# Multi-sensor semantic mapping and exploration of indoor environments


Islem Jebari, Stéphane Bazeille, Adriana Tapus, David Filliat
*Cognitive Robotics Theme, ENSTA ParisTech, Paris, France*

Ryad Benosman, Sio-Hoï Ieng, Cedric Meyer
*Institut Des Systemes Intelligents et de Robotique (ISIR-CNRS), Paris, France*

Benoit Pothier, Jean-Charles Mamanna, Eddy Cizeron
*GOSTAI S.A.S., Paris, France*

*Corresponding author: islem.jebari@ensta-paristech.fr Phone: 0033 1 45 52 64 80*
*Fax: 0033 1 45 52 83 27*


The human perception of the external world appears as a natural, immediate and effortless task. It is achieved through a number of “low-level” sensory-motor processes that provide a high-level representation adapted to complex reasoning and decision. Compared to these representations, mobile robots usually provide only low-level obstacle maps that lack such high-level information. This abstract presents the Panoramic and Active Camera for Object Mapping (PACOM) project. Its goal is to develop a mobile robot to participate in the “CAROTTE” competition [DefiCarotte] organized by the French research funding agency (ANR) and the French armament procurement agency (DGA). The goal is to autonomously explore an unknown indoor environment and to build a semantic map containing high-level information similar to those extracted by humans and that will be rapidly and easily interpreted by users to assess the situation.

## *System overview*

We developed a robot based on a *pioneer 3 dx* from Mobile Robots Inc. The robot was fitted with 2 scanning laser range finders (one horizontal and one vertical), a ring of sonar sensors, a Pan-Tilt-Zoom camera and three on-board computers. Software architecture uses the Urbi framework; a middleware for programming complex robotic systems developed and commercialized by Gostai. Urbi is composed of a distributed component architecture (UObject), and an innovative orchestrator language (urbiScript) to coordinate all components. This language incorporates high-level features that facilitate the development of parallel and event-based applications.

## *Software components*

The main software components are: mapping, path planning, exploration, object recognition and semantic mapping:

- “Mapping” performs 2D SLAM (*Simultaneous Localization And Mapping*) using the horizontal laser scanner and the Karto software library [Karto], which provides good performance and robustness in indoor environments. This library uses scan matching to correct the robot odometry drift and provides a 2D occupancy grid map.

- “Path planning” performs global path planning given the current map of the environment and carries out local obstacle avoidance of dynamic objects taking into account lasers and sonars. This module is based on the Karto library.

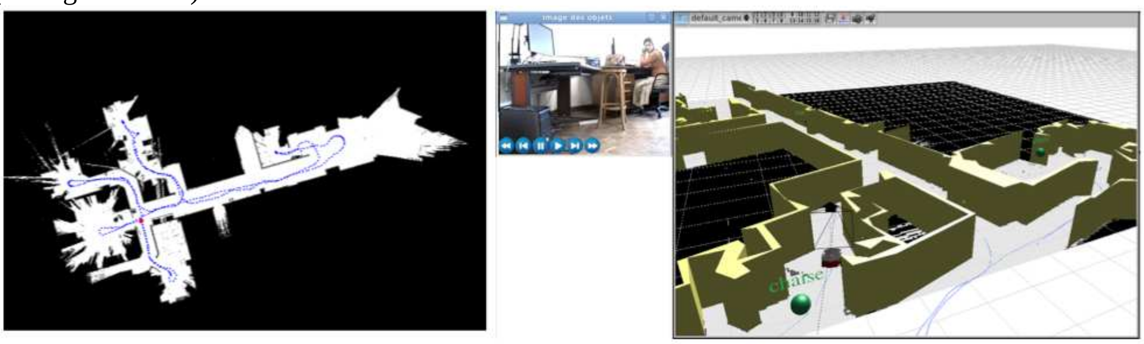
- "Exploration" integrates the need to map the environment structure and to detect objects. Global exploration focus on finding the next location in order to discover unknown areas through the laser scanner. It applies the Frontier Based Exploration method of Yamauchi [Yamauchi97]. The object search method is used during global exploration to select robot positions and camera directions in order to search objects in previously unseen areas. This method is based on a 2D view-map that records the area already perceived by the robot camera. A stochastic method inspired by the Randomized Art-Gallery Algorithm [Gonzalez-Banos01] is used to search the new position that discovers the more unseen area.

- "Object recognition" performs visual object detection using the bag of visual words approach [Sivic03]. This approach is based on local features whose descriptors are matched to a dictionary of visual words [Nister06]. Using the visual words occurrence histogram, a voting method is used to predict the object identity [Filliat07]. Two types of local features adapted to textured and coloured objects are used: SURF keypoints [Bay08] and local colour histograms [Filliat07]. This method is applied to each visually salient image area detected using [Butko08] in order to perform multiple object detection in an image. After detection, object distance is estimated by comparing the scale of the detected image to training images that have been recorded with known object distance. The objects to recognize (chairs, books, bottle, fan…) are either isolated or grouped together and they have variable sizes and colours.

- "Semantic mapping" adds meaningful information to the 2D map such as the 3D structure of the environment, the position, name and images of the detected objects and the rooms. The 3D structure is built as a point cloud using the second laser sensor, assuming that the "mapping" module gives a correct position. We estimate the object position by integrating multiple detections using Kalman filter [Smith86]. Rooms are detected in the occupancy grid through an algorithm that detects doors and analyses the resulting connected components of open space.

### *Mapping results*

This system participated in the first competition held in june 2010. It performed a mapping for artificial environments designed for the competition. An interface displaying the map, the robot trajectory and the object detection in 3D has been developed using the Peekabot library [Peekabot] (see figure below).




For the second year of the challenge, new constraints will be added to the competition rules like a composite ground (floor, carpet, grid, sand, grass…), more complex objects and multi-level environments.

### *Acknowledgement*

This project is funded by ANR under the subvention 2009 CORD 102.

## *References*


[Bay08] Bay, H.; Ess, A.; Tuytelaars, T. and Van Gool, L. (2008). "SURF: Speeded Up Robust Features". In Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346-359.

[Butko08] Butko, N.J.; Zhang, L.; Cottrel, G.W. and Movellan, J.R. (2008). "Visual Saliency Model for Robot Cameras". In International Conference on Robotics and Automation.

[DefiCarotte] http://www.defi-carotte.fr/

[Filliat07] Filliat, D. (2007). "A visual bag of words method for interactive qualitative localization and mapping". In Proceedings of the International Conference on Robotics and Automation (ICRA).

[Gonzalez-Banos01] Gonzalez-Banos, H and Latombe, J. (2001). "A randomized Art-Gallery Algorithm for Sensor Placement". In Annual Symposium on Computational Geometry.

[Karto] http://www.kartorobotics.com/

[Nister06] Nister, D. and Stewenius, H. (2006). "Scalable Recognition with a Vocabulary Tree". In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Volume 2.

[Peekabot] http://www.peekabot.org/

[Sivic03] Sivic, J. and Zisserman, A. (2003). "Video Google: A Text Retrieval Approach to Object Matching in Videos". In Proceedings of the International Conference on Computer Vision.

[Smith86] Smith, R.C. and Cheeseman, P. (1986). "On the Representation and Estimation of Spatial Uncertainty". In The International Journal of Robotics Research 5 (4): 56–68.

[Yamauchi97] Yamauchi, B. (1997). "A frontier-based approach for autonomous exploration". In proceedings of the 1997 IEEE International Symposium on Computational Intelligence in Robotics and Automation.