The Golem Team, RoboCup@Home 2016

Team Leader: Luis A. Pineda[1]
Caleb Rascon, Gibran Fuentes, Arturo Rodriguez, Hernando Ortega, Mauricio Reyes, Noé Hernández, and Ricardo Cruz
[1] lpineda@unam.mx
http://turing.iimas.unam.mx/~luis

Computer Science Department
Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS)
Universidad Nacional Autónoma de México (UNAM)
http://golem.iimas.unam.mx

Abstract. In this work we describe the Golem Team and the latest version of the robot Golem-III. This is the fourth time the Golem team participates on the RoboCup@Home Competition. The design of our robot is based on a conceptual framework that is centered on the notion of dialogue models with the interaction-oriented cognitive architecture (IOCA) and its associated programming environment, SitLog. This framework provides flexibility and abstraction for task description and implementation, as well as a high level modularity. The tasks of the RoboCup@Home competition are implemented under this framework using a library of basic behaviors.

1 Team Members

Robot: Golem-III.
Academics:

Dr. Luis A. Pineda. Dialogue Management, SitLog, Task Structure and Robotic Behavior, Knowledge Representation and Cognitive Architecture.
Dr. Caleb Rascon. Robot Audition, Navigation and Module Integration.
Dr. Gibran Fuentes. Vision and Object Manipulation.
M.Sc. Hernando Ortega. Robotic platform development and embedded software control.
M.Sc. Mauricio Reyes Castillo. Industrial Design.
M.Sc. Noé Hernández. Object Modeling and Technical Support.

Students:

M.Sc. Arturo Rodríguez Garcia. Person Recognition and Tracking, SitLog Behaviors, and Knowledge Base Programming.
M.Sc. Ricardo Cruz. Object Modeling and Health-care Applications.

2 Group Background

The Golem Group is a research group focused on robotics mainly on the cognitive modeling of the interaction between humans and robots. The group was created within the context of the project "Diálogos Inteligentes Multimodales en Español" (DIME, Intelligent Multimodal Dialogues in Spanish) in 1998 at IIMAS, UNAM where it has been established since. The goals of the DIME project were the analysis of multimodal task-oriented human dialogues, the development of a Spanish grammar, speech recognition in Spanish, and the integration of a software platform for the construction of interactive systems with spoken Spanish. By 2001 the group started the Golem project with the purpose of generalizing the theory for the construction of intelligent mobile agents, in particular the Golem robot. A first result was a version of a theory for the specification and interpretation of dialogue models which is still a corner stone in the group's philosophy [13].

Several versions of the Golem robot were demonstrated at the Universum Science Museum in which Golem interacted with visitors. In 2002, the robot had a simple conversation and followed movements commands.

In 2006, it guided a poster session. Finally, in 2009 we presented the module "Guess the card: Golem in Universum" in which children played a game [12].

In 2010, we started the development of Golem-II+ our current service robot. Golem-II+ incorporates an innovative explicit cognitive architecture that, in conjunction with the dialogue model theory and program interpreter, constitutes the theoretical core of our approach [14,17].

Since 2011, we have participated at the RoboCup@Home competition: Istanbul 2011, Mexico 2012 and Netherlands 2013. We have also participated on the local Mexican competitions in 2012 (1st place) and 2013, and German Open in 2012 (3rd place). All of which provided important feedback for the robot's performance. In particular, at the RoboCup@Home 2013 the team was awarded the Innovation Award of the league for our demo in which the robot uses its audio-localization system to perform a waiter role in a noisy enviroment. This demo incorporated the capability of spatial reasoning to the navigation subsystem, including the ability of facing the interlocutor during human-robot interactions [19] and provided the possibility for the robot to interact with more than one agent at a time [22].

During the span of 2014 and 2015, the team developed the iteration of the Golem here described, Golem-III, which expands our previous developments [15] and implements them in a new robotic base and body. This version uses a new set of modular behaviors programmed in SitLog[16], a new knowledge base system, a new system for detecting and tracking heads at the distance, and a new audio-activity tracker. In terms of hardware, we have added a new set of cameras to be used by the the computer vision module, and two new robotic arms with more degrees of freedom.

3 An Interaction-Oriented Cognitive Architecture

The behavior of our robot Golem-II+ is regulated by an Interaction Oriented Cognitive Architecture (IOCA) [14,17]. The IOCA architecture specifies the types of modules which integrate our system. A diagram of IOCA can be seen in Figure 1.

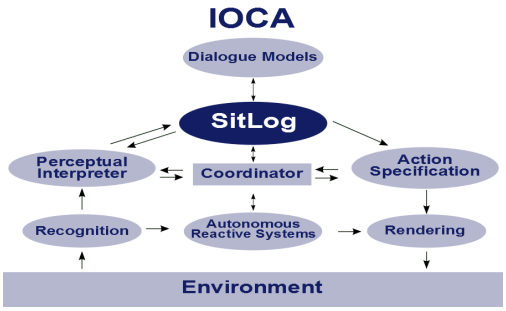

Fig. 1. Interaction Oriented Cognitive Architecture (IOCA).

Recognition modules encode external stimuli into specific modalities (e.g., speech into utterances transcriptions, images to SIFT features). *Interpreter* modules assign a meaning to those messages from different modalities (e.g., from utterances or SIFT features to a semantic representation, $name(golem)$ or $object(juice)$). On the other side, *Specification* modules specify global parameters into particular ones for the actions (e.g., *kitchen* the x,y points). *Render* modules are in charge to execute the actions (e..g, perform navigation actions to arrive to the kitchen). In the case of the dialogue manager module there is only one of its type. This is in charge to manage the execution of the task. A more detailed explanation is presented in section 4.1.

Reactive behavior is reached by tightly joining recognition and render modules into *Autonomous Reactive Systems* (ARSs). Example of these system are the Autonomous Navigation System (ANS) and the Autonomous Position and Orientation Source of Sound Detection System (APOS) to allow the robot to face its interlocutor reactively.

4 Software

We organize our software in modules for different skills and the IOCA architecture manages the connectivity among these modules. Additionally, for some of these skills we have abstracted some basic behaviors which are programmed in SitLog. These behaviors encapsulate skills and their recovery strategies. Example of such behaviors are to detect a person, to see an object or to visit a sequence of points. Next we present the main modules and its associated behaviors.

4.1 Dialogue Manager

The central communication and control structure of Golem is defined through modular schematic protocols that we called Dialogue Models (DM). DMs represent the structure of a given task. DMs are specified in SitLog (Situations and Logic) [16][1], a declarative programming language developed within the context of the project. DMs have a diagrammatic representation as Recursive Transition Networks, where nodes represent world situations and edges represent expectations and actions pairs, and situations can stand for fully embedded tasks. SitLog has also an embedded functional language for the declarative specification of control and content information. Expectations, actions and situations are specified through basic expressions of this language or through functions that are evaluated dynamically, supporting large abstraction in this dimension too. SitLog's interpreter is coded in Prolog and the specification of DMs follows closely the Prolog's notation. SitLog's interpreter is the central component of the IOCA architecture, and evaluates DMs continuously during the execution of the tasks, and also coordinates reactive and deliberative behavior.

4.2 Knowledge Representation

Golem has a central knowledge representation system consisting of a KB manager with its knowledge repository and administration procedures. Knowledge is specified as a class taxonomy with inheritance and supports naturally the expression of defaults and exceptions. The system permits the expression of properties of classes, relations between classes, and the expression of individuals of each class with their particular properties and relations. Conflicts between particular and general properties and relations are handled through the criteria of specificity, such that properties and relations of individuals have precedence over the properties and relation of their classes. All objects within the KB can be updated dynamically and the scheme behaves non-monotonically. The KB system is coded in Prolog and the KB-services can be used within the body of DMs directly. It has been fully design and developed within the context of the project.

4.3 Vision

Vision is carried out via various vision modules, described hereafter.

Face and Head Detection, Tracking and Recognition. OpenCV is used to perform face and head detection, tracking and recognition. Face detection is carried out by using the Viola-Jones Method [24]. Face recognition is based on the Eigenfaces technique [23]. For the head detection we use Histogram of Oriented Gradients (HOG) models created in house [3]. Face and head tracking is carried out via a technique based on the Hungarian Algorithm and Kalman Filtering [8].

For person detection and recognition, the face features are used when the person is close and head features when he/she is far. During the search, we take advantage of our 2-DOF neck movement capability to enhance the search.

Object Recognition. This capability is performed using the MOPED framework proposed by Collet et. al. [2]. During the training stage, several images are acquired from each object that is to be recognized, and SIFT features [9] are extracted from each image. Structure from Motion is applied to arrange all the features from all the images and obtain a 3D model of each object. In the recognition phase, the SIFT features are

[1] http://golem.iimas.unam.mx/sitlog

obtained from the visual scene and, with 3D model at hand, hypotheses are made in an iterative manner using Iterative Clustering Estimation [2]. Finally, Projective Clustering [2] is used to group the similar hypothesis and provide one for each of the objects being observed. A behavior associated to this capability is in charge of returning the position and orientation of a given object or the closest one in the scene.

Person Tracking, Gesture Estimation and Soft Biometric Identification. Kinect 2 SDK is used to detect persons and their respective skeletons. This information is used to estimate if persons are waving or pointing with their hands. The orientation of persons in relation with the robot is also estimated to determine if they are facing the robot. The skeleton is used to learn and identify persons based on their clothes. Different views of the same person indexed by their orientation angle are stored in the soft biometric database. The identification by clothes is intended to be used in situations where face recognition is not suitable. The current angle of the user to be identified is used to select the nearest view of each person stored in the soft biometric database, and then comparing small patches semantically tagged, extracted from different parts of the body (arms, chest, legs, etc.).

Plane detection. We use the Point Cloud Library, as well as the Kinect 2 SDK to detect planes. In particular, we focus on horizontal planes which could correspond to tables. We use this information to put safely an object on the tab table.

4.4 Arm and Neck Manipulation

The 5-DOF robotic arms were built in-house. These are mounted on the robot on its torso, the height of which can be controlled via two electronic pistons, providing a sixth DOF. The central upper part of the torso, a seventh DOF is provided for both arms, which acts as a clavicle that extends the length of the manipulation range. The robotic arms are based on Robotis Dynamixel motors for movement, and are controlled via a Servo Controller, which, in turn, accepts commands of the type of "park", "grasp object", and "offer object". The latter two are able to accept distance and angle arguments.

Associated to the arm there are the *take* and *deliver* behaviors. The *take* behavior will grasp an object taking into account the position and orientation information obtained by the object recognition module. Additionaly, it positions the robot so the object is in reach. The *deliver* behavior is in charge to deliver an object; it can do it to a specific position, or put it in a table using the plane detection module.

The 2-DOF robotic neck was also built in-house, and it is mounted over the upper base of the robot. This neck allows the range of the Kinect and the color camera to be shifted vertically and horizontally providing a wide area of recognition. In addition, a directional microphone is mounted over the horizontal DOF for the same purpose.

4.5 Speech Recognition and Synthesis

Based on the Windows Speech API, the ASR is able to switch between language models with which we hand-crafted a corpus for each of the tasks. The ASR switches from one language model to another, depending on the context of the dialogue (A yes/no language model for confirmation, a name language model for when the user is being asked their name, etc.). The ASR is kept idle until a recognition is requested by the Dialogue Manager, which triggers a 'bleep up' sound to signal the user that it is listening, and a 'bleep down' sound to signal that it is done listening. This module is a Recognizer in the IOCA framework [11].

In addition, the speech synthesis is also based on Windows Speech API, using the US Male voice. From the point of view of the IOCA framework, this is a Rendering module.

Both recognition and synthesis are an autonomous system so that the robot does not speak while listening or vice-versa.

4.6 Language Interpretation

In this version of the system, the language interpretation is based on a parser implemented in Prolog using Definite Clause Grammars. All production rules are objects stored in the knowledge base, which terminals are also extracted from the KB taxonomy.

4.7 Audio Localization

It is based on the GPL software JACK, which is used to create an all-encompassing simulated sound card that can be accessed by different audio clients at the same time, in real-time. It provides a robust direction-of-arrival estimation in near- real-time manner in mid-level reverberant environments, throughout the 360° azimuth range. The signals from the three microphones of the 8SoundsUSB external sound interface [1] are set in an equilateral triangle, which provide three measured delay-comparisons. This provides redundancy to the direction-of-arrival estimation, as well as a close-to-linear mapping between delay measurements and direction-of-arrival estimations [19].

This module, in conjunction with the Reactive Navigation Module (described later), compose the Autonomous Position and Orientation Source of Sound Detection System (APOS). In addition, a multi-DOA estimation is employed if there are more than one user in the environment [21,22,20].

4.8 Navigation

The Autonomous Navigation System (ANS) inside IOCA is based on the ROS Navigation Stack, which is divided in several parts: the Global Planner, the Local Planner, the Auto-Localization Module, and the Map Server. The Map Server carries out GMapping for map creation [7] and Adaptive Monte-Carlo Localization (AMCL) is used in the Auto-Localization module [4]. The Local Planner uses the Dynamic Window Approach for collision avoidance [6] and the Global Planner uses a combination of grid-based optimal route seeker (similar to the Dijkstra Algorithm[5]), interpolation between the points for path smoothening, and the use of a costmap (fed by the Map Server) to avoid planning paths through walls [10].

We also implemented a Semantic Proxy that carries out topological translation between a label of a custom location and its coordinates and robotic pose. To do this, a set of labels are inputed and stored using a custom GUI that uses the Map Server to provide a visual status of the environment. These labels are then used by a custom ROS node that carries out the translation. This node also provides simple moving capabilities such as turn θ degrees, move Z meters to the front, etc. to the Dialogue Model.

The ANS has associated one behavior which coordinates the movement of the robot in its different versions: relative or absolute, topological places or coordinates, normal or fine movement, using a pre-made map or carry out automatic mapping.

4.9 Software Libraries

Both the robot internal computer and the external laptop run the Ubuntu 14.04 operating system, and inter-modular communication is done using ROS [18]. Table 1 shows which software libraries are used by the IOCA modules and Golem-II+'s hardware.

Table 1. Software Libraries used by the IOCA Modules and Hardware of Golem-II+

Module	**Hardware**	**Software Libraries**
Dialogue manager	–	SitLog, SWI Prolog
Knowledge-base	–	SWI Prolog
Vision	Kinect 2 and Flea3 Camera	SVS, OpenCV, PCL, and Kinect 2 SDK
Speech Recognition	Rode VideoMic Directional Microphone	Windows Speech API
Robot Audition	8SoundsUSB External Sound Card	JACK
Voice synthetizer	Speakers	Windows Speech API
Navigation	Bumpers, Laser, Odometric Sensors	ROS Navigation Stack
Object Manipulation	Custom Robotic Arms and Torso	Dynamixel RoboPlus
Camera/Mic. Movement	Custom Robotic Neck	Dynamixel RoboPlus

5 Description of the Hardware

The "Golem-III" robot (See Fig. 2) will be used, which is composed by the following hardware:

- PatrolBot^{TM} robot (Mobile Robots Inc.)
  - 8-sensor sonar array
  - Two protective 5-bumper arrays
  - Infinity 3.5-Inch Two-Way loudspeaker
  - On-board computer Cobra EBX-12
  - Sick LMS-500 Laser
- Dell Precision M4600 laptop computer x 2
- Black Box 5-port ethernet switch
- Microsoft Kinect 2 camera
- Point Grey Flea USB 3 camera
- 8SoundsUSB audio interface (3 microphones)
- RODE VideoMic directional microphone
- In-house robotic arms, gripper and neck

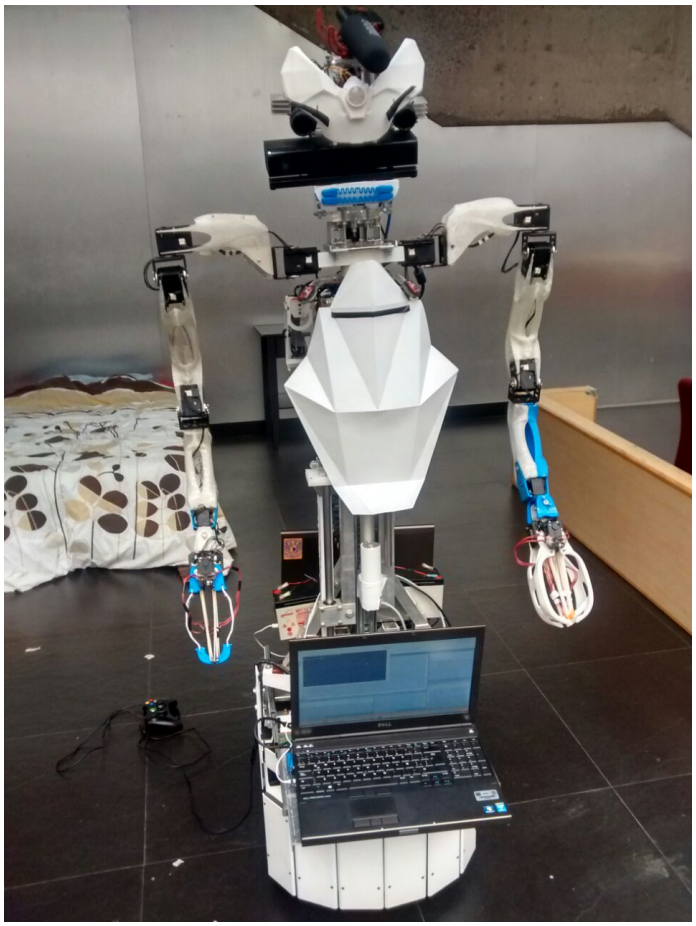

Fig. 2. The Golem-III robot.

Acknowledgments

Golem-III was financed by CONACyT project 178673, by PAPIIT-UNAM project IN107513 and SECITI project ICyTDF/209/2012. We would like to thank the students and academic personnel that have provided us support outside the competition:

- Raul Peralta. Navigation and manipulation development.
- Uriel Ortiz. CAD modeling and manipulation development.

References

1. Abran-Cote, D., Bandou, M., Beland, A., Cayer, G., Choquette, S., Gosselin, F., Robitaille, F., Kizito, D.T., Grondin, F., Letourneau, D.: USB Synchronous Multichannel Audio Acquisition System.pdf, http://sourceforge.net/projects/eightsoundsusb/files/Technical%20Paper/USB%20Synchronous%20Multichannel%20Audio%20Acquisition%20System.pdf
2. Collet, A., Martinez, M., Srinivasa, S.S.: The MOPED framework: Object Recognition and Pose Estimation for Manipulation. The International Journal of Robotics Research 30, 1284–1306 (2011)
3. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. vol. 1, pp. 886–893 vol. 1 (2005)
4. Dellaert, F., Fox, D., Burgard, W., Thrun, S.: Monte carlo localization for mobile robots. In: Robotics and Automation, 1999. Proceedings. 1999 IEEE International Conference on. vol. 2, pp. 1322–1328 vol.2 (1999)
5. Dijkstra, E.W.: A note on two problems in connexion with graphs. Numerische Mathematik 1, 269–271 (1959), http://dx.doi.org/10.1007/BF01386390, 10.1007/BF01386390
6. Fox, D., Burgard, W., Thrun, S.: The dynamic window approach to collision avoidance. Robotics Automation Magazine, IEEE 4(1), 23–33 (Mar 1997)
7. Grisetti, G., Stachniss, C., Burgard, W.: Improved techniques for grid mapping with rao-blackwellized particle filters. Robotics, IEEE Transactions on 23(1), 34–46 (Feb 2007)
8. Kalman, R.E.: A New Approach to Linear Filtering and Prediction Problems. Transactions of the ASME–Journal of Basic Engineering 82(Series D), 35–45 (1960)
9. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60, 91–110 (2004)
10. Lu, D.V.: Contextualized Robot Navigation. Ph.D. thesis, WASHINGTON UNIVERSITY IN ST. LOUIS, Saint Louis, Missouri (2014)
11. Meza, I., Rascon, C., Pineda, L.: Practical Speech Recognition for Contextualized Service Robots, Lecture Notes in Computer Science, vol. 8266, pp. 423–434. Springer Berling Heidelberg (2013)
12. Meza, I.V., Salinas, L., Venegas, E., Castellanos-Vargas, H., Alvarado-González, M., Chavarría-Amezcua, A., Pineda, L.A.: Specification and Evaluation of a Spanish Conversational System Using Dialogue Models. Advances in Artificial Intelligence - IBERAMIA 2010 6433 (2010)
13. Pineda, L.A.: Specification and Interpretation of Multimodal Dialogue Models for Human-Robot Interaction. In: Sidorov, G. (ed.) Artificial Intelligence for Humans: Service Robots and Social Modeling, pp. 33–50. SMIA, Mexico (2008)
14. Pineda, L.A., Castellanos, H., Cuétara, J., Galescu, L., Juárez, J., Llisterri, J., Pérez, P., Villaseñor, L.: The Corpus DIMEx100: Transcription and Evaluation. Language Resources and Evaluation 44, 347–370 (2010)
15. Pineda, L., Group, G.: The Golem Team, RoboCup@Home 2013. In: Proceedings of RoboCup 2013 (2013)
16. Pineda, L., Salinas, L., Meza, I., Rascon, C., Fuentes, G.: SitLog: A Programming Language for Service Robot Tasks. International Journal of Advanced Robotic Systems 10(538) (2013)
17. Pineda, L.A., amd Héctor H. Avilés, I.V.M., Gershenson, C., Rascón, C., Alvarado, M., Salinas, L.: IOCA: An Interaction-Oriented Cognitive Architecture. Research in Computer Science 54, 273–284 (2011)
18. Quigley, M., Conley, K., Gerkey, B.P., Faust, J., Foote, T., Leibs, J., Wheeler, R., Ng, A.Y.: ROS: an open-source Robot Operating System. In: ICRA Workshop on Open Source Software (2009)
19. Rascón, C., Avilés, H., Pineda, L.A.: Robotic Orientation towards Speaker for Human-Robot Interaction. Advances in Artificial Intelligence - IBERAMIA 2010 6433, 10–19 (2010)
20. Rascon, C., Fuentes, G., Meza, I.: Lightweight multi-DOA tracking of mobile speech sources. EURASIP Journal on Audio, Speech, and Music Processing 2015(11) (2015)
21. Rascon, C., Pineda, L.: Lightweight Multi-DOA Estimation on a Mobile Robotic Platform. In: Lecture Notes in Engineering and Computer Science: Proceedings of The World Congress on Engineering and Computer Science 2012, WCECS 2012, 24-26 October, 2012, San Francisco, USA. pp. 665–670 (2012)
22. Rascon, C., Pineda, L.: Multiple Direction-of-Arrival Estimation for a Mobile Robotic Platform with Small Hardware Setup, Lecture Notes in Electrical Engineering, vol. 247, pp. 209–223. Springer Netherlands (2014)
23. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience 3(1), 71–86 (1991)
24. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. vol. 1, pp. I–511–I–518 (2001)