2020 SinfonIA Pepper Team Description Paper

Carolina Higuera Eyberth Rojas Fabián Pérez
Carlos A. Quintero Saith Rodríguez Uniandes Members
Santo Tomás Members Unimagdalena Members
Bancolombia Members

October 27, 2019

Abstract. In this paper we introduce the SinfonIA Pepper Team for RoboCup Social Standard Platform League (SSPL). Our team is a collaborative effort between Colombian institutions whose main objective is to lead the development of robotics and artificial intelligence in our country. Our current research interest is to improve the human robot interaction by a) customer service applications based on reliable face detection and person identification and b) entertainment services offering two-player games, based on reinforcement learning. In this paper, we also show our results implementing a custom version of the Stage 2 tests *where is this* and *stickler for the rules*. We use our Pepper robot to demonstrate several abilities such as human-robot interaction, person recognition, localization, navigation and task planning capabilities.

1 Introduction

The SinfonIA Pepper Team is a collaborative effort between four different institutions in Colombia, i.e., Universidad de los Andes, Universidad Santo Tomás, Universidad del Magdalena and Bancolombia, whose main objective is to lead the education and research on social robotics, machine learning and artificial intelligence in the country.

Our main interest with this educational initiative is to promote the usage of robots and systems capable of providing services to people, in areas such as information retrieval, guidance and entertainment. With this demonstration, we intend to sensitize the Colombian society about the advantages of using social robots and AI systems in people's daily lives. For this purpose, we have defined three research lines to develop the self-learning ability of the robot. Which are computer vision, human-robot interaction and decision-making systems.

In order to test our research outcomes, we have chosen RoboCup@Home Social Standard Platform League as test-bed. This competition provides our team with an excellent opportunity to learn and enrich our work from others in the RoboCup community, and to contribute to the international scientific advances in the developing of social robotics abilities for their use in future real world applications.

2 Background and previous experiences in RoboCup

While the SinfonIA Pepper team and the collaborative work between the four institutions started recently (we have been working for less than a year now), the team consists of people and universities that have an academic trajectory in research groups and that have also been part of the RoboCup community for several years. Some of the team members were part of the STOx's team, which participated in the Small Size League (SSL) from 2011 to 2017, becoming the first and only latin-american team to enter the league's hall of fame up to date [2,3,4,5]. However, our most important involvement with the @Home League is the 2019 participation in the Social Standard Platform League (SSPL), where we made our debut as the SinfonIA Pepper Team [1].

At Sydney, in this year's participation, we were able to show some of Pepper's autonomous skills such as: speech interaction (recognition and speaking), object and person recognition and simple navigation, where we finished 4th tied with the KameRider and LiU@HomeWreckers teams with 350 points. Even though, at that time, we had our Pepper robot for less than 6 months, we managed to present the tools that we had been developing as well as the first iteration of our architecture. Nonetheless, we found that we could not display our full potential due to failures in some of these skills that did not let us accomplish the tasks that we were aiming for.

More specifically during the competition we attempted the *receptionist* and *clean up* tests, where we encountered problems mainly in navigation and speech recognition (e.g., the robot was not able to go to the place it was supposed to or the robot did not understand what the operator was saying). Furthermore, these problems were the result of our limited experience with the robot and the insufficient time we had to develop more robust tools. However, thanks to the knowledge that we acquired from this competition and from other SSPL teams, we have been able to improve upon our failures. The following sections describe in further detail our current research and the developments that our team have made to make our architecture and tools more robust and better overall.

3 Achievements and Research Interests

Regarding the problems we encountered in Robocup 2019, we focused our work on improving the tools that were not robust enough. In first place, all the architecture of the system was reformulated and we are currently working on the implementation of redundant tools. Moreover, we were able to install the Navigation Stack tool in our robot, which allows us to control easily the autonomous navigation of the robot. Besides, we have decided to incorporate the use of a Docker into our development environment after discussions with other teams in the league. This has significantly improved our development efficiency and the integration with new tools.

Our major interest research interest is to develop data-driven models for social robots that allow them to improve their behavior as they interact with humans. We are interested in providing social robots with the ability to automatically learn new ways to interact with humans without being explicitly programmed to do so. We also aim to come up with new methods to improve the robot's understanding of humans. We have tackled some of these problems in projects, as described below.

3.1 Customer service

This project was formulated to make the university community acquainted with the services a robot can provide in an academical environment. Here by, the project is focused on the contributions Pepper could have in the library of a university. With these means, the robot must give the students a good assistance when entering the library.
To achieve so, the robot has to identify, recognize and interact with the students that enter the library. First of all, the identification of people was done with an algorithm called Viola-Jones, which is based on Adaptive Boosting. With this technique the robot is capable of identifying people in real-time and counting how many people are nearby. Secondly, the method that was implemented to recognize people was Eigen Faces, which is based on principal component analysis (PCA) to pre-process the images, and support vector machines (SVM) to classify them. Moreover, to simulate a real life situation, a database of 30 people was created, in order to recognize them at any time. While an unknown person can be memorized by the robot during the interaction with him or her; while there is an interaction between the robot and the student, the robot must take approximately 30 pictures to learn the face of the student, and learn to connect those pictures and features with his or her name. For future work, once the robot identifies who is the person with which it is interacting, it must be able to talk to the student in a fluent and natural way, maybe even guiding the student inside the library and giving him or her information about the books or material that can be loaned, or if the student owes some material in the library, to remind him or her to return it as soon as the student is done working with it.

3.2 Board games with reinforcement learning

In contrast with the Customer Service project, this project is focused on the entertainment that the robot can give to any person. Through reinforcement learning, Pepper must learn how to win his adversary in board games such as Four in a Row or Othello. To achieve this, a whole user interface was developed, which is projected in Pepper's tablet, in which the user who wishes to play has the chance to select the game he or she wants to play and the difficulty of the game. Through the game, the robot interacts with the person through animations and dialogues to have a better competitional environment between the human and the robot. In future work, the emotions and reactions of Pepper's adversary will be studied; in order to guarantee that the person is having a good time playing, so the performance of the robot can be modified. This guarantees that the entertaining people goal can be achieved with success.

3.3 Multimodal person detection based on laser and depth data

We developed a system to detect people legs using laser data. Our approach used raw laser data to detect if there are people legs or not within the robot boundaries. This detector allows the robot to keep track of the person when guiding him or her. The detector architecture is intended to be platform independent, which is why the system was tested on a Pepper robot and on a Pioneer P3-DX. This was the research project that was conferred in our team's poster for Robocup 2019. Nevertheless, we used the knowledge adquired in this area to develop a project for a master's subject, called Learning Machine Learning, with which the students had to work with raw laser data to identify people nearby the robot.

4 Case study: Pepper's Where is this and Stickler for the rules

In order to illustrate the abilities developed by our team, we decided to make a customized version of *Where is this* and *Stickler for the rules* from the 2019 edition rulebook's Stage 2 tests. In the *Where is this* test, Pepper has to provide directions to a student in order to find the office of the professor that he or she is looking for, and to offer its guidance service. In the *Stickler for the rules* test, Pepper has to enforce the library rules, such as no eating inside and silence to be maintained. To accomplish these challenges, we had to handle several tasks, including person identification and recognition, object recognition, speech interaction for gathering orders, noise detection, self-localization and path-planning and navigation with obstacle avoidance.

We are working on improving our general architecture, in order to provide basic abilities as tools, so that later the robot's general decision-making system will be able to call them. We are also applying the recommendations that were given to us by other teams during RoboCup 2019. For instance, having backup or redundant services, as well as using the open tools provided in other teams repositories. In figure 1 we present our current architecture to solve all the previously mentioned tasks. The core of the solution lays in the Scheduler implemented in ROS (Robotic Operating System). Also, the Robot Toolkit provides methods and functions that allows Pepper's tools access to all the robot's sensors, cameras and actuators.

Pepper's tools cover interaction, navigation and retrieval information tools. Some of them are based on machine learning models, such as speech, object and person detection and recognition. In order to make the whole system reliable and fault tolerant, we are working on the implementation of redundant tools. This means that some services are going to run locally in the robot, with light models, and simultaneously, they are going to run in the cloud with more accurate models.

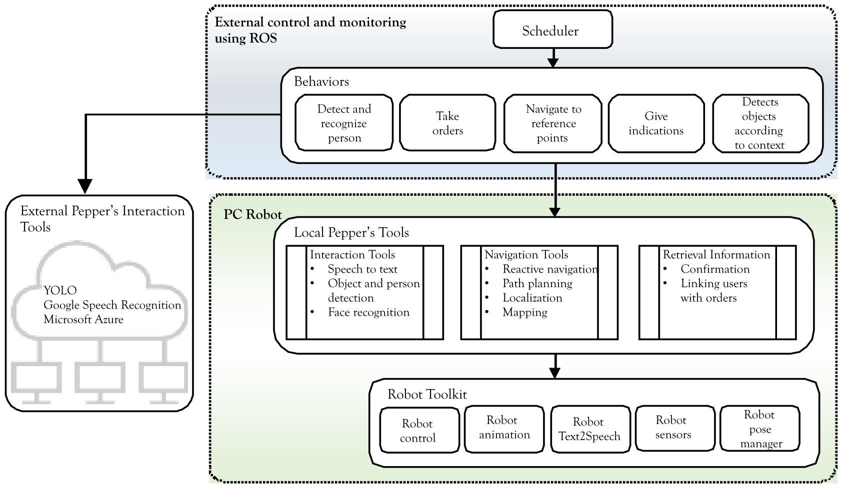

Fig. 1: General solution architecture for Pepper

The application layer contains general behaviors that will allow us to build our case studies, such as reaching a specific location, identifying a person, understanding orders, among others. The Scheduler is our decision-making system. It gathers the needed behaviors in a SMACH state machine and decides when and which function to execute, reaching the construction of a successful Where is this and Stickler for the rules. It is worth mentioning that all the tools and behaviors were developed with the objective of using them for future applications in the universities, especially for customer service.

4.1 Speech interaction

This ability allows our robot to reach out to people in a kind way and to get information from the users. Below, we explain in further detail the speaking skill and the speech recognition.

Speaking skill: To allow our robot to speak, we are using the robot interaction functionalities of our robot toolkit, which interfaces with the NAOqi driver. We implemented an algorithm capable of analizing all the possible answers and obtaining the desired words. To integrate all these possible options and to diversify the answers for each user, we implemented a method capable of choosing a different and appropriate response for the interaction stage.

Speech recognition: We are working on using pocketsphinx as our local speech recognition to get a a human speech transcription. However, in order to broaden the possibility to achieve a more fluent interaction, we are using the Google Cloud API Speech-To-Text externally, which allows the robot to transcript words of the person who is talking into a text file. We developed an algorithm to split and analyze all the words contained in the original speech. Therefore, according to the meaning and position of the word, the robot is able to change its answer. The speech recognition works together with the speaking skill, in order to communicate in a successful way with a client. The algorithm is also customized for our Where is this and Stickler for the rules challenges. In order to find the source of noise, we have created a service able to publish the azimuth, elevation and energy level, which allow us to find the person's position.

4.2 Visual skills

Two important visual skills are required for social robots, namely person recognition and object detection. For this purpose, we used the two robot's on-board 2D cameras to capture the images. However, since the robot and the people are in constant motion during the interaction, some of the captured images may be blurry and not adequate for a proper learning and detection of faces and/or objects. Anyhow, we already had implemented the methodology in [**?**,**?**], to compute the focus level and to select the best images from a set of pictures taken by the robot and use them to learn new faces or objects.

Person recognition and characterization: As an external tool for the face recognition task we are using the Face Recognition ROS package, which is a cloud Azure Cognitive Service. This allows Pepper to identify, recognize and extract some characteristics of a person from a single photo. Even so, to provide this function locally, we have implemented a method for people identification, and another one for people recognition. The decision about which methods to implement was based on the knowledge acquired through the years on some Machine Learning techniques. In the first place, the people identification was based on Viola-Jones, a widely used method for real-time object and people identification, based on the Ada-Boost technique. On the other hand, the algorithm for people recognition was based on Eigen Faces, which analyses the image of the person that was detected by the robot, and through PCA and SVMs the person can be labeled with his or her name, in case the robot knows this person, and in the opposite case, the person will be labeled as unknown.

Object detection and recognition: We are using an interface called LightNet, which is based on the object detection models of YOLO. This tool uses a picture of the objects taken by the robot that is then used to create labelled bounding boxes on the image. To obtain an accurate result, we are working on retraining a pre-trained YOLO network, reducing the number of classes or objects that the robot has to recognize. Through the training process, the weights of the network are modified in order to get a higher detection score. We are using prior information and domain knowledge for retraining the network.

4.3 Localization and Navigation

Using the recommendations and tools provided by the UTS Unleashed team in their repository, now we are able to run the navigation ROS package in the robot. Currently, we are working on improving the localization and adjusting the parameters for the global and local path planners.

Localization We are using the amcl ROS package for localizing the robot in a known map, based on the readings of the laser scan and transform tree messages. However, we have noticed that when the robot has to navigate a long path, the output pose estimation of the amcl node differs from the real location in the environment. As an approach for improving the localization ability of the robot, we are working on running rtabmap ROS node internally. The hypothesis is that once we have a 3D point cloud of the environment, it would be easier for the robot to localize itself when a loop closure is found against the RGB bag-of-images taken during the mapping.

Navigation We are running the navigation stack inside the robot, providing the odometry, the laser scan readings and the transformation tree of the robot. To improve the performance, we are working on methods for cleaning the global costmap periodically. The reason behind this is to eliminate the navigation's current map contamination with obstacles that do not exist. We think that this is due to the inaccuracies that eventually occur in the location of the robot. Additionally, we are working on the reduction of the complexity of the tf tree, so we can have a better transfer rate, especially for monitoring purposes.

5 Conclusions and future work

Despite the fact that our team has only nearly a year of experience in social robotics research and development, our dedication scientific interest on the subject has allowed us to learn and to develop, in a short period of time, useful tools and interfaces that propose a custom approach for the human robot interaction in the context of social robots. We see RoboCup @Home League as a remarkable opportunity to test our developments and to keep learning and sharing experiences with this scientific community.

It is worth saying that the new tools related to human-robot interaction are being developed by our students and researchers. This is important for us due to the fact that our team's main objective is to lead the development of robotics and artificial intelligence in our country, and a way to achieve this goal is by training the future engineers in this field.

Future work will include research oriented to human-robot interaction in the field of natural language, in order to enable a natural and fluent conversation. Also, we would like to explore other possibilities for the robot's navigation's ability, such a reinforcement learning approach for moving in unknown environments.

References

1. C. Quintero, C. Higuera, E. Rojas, F. Pérez, S. Rodríguez, Uniandes members, Santo Tomas members, Unimagdalena members and Bancolombia members. Alianza SinfonÍA 2019 Team Description Paper. Sydney, Australia 2019. Available for download in https://alianzasinfonia.net.co/resources.html.
2. S. Rodríguez, E. Rojas, K. Pérez, J. López, C. Quintero, and J.M. Calderón. STOx's 2014 Extended Team Description Paper. Joao Pessoa, Brazil 2014. Available for download in http://robocupssl.cpe.ku.ac.th/_media/robocup2014:etdp:stoxs_2014_etdp.pdf
3. S. Rodríguez, E. Rojas, K. Pérez, C. Quintero, H. Báez, O. Peña, J.M. Calderón. STOx's 2016 Extended Team Description Paper. Leipzig, Germany 2016. Available for download in http://wiki.robocup.org/images/0/0d/Small_Size_League_-_RoboCup_2016_-_TDP_STOx%27s.pdf
4. S. Rodríguez, E. Rojas, K. Pérez, C. Quintero, O. Peña, A. Reyes and J.M. Calderón. STOx's 2017 Team Description Paper. Nagoya, Japan 2017. Available for download in http://wiki.robocup.org/images/d/d7/Robocupssl2017-final16.pdf
5. S. Rodríguez, E. Rojas, K. Pérez, C. Quintero, O. Peña and A. Reyes. 2018 STOx's Extended Team Description Paper. Montreal, Canada 2018. Available for download in http://wiki.robocup.org/images/6/67/Robocupssl2018-STOxs.pdf
6. Bouchard S., Lachapelle F., Lebel P. and Verret B. (2018) ULtron 2018 Team Description Paper, RoboCup 2018.
7. Rodríguez S., Quintero C.A., Pérez A.K., Rojas E., Peña O., De La Rosa F. (2018) Methodology for Learning Multimodal Instructions in the Context of Human-Robot Interaction Using Machine Learning. In: Brito-Loeza C., Espinosa-Romero A. (eds) Intelligent Computing Systems. ISICS 2018. Communications in Computer and Information Science, vol 820. Springer, Cham
8. M.A Williams, B. Johnston, S. Pfeiffer, J. Vitale, J. Clark, L. Kang, D. Ebrahimian, T. Vu, S. Leong, R. Billingsley, M. Tonkin, S. Herse, S. Alam, S. Leela, K. Chand and S. Ojha. UTS Unleashed! 2019 Team Description Paper, Robocup 2019.
9. Y. Qing Wang. An Analysis of the Viola-Jones Face Detection ALgorithm. CMLA, ENS Cachan, France 2013.
10. M.Turk, A.Pentland. Eigenfaces for Recognition. Journal of Cognitive Neuroscience, Volume 3, Number1. Vision and Modeling Group, Massachusetts Institute of Technology 1991.
11. R. S. Sutton and A. G. Barto. Reinforcement Learning: An Introduction. The MIT Press. Cambridge, Massachusetts. London, England 2017.

Pepper Software and External Devices

We use a standard *SoftBank Robotics* Pepper robot unit.

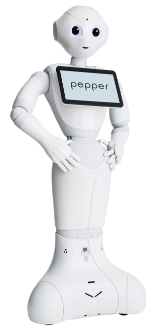

Fig. 2: Pepper Robot

Robot's Software Description

For Pepper robot we are using the following software:

- Platform: Ubuntu 9.1
- Face recognition: Based on Microsoft Azure Face API (See previous sections).
- Object recognition: Based on Lightnet and Yolo (See previous sections).
- Speech Interaction: Based on Google Speech Recognition. (See previous sections).

External Devices

Pepper relies on the following external hardware:

- Atom E3845 Quad core 1.91 GHz
- Intel HD graphics up to 792 MHz
- 4 microphones, 2 RGB HD cameras, 5 tactile sensors, touch screen on the breast

Cloud Services

Pepper connects the following cloud services:

- Object detection and recognition: Yolo and Lightnet.
- Face Recognition: Microsoft Azure Face API.
- Speech recognition: Google Speech Recognition Service.

Team Members

Uniandes	**Santo Tomás**	**Unimagdalena**	**Bancolombia**
Adelaida Zuluaga	Armando Mateus	Christian Carpio	Diana Arismendy
César Daniel Garrido	Bryan Betancur	Dilan David Solar	Eduardo González
Jorge Sebastian Mora	Camilo Camacho	Ervinson Plata	Estefany Montoya
Juan David García	Diego Ibáñez	Eynerth Rojas	John Alexander
Juan José García	Lina Plazas	José Ricardo Romo	Manuel Ríos
Nicolás Rocha	Nicolás Garzón	Mario Acuña	Óscar Salinas
	Sindy Amaya		Oswaldo Peña