CAMBADA@Home'2013: Team Description Paper

J. Cunha, J. L. Azevedo, M. B. Cunha, L. Ferreira, P. Fonseca, N. Lau, C. Martins, A. J. R. Neves, E. Pedrosa, A. Pereira, L. Santos, A. J. S. Teixeira

Transverse Activity on Intelligent Robotics
IEETA/DETI – University of Aveiro
3810-193 Aveiro, Portugal

Abstract. This paper presents the CAMBADA@Home team for the purpose of integrating the qualification materials for the RoboCup@Home competition in RoboCup 2013. The CAMBADA@Home team was created in 2011 and builds on the success of the CAMBADA robotic soccer team which competes in the Middle Size League. Based on the lessons learned from the last two participations, the team is focusing this year on completing the interaction of the ROS framework, on the conclusion of the new platform with the conclusion of manipulators and robot/human communication devices and on the robot/human iteration software modules.

1 Introduction

The CAMBADA@Home[1] is the University of Aveiro RoboCup@Home team. The project was created in January 2011 following the team past experience in the CAMBADA [1] robotic soccer team. The development of the CAMBADA soccer team started in 2003 and a steady progress was observed since then.

CAMBADA has participated in several national and international competitions, including RoboCup world championships (5^{th} place in 2007, 1^{st} in 2008 and 3^{rd} in 2009, 2010 and 2011), the European RoboLudens, German Open (2^{nd} in 2010) and the annual Portuguese Open Robotics Festival (3^{rd} place in 2006, 1^{st} in 2007, 2008, 2009, 2010, 2011 and 2012). Moreover, the CAMBADA team achieved excellent results in the mandatory technical challenge of the RoboCup MSL: 2^{nd} place in 2008 and 1^{st} place in 2009 ans 2012 along with the 1^{st} place in the free challenge in 2011 and a 2^{nd} place in 2012.

The CAMBADA@Home team is comprised of students from the Department of Electronics, Telecommunications and Informatics and researchers from the Institute of Electronics and Telematics Engineering of Aveiro. The team participated in the last RoboCup@Home edition, where, for the first time, it performed according its level of expectations. This two experiences, however, also provided

[1] CAMBADA is an acronym of Cooperative Autonomous Mobile roBots with Advanced Distributed Architecture.

a valuable insight into future research directions and to the level of work required to achieve higher aims. This means that, starting in 2014, team is motivated to achieve the level of performance of the top teams in the competition.

In the following section we shortly present the robotic platform which is still under construction, and that will be the test bed to carry out future research as well as future participations in the Robocup@Home. Section 3 discusses the new developments in human detection, identification and tracking for tasks such as “Follow me” and “Who is Who”. Section 4 discusses the future developments in the field of high-level reasoning. The Human-Robot Interaction system is presented in section 5. Finally, section 6 concludes the team description paper.

2 New Robot Platform

The new CAMBADA@Home platform is designed as a three layer mechanical/electronicall platform which can accommodate in an effective way the number of sensors and actuators needed to perform the RoboCup@Home challenges.

Bottom layer adopts a four motor system which drives a symmetrical set of omni-wheels ensuring a stable and versatile motion solution. This layer, which encompasses a sandwich like mechanical structure also includes three Li-Po batteries, all the low level control hardware (adopting a distributed approach over a CAN network), a SICK LMS100 and the support for a vertical linear actuator. The top part of the sandwich allows a standard laptop to be located in the back side and has room for other needed equipment such as a switch for a robot local network. This approach lowers the centre of the gravity of the robot increasing the overall stability.

The second layer, which can be moved up and down by means of the linear actuator, represents the torso of the robot. The linear actuator allows a 450mm movement allowing the size of the robot to be continuously adjusted between 0.95m and 1.40m high.

On top of the torso, a third layer is developed upon a pan and tilt structure that holds a Kinect plus a thermal vision system. This layer symbolically represents the head of the robot. If will also include a set of small programmable oLed screens together with a carbon fiber encapsulation to allow the creation of dynamic facial expressions.

2.1 Anthropomorphic manipulation

Manipulators are crucial to provide the robot with the capability to perform manipulation tasks such as grabbing objects, openning bottles, manipulating device interfaces, among others.

The development of a set of two anthropomorphic arms, each with 7 degrees of freedom (DOF), aims to execute the above mentioned tasks, contributing to the evolution of current version of CAMBADA@Home robot.

Each robotic arm will be equipped with Dynamixel servos (MX-106R, RX-28 and RX-64) plus a BLDC motor to the shoulder. Such structure will be designed to provide a 1Kg payload at the hand, which is formed by a gripper with 2 DOF and a wrist with 3 DOF.

The final platform design is shown in Figure 1.

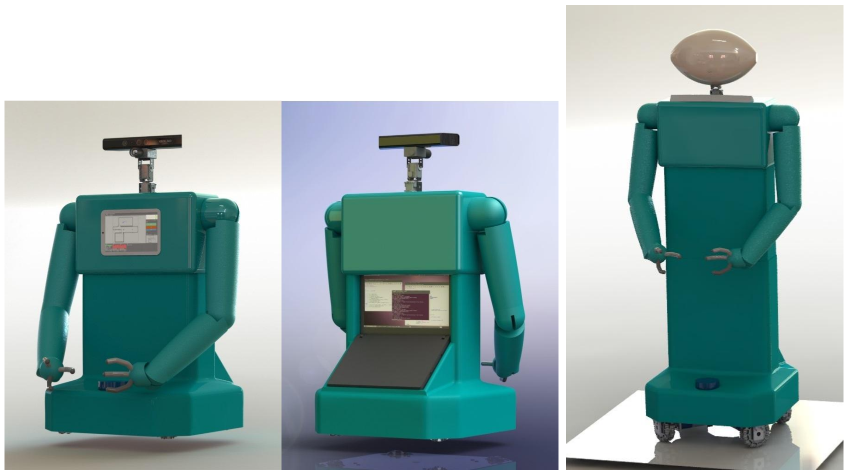


Fig. 1. The CAMBADA@Home platform.

3 Vision system

In recent years, robots started using alternatives to traditional cameras in their vision systems, due to the restrains that the best human detection algorithms using color images impose such as: stationary camera, a model of the background or environments with decent illumination. These alternatives range from Time-of-Flight cameras, to Infrared cameras, culminating in a mixture of cameras in order to minimize the disadvantages inherent to each type of camera.

The CAMBADA@Home robot will be equipped with a RGB-D and a Thermal camera for the purpose of obstacles, people and objects detection (see Figure 2). To perceive the world that surrounds the robot a Microsoft Kinect camera will be used to capture color and depth images, and a Xenics GOBI Long-wavelength infrared camera (LWIR) will provide thermal imaging. By combining these three types of images we intend to overcome many of the disadvantages inherent to each type of camera, being one of the most common the changes in lighting conditions that affect the color camera.

3.1 Obstacles detection

Besides the use of a LRF for mapping and obstacle detection, using the depth sensor it is possible to improve the occupation map of the environment, which is later used to obstacle detection. By creating a correspondence between the

(a) Color image capture (b) Depth image capture (c) Thermal image capture

Fig. 2. Example of the three types of images available with the used cameras, (a) and (b) correspond to color and depth images respectively and were captured using Microsoft Kinect RGB-D camera, (c) presents an infrared image captured by the Xenix Gobi LWIR camera

colour image and the depth image, it is also possible to add colour to the point cloud obtained from the depth sensor alone.

The kinect camera is used to detect the nearest obstacles. This detection includes selection of likely points and later filtering of those points running only over a predefined set of interest points allowing the reduction of the needed processing time. The used methods are based on the analysis of the image by column, where the farthest point is considered as wall and the closest point is considered as obstacle. A point is effectively an obstacle if its height, relative to the floor, is less than the robot height.

3.2 Objects detection

Another related challenge regarding RoboCup@HOME and the use of artificial vision by robots is its capability to be able to learn about objects and later to recognize them. To achieve this capabilities, the ability to learn new objects, preferably on-line, and to detect them when coming upon them in the world will be developed.

The depth sensor allows a segmentation method based in brightness discontinuity and depth, which can be used to detect/extract individual objects from an image. Later, the segmented objects can be processed using RGB or grayscale information. The objects can be classified and matched with the proper category. The techniques that we are developing are based on feature extraction followed by classifiers.

In pattern recognition and in image processing, feature extraction is a special form of dimensionality reduction. The idea is to transform the input image into a reduced representation set of features (also named features vector). Typically, feature detection is a low-level image processing operation. Several types of features can be taken into consideration: edges, corners, blobs, scale-invariant features, optical flow, among others.

We are exploring the use of several methods for feature extraction in the RGB and depth images, followed by the use of some type of classifiers to detect objects of interest. Among other, we are exploring the use of the following algorithms: FAST, STAR, SIFT, SURF, ORB and MSER.

3.3 Human Detection, Recognition and Tracking

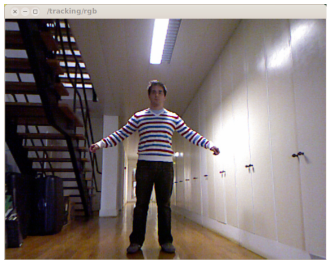
The vision system objectives are not only to help the interaction of the robot with the world of people but also with the people of the world. In order to complete many of the challenges presented in the RoboCup@Home competition, and also to enable research in the area of Human Robot Interaction (HRI) in University of Aveiro, algorithms to detect, identify and track humans are under development. To enable proper interaction between the robot and a person, it has first to be able to detect and possibly identify the person who it is communicating, if there are several people in the same room. In order to do this there is a project under development supported by state of the art methods and hardware.

For the purpose of human object detection the depth images play a crucial part, where these enable spatial object segmentation and are invariant to lighting conditions. Through the analysis of the depth image histogram the system is able to detect significant objects as local maxima, inspired by the method used in [2]. These candidate objects are then segmented from the rest of the image according to the pixel values of the image, and shape matching algorithms are used to identify human and non-human objects. The use of thermal images as a means of further confirmation of human detection is planned. In thermal imaging "the intensity value of a pixel is proportional to the thermal radiation of the scene associated with it" [3], enabling the detection of human objects.

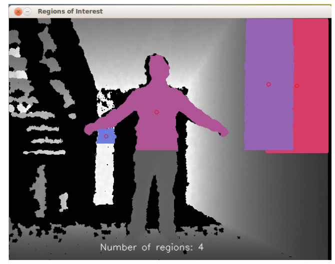
In Figure 3 its possible to see the output of the human detection process until the template matching step. There are 4 regions of interest detected, one for each local maxima on the histogram, and each is cut at a defined waist height to reduce processing time in the subsequent stages since the system only collects information from the face and upper body. These regions are then processed to enable the segmentation of the final human candidate regions that will be tested using a *head-shoulder template*.

After a successful detection the system will employ a human identification mechanism using infrared imaging. In recent years there has been many research on this area with useful comparative studies ([4], [5]) based on different techniques. Among the existing, feature based techniques such as Local Binary Patterns (LBP) Histograms, Gabor Jet Descriptors and SIFT Descriptors are among the most successful, whereas LBP Histograms presented an overall best result for HRI applications according to the studies performed in [4] and [5]. Using these features to train classifiers with the Support Vector Machines (SVM) method the system is capable of identifying people robustly.

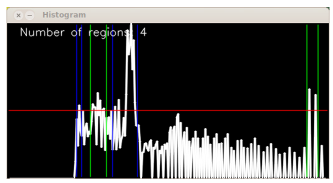
Using a novel method Hermosilla, G., et al. studied the effectiveness of vascular networks for recognition purposes in [3], proving that these preserve important discriminative information about the original thermal images and are robust to variable face expressions. The researcher chose to use SIFT descriptors and the Loncomilla and Ruiz-del-Solar (L&R) system to relate the vascular networks.



(a) Captured color image



(b) Captured color image



(c) Depth image histogram with analysis

Fig. 3. Output of the detection process without human object confirmation. (a) presents the captured color image, (b) shows the captured depth image with an overlay of the regions of interest detected, and (c) the histogram of the depth image

Further study is needed to determine if the method presented by Hermosilla, G., et al. is adaptable to LBP features and SVM classification.

Furthermore, for identification purposes, color images of the clothes of the person will be used to aid identification when the person's face is not visible (e.g. the person's back is facing the camera).

To complete the system, after a person has been detected and identified, the robot will be able to track his or her movements using motion and probabilistic models allowing for continuous detection even if the person is partially or totally occluded for short periods of time (e.g when moving behind a pillar or wall).

4 High-level reasoning

A domestic service robot requires a high level of reasoning about the environment and must be able to communicate with Humans using high level concepts. The robot should be able to autonomously learn and reason about locations and objects, i.e., it should be able to autonomously build a sematic view of its environment. Using this semantic viewpoint of the environment, the robot will efficiently navigate when asked to perform a certain task that requires more than just a motion to a location point. Our intent is to research and develop high level navigation techniques for a mobile robot in an indoor environment with focus on using a semantic view [6,7] to accomplish high level navigation tasks.

Moving from one point to another while pursuing a planned trajectory and avoiding obstacles is the basis of robotic navigation. Typical path planners such as A*, only use metric information of the environment for its calculation. A* discards useful information and it lacks the abstraction that we, Humans, use when performing a similar action. A task that requires planning and motion is usually goal-directed and it has a meaning of why to move to that specific location. For example the command *go fetch a coke* requires the robot to reason about its goal, which includes what is a coke, in which room the coke may be, plan the shortest path to coke and have a backup plan in case the coke is not where expected [8,9].

5 Human-Robot Interaction

Spoken language is a natural way - possibly the most natural - to control and process human- robot interaction. It has some important advantages: eyes and hands free; communication from a distance, even without being in line of sight; no need for additional learning for humans.

Therefore, we integrated in our mobile service robot some interaction facilities by means of three spoken and natural language processing components: an Automatic Speech Recognition (ASR) component to process the human requests (in form of command-like small sentences), a Text-to-Speech (TTS) component to generate more natural responses from the robot side, and a semantic framework (dialog manager) to control how these two components work together.

In terms of hardware, a microphone array framework with noise reduction and echo cancellation is being used. To deal with the high amount of non-stationary background noises and background speech, a close speech detection framework is applied in parallel to noise robust speech recognition techniques.

Speech recognition is accomplished through the use of CMUSphinx, an Open Source Toolkit for Speech Recognition project by Carnegie Mellon University. For this purpose, a specific grammar for command interaction is being defined, with each command-like sentence preceded by a predefined prefix (robot's name).

For robot speak-back interaction and user feedback, external speech output devices (external speakers) will be used. For the the speech synthesis component we are using the FESTIVAL Speech Synthesis system developed at the Edinburgh University. We are trying to implement some adaptation features as using the information on distance from robot to user to change dynamically the output volume, and change the TTS rate from normal to slower according to user's age.

These components will be integrated in the Olympus framework [10], developed at Carnegie Mellon University, which provides a domain independent voice based interaction system ruled by the Ravenclaw dialog manager [11]. With the transition to ROS we are adapting the Olympus framework by implementing its tools as ROS nodes.

6 Conclusion

This paper presents an overview of the current stage of development of the CAMBADA@home team. For this year the team focused on a series of new developments. Last year the team presented the basis for a new hardware platform and adopted the ROS framework. This year the team is focused on building on the previous developements. We are completing the hardware platform and building on developing better human detection, higher level reasoning layers using semantic information and improving speech-based human-robot interaction.

References

1. Neves, A., Azevedo, J., B. Cunha, N.L., Silva, J., Santos, F., Corrente, G., Martins, D.A., Figueiredo, N., Pereira, A., Almeida, L., Lopes, L.S., Pedreiras, P.: 2. In: CAMBADA soccer team: from robot architecture to multiagent coordination. I-Tech Education and Publishing, Vienna, Austria (In Vladan Papic (Ed.), Robot Soccer, 2010)
2. Guan, F., Li, L., Ge, S.S., Loh, A.P.: Robust human detection and identification by using stereo and thermal images in human robot interaction. International Journal of Information Acquisition **4**(2) (2007) 161–183
3. Hermosilla, G., Loncomilla, P., del Solar, J.R.: Thermal face recognition using local interest points and descriptors for HRI applications. In: RoboCup 2010: Robot Soccer World Cup XIV. Volume 6556 of Lecture Notes in Computer Science., Springer Berlin Heidelberg (2011) 25–35
4. Hermosilla, G.: Face recognition using thermal infrared images for human-robot interaction applications: a comparative study. In: Robotics Symposium (LARS), 2009 6th Latin American, Ieee (October 2009) 1–7
5. del Solar, J.R., Verschae, R., Correa, M.: Recognition of faces in unconstrained environments: A comparative study. EURASIP Journal on Advances in Signal Processing **2009**(iv) (2009) 1–20
6. Nüchter, A., Hertzberg, J.: Towards semantic maps for mobile robots. Robotics and Autonomous Systems **56**(11) (August 2008) 915–926
7. Ranganathan, A., Dellaert, F.: Semantic modeling of places using objects. In: Proceedings of the 2007 Robotics: Science and Systems Conference. (2007) 27–30
8. Fikes, R.E., Nilsson, N.J.: Strips: a new approach to the application of theorem proving to problem solving. In: Proceedings of the 2nd international joint conference on Artificial intelligence. IJCAI'71, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc. (1971) 608–620
9. Carbonell, J., Etzioni, O., Gil, Y., Joseph, R., Knoblock, C., Minton, S., Veloso, M.: Prodigy: an integrated architecture for planning and learning. SIGART Bull. **2**(4) (July 1991) 51–55
10. Bohus, D., Raux, A., Harris, T.K., Eskenazi, M., Rudnicky, A.I.: Olympus: an open-source framework fro conversational spoken language interface research. In: proceedings of HLT-NAACL 2007 workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technology. (2007)
11. Bohus, D., Rudnicky, A.I.: The ravenclaw dialog management framework: Architecture and systems. Computer Speech & Language **23**(3) (2009) 332–361