The Golem Team, RoboCup@Home 2020

Luis A. Pineda (Team Leader), Gibran Fuentes, Arturo Rodríguez, Hernando Ortega, Mauricio Reyes, and Noé Hernández

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

Abstract. In this paper we describe the Golem Team and the latest version of its service robot Golem-III. This is the fifth participation of the Golem team in the RoboCup@Home Competition. The operation 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 high modularity. The tasks of the RoboCup@Home competition are implemented under this framework using a library of basic behaviors. In addition, over this framework, a system that carries out diagnostics, decision making and planning has been developed as an inference machine platform with which error detection and recovery is possible.

1 Team Members

Robot: Golem-III.
Academics:

- **Dr. Luis A. Pineda.** SitLog, Knowledge Representation, Inference and Cognitive Architecture.
- **Dr. Gibran Fuentes.** Vision and Object Manipulation.
- **Dr. Arturo Rodríguez Garcia.** Person Recognition and Tracking, SitLog Behaviors, and Knowledge Base and Inference Programming.
- **M.Sc. Hernando Ortega.** Robotic Platform Development and Embedded Software Control.
- **Dr. Mauricio Reyes Castillo.** Industrial Design and Emotion Expression.
- **M.Sc. Noé Hernández.** Object Modeling and SitLog Behaviors.
- **Dr. Ricardo Cruz.** Object Modeling and Health-care Applications.

Students:

- **M.Sc. Dennis Mendoza.** Vision, Object Manipulation and Robot Navigation.
- **Rocío Aldeco-Pérez.** Project Management and SitLog behaviors
- **Mario Rosas.** Human Machine Interaction

Emmanuel Maqueda. Human Machine Interaction
Juan Álvarez. Industrial Design
Karla González-Carreón. Project Management and Object Modeling

2 Group Background

The Golem Group is a research group focused on service 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. 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 [6].

Since 2011, we have participated in the following RoboCup@Home competitions: Istanbul 2011, Mexico 2012, Netherlands 2013, Germany 2016 and Japan 2017. We have also participated in the local Mexican competitions in 2012 (1st place), 2013, 2014, 2016 (2nd place) and 2017, and German Open in 2012 (3rd place), 2018 (3rd place) and 2019. These competitions have 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 environment.

During the span of 2014 and 2015, the team developed Golem-III, the iteration of the robot presented at RoboCup@Home in Germany 2016. This version uses a set of modular behaviors programmed in SitLog [8], a knowledge base system [10], a system for detecting, identifying and tracking persons, and an audio-activity tracker. In terms of hardware, this implementation uses a robotic torso, which includes a 2-DOF robotic neck and two 5-DOF robotic arms.

3 An Interaction-Oriented Cognitive Architecture

Golem-III is a realization of the conceptual model presented in [11] and its behavior is regulated by an Interaction Oriented Cognitive Architecture (IOCA) [7,9]. The IOCA architecture specifies the types of modules which integrate our system. A diagram of IOCA can be seen in Figure 1.

Recognition modules encode external stimuli into specific modalities (e.g., speech into utterances transcriptions, images to convolutional neural network (CNN) features). *Interpreter* modules assign a meaning to those messages from different modalities (e.g., from utterances or CNN features to a semantic representation). On the other side, *Specification* modules specify global parameters

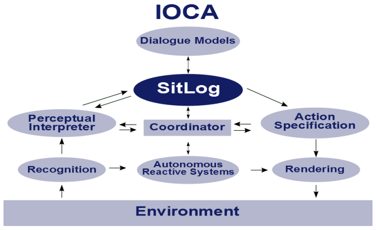

Fig. 1. Interaction Oriented Cognitive Architecture (IOCA).

into particular ones for the actions (e.g., *kitchen* the x,y points). *Render* modules are in charge of executing the actions (e..g, to 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 of managing the execution of the task. A more detailed explanation is presented in section 4.1.

A reactive behavior is reached by tightly joining recognition and render modules into *Autonomous Reactive Systems* (ARSs). Two examples of these systems 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. In this section, we present these modules and their associated behaviors.

4.1 Dialogue Manager

The central communication and control structure of Golem-III is defined through modular schematic protocols that we call Dialogue Models (DM). DMs represent the structure of a given task and are specified using SitLog (Situations and Logic) [8][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 abstractions 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

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

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-III has a central knowledge representation system [10], which consists of a KB manager with its knowledge repository and administration procedures. Golem-III's KB allowas the expression and reasoning about preferences. In this system, knowledge is specified as a class taxonomy with inheritance and it 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 relations of their classes. All objects within the KB can be updated dynamically and the scheme behaves non-monotonically. The KB also allows the specification of user preferences and offers inference mechanisms to exploit them [18]. 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 Opportunistic Symbolic Reasoning

Golem-III is able to invoke a reasoning engine on demand to perform deliberative inferences either in particular situations within the task structure or when its expectations do not match the current situation of the world [12]. The inference cycle in the latter case involves performing diagnosis, decision making and planning. The first generates a set of plausible explanations of the observed situation in relation to a set of causal rules; the second finds an action to be made to reach a desired situation in relation to the diagnosis and a set of preferences that guide the action of the robot at a higher level; finally, the third induces a plan to achieve such situation. This inferential cycle is used by the robot to perform tasks in dynamic environments, and also to recover from errors. The machine is independent of particular inference methods, like the explicit definition of a problem space that is explored through standard heuristic search or Answer Set Programming, both of which have been implemented successfully in Golem-III. This functionality is illustrated in the qualification video at `https://youtu.be/diQWdCzberE`.

4.4 Vision

Object Recognition. Object recognition and localization is based on the *You only look once* (YOLO) version 3 [16] real-time object detector. For specific objects, a customized detector is built by fine-tuning the YOLOv3 detector pretrained on the MSCOCO dataset [5]. The world coordinates of the objects is calculated by extracting the depth of their corresponding bounding boxes from the Kinect v2.

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 and to classify if they are standing, sitting or laying on the floor. The orientation of persons in relation to 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.). Microsoft Cognitive Services is used to obtain description data from the person, such as gender, age, glasses and facial hair. For person tracking, a global nearest neighbor based association approach from the world coordinates of OpenPose's detections [2] is used.

4.5 Arm and Neck Manipulation

The 5-DOF robotic arms were built in-house and are controlled via a Servo Controller. These are mounted on the robot's 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 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.

4.6 Speech Recognition and Synthesis

Based on the Windows Speech API, the ASR is able to switch between language models 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. In addition, the speech synthesis is also based on Windows Speech API, using the US Male voice. Both recognition and synthesis are an autonomous system so that the robot does not speak while listening or vice-versa.

4.7 Language Interpretation

In this version of the system, the language interpretation is based on a parser implemented in Prolog using Definite Clause Grammars, mounted over a tree-based structure for re-usability. All rules and terminals are objects stored in the knowledge base.

4.8 Audio Localization

It provides a robust direction-of-arrival estimation in mid-reverberant environments, throughout the 360° azimuth range, from a 3-microphone array [1] via a redundant direction-of-arrival estimation [14]. In addition, a multi-DOA estimation is employed if there are more than one user in the environment [15].

4.9 Navigation

It is based on the ROS Navigation Stack that uses GMapping for map creation [4] and Adaptive Monte-Carlo Localization (AMCL) [3]. We also implemented a Semantic Proxy that carries out topological translation between a label of a custom location and its coordinates and robotic pose. The Navigation system can provide several versions of movement: relative or absolute, topological places or coordinates, normal or fine movement, using a pre-made map or carry out automatic mapping.

4.10 Emotion Expression

Golem-III has a robotic face that can express basic emotions according to the interpretations made along the execution of the task. The expression of emotions in the robot is specified in SitLog as an intentional action in the appropriate situations [17].

4.11 Software Libraries

Both the robot internal computer and the external laptop run the Ubuntu 16.04 operating system, and inter-modular communication is done using ROS Kinetic [13]. Table 1 shows which software libraries are used by the IOCA modules and Golem-III's hardware.

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

Module	**Hardware**	**Software Libraries**
Dialogue manager	–	SitLog, SWI Prolog
Knowledge-base	–	SWI Prolog
Vision	Kinect 2 and Flea3 Camera	YOLO, PCL Kinect 2 SDK, MS Cognitive Services
Robot Audition	Kinect v2 microphone array, 8SoundsUSB Sound Card	Windows Speech API, JACK
Voice synthetizer	Speakers	Windows Speech API
Navigation	Lasers, Odometric Sensors	ROS Navigation Stack
Object Manipulation	Custom Robotic Torso	Dynamixel RoboPlus
Camera/Mic. Movement	Custom Robotic Neck	Dynamixel RoboPlus

5 Description of the Hardware

The "Golem-III" robot (see Fig. 2) is composed by the following hardware:

- PatrolBot^{TM} robot
  - 8-sensor sonar array
  - Two 5-bumper protective arrays
  - Infinity 3.5-Inch two-way loudspeakers
  - On-board computer Cobra EBX-12
  - Sick LMS-500 Laser
- Two Dell Precision M7510 laptop computers
- Hokuyo SOKUIKI laser range finder
- Black Box 5-port usb-powered ethernet switch
- Microsoft Kinect v2 camera
- Point Grey Flea USB 3 high-resolution camera
- 8SoundsUSB audio interface
- 3 miniature microphones
- In-house robotic torso, arms, hands and neck

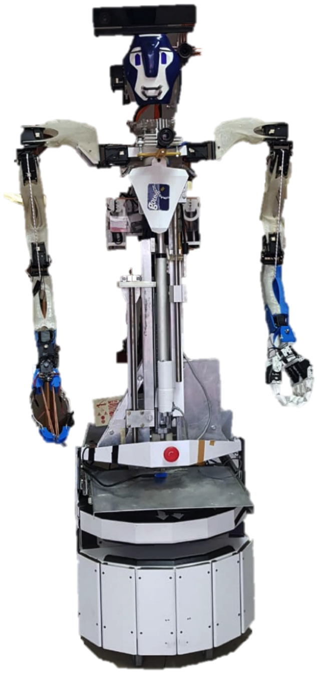

Fig. 2. The Golem-III robot.

Acknowledgments

Golem-III was financed by CONACyT project 178673, by PAPIIT-UNAM projects IN107513 and IN112819.

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
2. Cao, Z., Hidalgo, G., Simon, T., Wei, S.E., Sheikh, Y.: Openpose: Realtime multi-person 2d pose estimation using part affinity fields (2018)
3. 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)
4. 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)

5. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) Proceedings of the European Conference on Computer Vision. pp. 740–755 (2014)
6. 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)
7. 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)
8. 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)
9. Pineda, L.A., Meza, I.V., Avilés, H., Gershenson, C., Rascon, C., Alvarado-González, M., Salinas, L.: IOCA: Interaction-Oriented Cognitive Architecture. Research in Computer Science 54, 273–284 (2011)
10. Pineda, L.A., Rodríguez, A., Fuentes, G., Rascón, C., Meza, I.: A light non-monotonic knowledge-base for service robots. Intelligent Service Robotics pp. 1–13 (2017)
11. Pineda, L.A., Rodríguez, A., Fuentes, G., Rascon, C., Meza, I.V.: Concept and Functional Structure of a Service Robot. International Journal of Advanced Robotic Systems 12(2), 6 (2015)
12. Pineda, L.A., Rodríguez, A., Fuentes, G., Hernández, N., Reyes, M., Rascón, C., Cruz, R., Vélez, I., Ortega, H.: Opportunistic inference and emotion in service robots. Journal of Intelligent Fuzzy Systems 34(5), 3301–3311 (2018)
13. 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)
14. 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)
15. 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)
16. Redmon, J., Farhadi, A.: Yolov3: An incremental improvement (2018)
17. Reyes, M.E., Meza, I.V., Pineda, L.A.: Robotics facial expression of anger in collaborative human–robot interaction. International Journal of Advanced Robotic Systems 16(1), 1–13 (2018)
18. Torres, I., Hernández, N., Rodríguez, A., Fuentes, G., Pineda, L.A.: Reasoning with preferences in service robots. Journal of Intelligent Fuzzy Systems 36(5), 5105–5114 (2019)