *Clips reference manual version 6.0.*, tech. rep., NASA-Johnson Space Center, 1994.

Year	Event	Venue	Award
2007	RoboCup 2007	Atlanta, USA	3rd Place
2015	RoCKIn 2015	Lisbon, Portugal	2nd place Object Perception 2nd place Speech Understanding
2017	RoboCup 2017	Nagoya, Japan	4th place @Home League Best in speech recognition and Natural Language Understanding
2018	RoboCup 2018	Montreal, Canada	2nd place @Home DSLP 2nd place @Home OPL
	IROS 2018	Madrid, Spain	1st place @TBM1: Getting to Know My Home 1st place TBM2: Welcoming Visitors 1st place TBM3: Catering for Granny Annie's Comfort 1st place TBM4: Visit My Home
	World Robot Summit 2018	Tokyo, Japan	4th Place Virtual Space Category 6th place Parthner Robot Challenge
2019	RoboCup 2019	Sydney, Australia	2nd place @Home OPL

Table 1: Competitions and events results

[9] N. Silberman, D. Hoiem, P. Kohli and R. Fergus, *Indoor segmentation and support inference from rgbd images*, in *Computer Vision – ECCV 2012*, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato and C. Schmid, eds., (Berlin, Heidelberg), pp. 746–760, Springer Berlin Heidelberg, 2012.

[10] E. Kolve, R. Mottaghi, W. Han, E. VanderBilt, L. Weihs, A. Herrasti et al., *AI2-THOR: An Interactive 3D Environment for Visual AI*, *arXiv* (2017) .

[11] Y. Zhu, R. Mottaghi, E. Kolve, J. J. Lim, A. Gupta, L. Fei-Fei et al., *Target-driven Visual Navigation in Indoor Scenes using Deep Reinforcement Learning*, in *IEEE International Conference on Robotics and Automation*, 2017.