Experiences with CiceRobot, a museum guide cognitive robot

I. Macaluso[1], E. Ardizzone[1], A. Chella[1], M. Cossentino[2], A. Gentile[1], R. Gradino[1], I. Infantino[2], M. Liotta[1], R. Rizzo[2], G. Scardino[1]

[1] Dipartimento Ingegneria Informatica – Università degli Studi di Palermo – Viale delle Scienze Ed. 6, 90128 Palermo
[2] Istituto di Calcolo e Reti ad Alte Prestazioni - Consiglio Nazionale delle Ricerche – Viale delle Scienze Ed. 11, 90128 Palermo

Abstract. The paper describes CiceRobot, a robot based on a cognitive architecture for robot vision and action. The aim of the architecture is to integrate visual perception and actions with knowledge representation, in order to let the robot to generate a deep inner understanding of its environment. The principled integration of perception, action and of symbolic knowledge is based on the introduction of an intermediate representation based on Gärdenfors conceptual spaces. The architecture has been tested on a RWI B21 autonomous robot on tasks related with guided tours in the Archaeological Museum of Agrigento. Experimental results are presented.

Introduction

The current generation of autonomous robots has showed impressive performances in mechanics and control of movements, see for instance the ASIMO robot by Honda or the QRIO by Sony. However, these state-of the-art robots are rigidly programmed and they present only limited capabilities to perceive, reason and act in a new and unstructured environment.

We claim that a new generation of cognitive autonomous robots, effectively able to perceive and act in unstructured environments and to interact with people, should be aware of their external and inner perceptions, should be able to pay attention to the relevant entities in their environment, to image, predict and to effectively plan their actions.

In the course of the years, we developed a cognitive architecture for robots [3][4]. The architecture is currently experimented on an autonomous robot platform based on a RWI B21 robot equipped with a pan-tilt stereo head, laser rangefinder and sonars (Fig. 1). The aim of the architecture is to integrate visual perception with knowledge representation to generate cognitive behaviors in the robot. One of the main features of our proposal is the principled integration of perception, action and of symbolic knowledge representation by means of an intermediate level of representation based on conceptual spaces [7].

Fig. 1. CiceRobot at work.

We maintain that the proposed architecture is suitable to support robot perception, attention, imagination and planning; in other words, the claim of this paper is that our architecture is a good candidate to achieve an effective overall cognitive behavior.
In order to test our robot architecture in non trivial tasks, we employed it in the CiceRobot project, a project aimed at developing a robotic tour guide in the Archaeological Museum of Agrigento. The task is considered a significant case study (see [2]) because it involves perception, self perception, planning and human-robot interaction.
The paper is organized as follows. Sect. 2 presents some remarks of the adopted cognitive framework; Sect. 3 deals with the implemented cognitive architecture; Sect. 4 is devoted to describe the robot vision system and Sect. 5 describes the implemented human-robot interaction modalities. Finally, Sect. 6 is a detailed description of an example of the operations of the robot at work.

Remarks of the adopted theoretical framework

The cognitive architecture of the robot is based on previous works ([3], [4]) and it is organized in three computational areas (Fig. 2). The Subconceptual Area is concerned with the processing of data coming from the robot sensors and in the considered case it is also a repository of behavior modules, as the localization and the obstacle avoidance module. This allows for standard reactive behaviors in order to face unpredictable situations in real time.
In the Linguistic Area, representation and processing are based on a logic-oriented formalism. The Conceptual Area is intermediate between the Subconceptual and the Linguistic Areas. This area is based on the notion of conceptual spaces CS [7], a metric space whose dimensions are related to the quantities processed in the subconceptual area.

In the implemented robot system, in the case of static scenes, a knoxel corresponds to geometric 2D, 2D and ½ and 3D primitives according to the perceived data. It should be noted that the robot itself is a knoxel in its conceptual space. Therefore, the perceived objects, as the robot itself, other robots, the surrounding obstacles, all correspond to suitable sets of knoxels in the robot's CS. In order to account for the representation of dynamic scenes, the robot CS is generalized to represent moving and interacting entities [4]. The dynamic Conceptual Space lets the agent to imagine possible future interactions with the objects in the environment: the interaction between the agent and the environment is represented as a sequence of sets of knoxels that is imagined and simulated in the conceptual space before the interaction really happens in the real world. The robot can imagine itself going through the environment and refining the plan if necessary. Once a correct plan is generated, the ideal trajectory can be sent to the robot actuators.

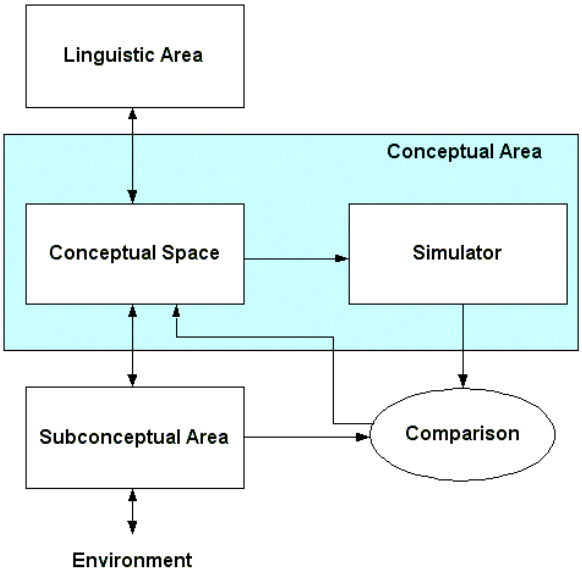

Fig. 2. The three computational areas.

The cognitive architecture of the robot

The described architecture has been implemented in the working robot and it is shown in Fig. 3. It should be noted that the three areas are concurrent computational components working together on different commitments.

The linguistic area acts as a central decision module: it allows for high level planning and contains a structured description of the agent environment. We adopted the Cyc ontology [6] extended with specific domain assertions and rules that allow common sense reasoning.

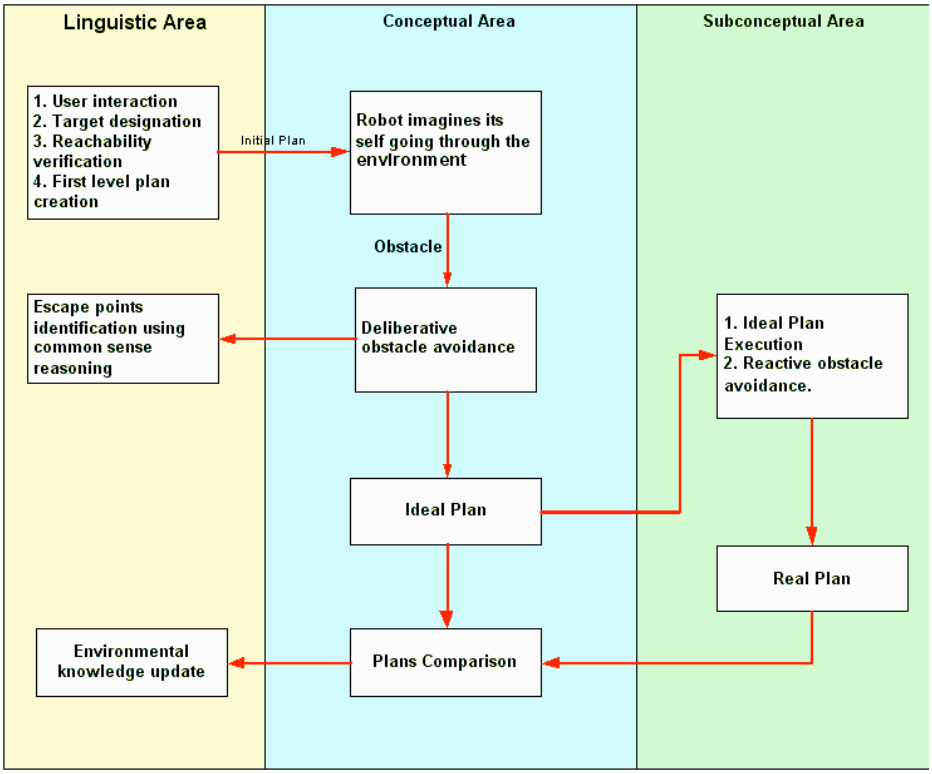

Fig. 3. The cognitive architecture.

At this level, the Information Retrieval Module (IRM) allows to understand visitors queries and to find related information in an interactive way, as described in subsequent Sect.

The linguistic planner receives the IRM results and converts them into knowledge base queries in order to obtain an initial plan which leads the robot to the selected targets, starting from the current configuration. This plan is the optimal solution to navigate the connectivity graph that describes the environment.

The planner, working on the conceptual area, verifies the applicability of the initial plan and, if necessary, modifies it. The planner operates on a 3D environment simulator (Fig. 4), based on VRML description of the robot environment in order to simulate and test the generated plans.

As previously stated, the dynamic conceptual space lets the agent to imagine possible future interactions with the objects. Therefore, by simulation it is possible to continually check the plan execution and, if it is not applicable (for instance because of the presence of some obstacles) a local re-planning task is started, asking the knowledge base a new optimal path to avoid the obstacle. An interesting feature of this approach is that it does not present local minima problems.

At this level, the robot imagines to go through the environment; when the robot finds an obstacle, then it refines the plan. The result is a plan that could be safely executed

by the real robot in a static environment. Such a plane can be considered a sequence of expected situations.

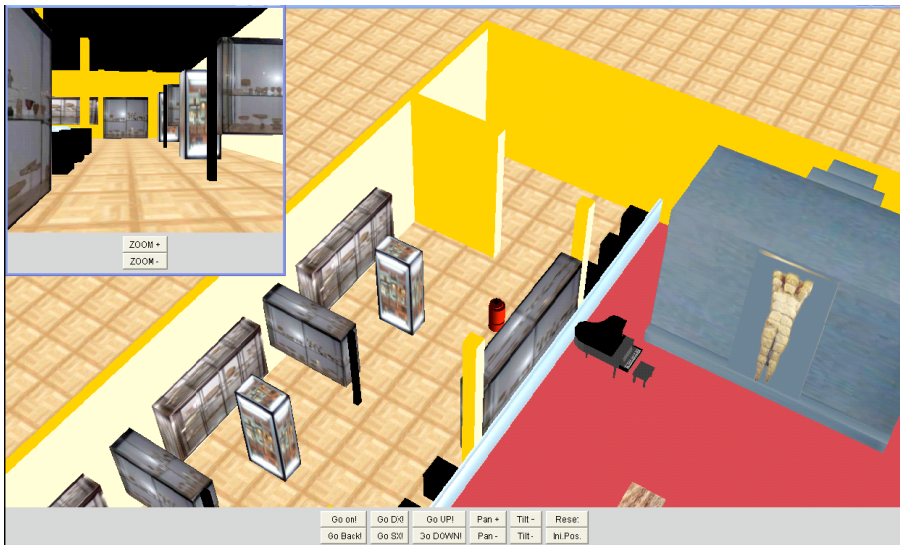

Fig. 4. Snapshot of the 3D simulator from the external and from the robot point of view.

Of course the robot will not be able to exactly follow the geometrical path generated by the planner both because the environment is populated by unknown moving and interacting obstacle and also because of the sensory motor errors. In order to deal with such unexpected situations, the reactive modules are activated to control the robot at execution time.
Therefore, the effective trajectory followed by the robot could be different from the planned one. In our model, we propose a direct comparison between the expected results of the action as they were simulated during plan generation, and the effective results, according to the current perception.
The outcome of the comparison process is used to update the knowledge of the environment and to decide whether or not it is necessary to re-plan . This is quite useful when the robot, during the real trajectory, get trapped in a local minimum, for instance because of the collision avoidance module. In that case, the plan is interrupted, the current model of the world is revised and the system starts up the re-planning module.

The Vision System

The Vision System is responsible of the knowledge processing arising from vision sensors [9]. A calibrated stereo head and its pan-tilt allows the acquisition of the images of the museum environment during the execution of the robot task. The main task of the vision system is to perform the self-localization allowing to detect and

correct the position of the robot in the scene. Preliminary stereo calibration is performed using the Calibration routine that processes the images of a grid.

Camera calibration. The pin-hole camera model requires the estimation of intrinsic parameters in order to allow three-dimensional reconstruction from stereo pairs. Given a world point X, the relation with its projection w on the image is

$$\lambda w = K\begin{bmatrix} R & t \end{bmatrix} X \tag{1}$$

where K is the calibration matrix

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

(c_x,,c_y) are the coordinates of the central point, and (f_x,f_y) are focal lengths along image dimensions u and v. Moreover, the radial distortion introduced by lens is corrected estimating the coefficients k_1,k_2,p_1, and p_2 that are involves in the following model

$$\begin{aligned} u_D &= u + u[k_1 r^2 + k_2 r^4] + [2p_1 uv + p_2(r^2 + 2u^2)] \\ v_D &= v + v[k_1 r^2 + k_2 r^4] + [2p_2 uv + p_2(r^2 + 2v^2)] \end{aligned} \tag{3}$$

where (u,v) are coordinates of the undistort image, (u_D,,v_D) are coordinates of the distort image, and $r^2 = u^2+v^2$. The calibration matrix is estimated by the standard algorithm described by Zhang [13] using a grid, and radial distortion coefficients by non-linear minimization algorithm [11] (see Fig. 5). Some filtering operations are performed on the acquired images in order to limit the influence of changing illumination during the robot movement.

Robot's self localization. Two critical issues are to be considered when the robot moves in the real world: the starting point and the initial orientation of the robotic platform are not exactly known, wheels friction and floor defects could be introduce great imprecision in the estimate of the current position of the robot taking in account the odometer. As consequence of that, the position of the robot needs to be updated during robot's movement using visual localization of known objects. The implemented vision system uses as landmark the same planar grid employed in the calibration phase: they are placed on the various walls of the museum and are visible to the robot also in presence of visitors.
When the self-localization is requested, the following operations are executed:

1. The coordinates of the nearest marker are obtained;
2. Camera pan-tilt movements are performed to view the marker by stereo head;
3. Additional corrections to pan-tilt position are computed to place the marker at the image centers;
4. 3D reconstruction of the marker points are performed by triangulation [hartley04];
5. The new estimated position of the robot respect to the landmark is computed;
6. The robot position is updated.

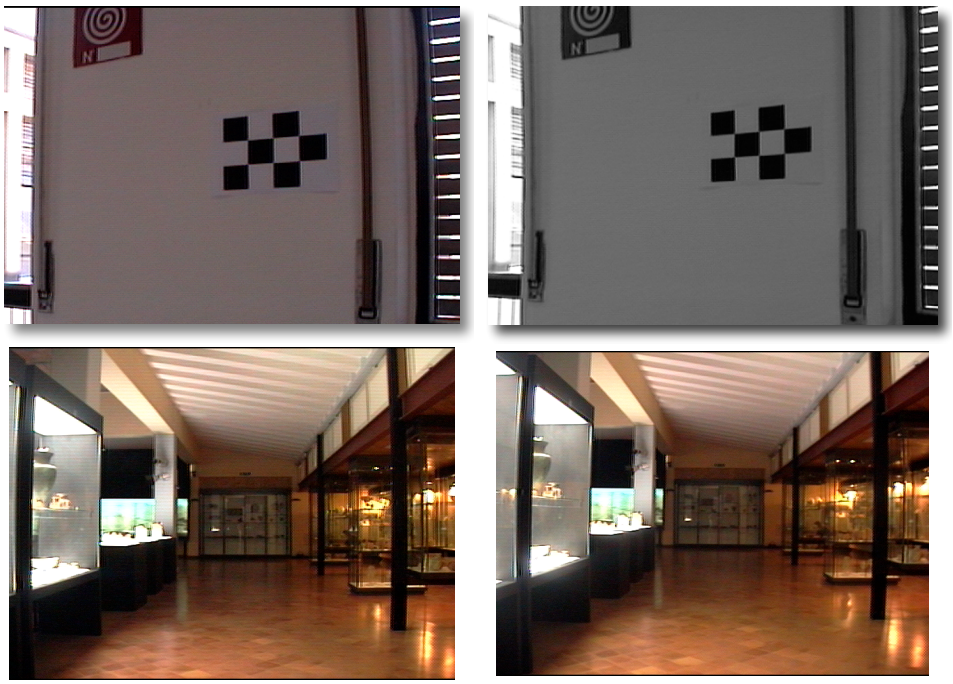

Fig. 5. In the first row, it is showed the calibration grid and the result of radial distortion removal. In the second row, a typical museum scene viewed by robot cameras. The light conditions and the structure of the scene are difficult to manage at the vision level.

Scene matching. The images acquired by the camera that represent the current perception of the robot are now matched with the corresponding expected scene generated by the planner and rendered by the 3D simulator previously described. Currently, this operation is simply performed by matching the corresponding landmarks in the acquired image and in the simulated one.
As previously stated, this comparison process has the role to synchronize the effective, external robot perception of the world with the inner robot expectations about the environment.

Human-Robot Interactions

Interaction between robot and visitors is a praiseworthy aspect of Cicerobot. The robot, not only explains to visitors the contents of windows but it also enables them to do queries to deepen topics related to objects that the windows themselves contain.
To this purpose, the robot is provided with an Information Retrieval Module (IRM) that finds information for an interactive presentation.
The main task of the IRM [9] is to provide the user with relevant information as a result of a request; at the same time, the quantity of less interesting information has to be minimized. Traditional search engines retrieve information through lexical criterions; so doing, all and only documents, containing the set of specified words in queries, or some logical combinations, are recovered. Then, information returned with this approaches is strongly dependent by the user request formulation.

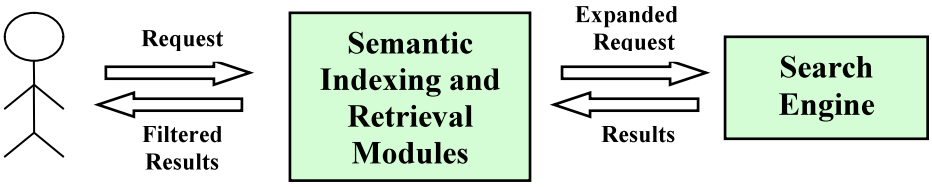

Fig. 6. The Information Retrieval Module.

Well known problems of lexical information retrieval systems are the polysemy, i.e., a term that can have many meanings, and synonymy, a concept that can be expressed by different terms. Interposing semantic modules between user and traditional search engine may help to retrieve higher level of interesting information. In the current implementation, the interposed modules enable semantic indexing and search. Both of them are based on the Latent Semantic Analysis (LSA), a theory related to knowledge representation and induction [9] (see Fig 2). User Request is converted in an Expanded Request, so that the search engine is questioned not only on words contained in the user request but also on words semantically closed to those which build the initial query. The interposed semantic module is then used to semantically filter the Results.
The results of semantic analysis is a list of documents and a windows sequence semantically connected to the user query.
Another feature of the IR system is that afterwards the interaction with the user it can increase its own knowledge base, inserting in the local repository also new documents, obtained in research on Internet or in the offline research and coding them in the semantic space: in this way the system knowledge base is improved during time basing on the experience.
The robot speech generation is enabled by a Vocal User Interface (VUI) based on the VoiceXML standard [9]. The environment is based on the IBM™ WebSphere Voice Response Server, which runs on a laptop sitting on the robot and equipped with speakers.

CiceRobot at Work

The robot operates in the "Sala Giove" of the Archaeological Museum of Agrigento. The user of the robot inserts a query, and the IRM returns the corresponding list of documents shown by a web interface. Moreover, a sequence of related windows are returned. Starting from this information, the system generates the plan of the mission task.
Before the execution, the robot, as previously described, simulates the generated plan by using the 3D simulator, in order to check the plan itself and eventually to refine it.
After this step, the robot starts the visit. During the visit, the robot gives some general information about the museum. When the robot reaches one of the selected windows, it stops and it gives the related information previously retrieved by the IRM.
In the case that around one of the selected windows there are some visitors so that the robot is not able to reach it, CiceRobot will continue the visit and it will reschedule the skipped window. When, during the visit, CiceRobot perceives a visitor in its trajectory, the robot stops and it sends an alarm sound. During the whole visit, the images acquired by the camera are also sent via Internet to other computers in order to perform a sort of remote visit of the museum.

References

1. R. Baeza-Yates and B. Ribeiro-Neto. Modern Information Retrieval. Addison Wesley, 1999.
2. W. Burgard, A. B. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, S. Thrun, Experiences with an interactive museum tour-guide robot, Artificial Intelligence 114 (1999) 3–55.
3. A. Chella, M. Frixione, and S. Gaglio. A cognitive architecture for artificial vision. Artificial Intelligence, pp. 89:73–111, 1997.
4. A. Chella, M. Frixione, S.Gaglio, Understanding dynamic scenes, Artificial Intelligence, Vol 123, pp. 89-132, 2000.
5. A. Chella, M.Cossentino, R. Pirrone, A. Ruisi, Modeling Ontologies for Robotic Environments, The Fourteenth International Conference on Software Engineering and Knowledge Engineering - July 15-19, 2002 - Ischia, Italy.
6. Cyc Home Page, Cycorp Inc., Austin, TX. http://www.cyc.com.
7. P. Gärdenfors, Conceptual Spaces, MIT Press, Bradford Books, Cambridge, MA, (2000).
8. R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2004.
9. I. Infantino, M. Cossentino, A. Chella, "An agent based architecture for robotic vision systems", Agentcities iD3, Barcelona, Spain, 6-8 February 2003.
10. T. K. Landauer, P. W. Foltz, D. Laham, An introduction to Latent Semantic Analysis, Discourse Processes, pp. 259- 284, 1998.
11. G. Stein, "Lens distortion calibration using point correspondences", in proc. Computer Vision and Pattern Recognition Conference, pages 602-608, 1997.
12. Voice Extensible Markup Language Home Page, http://www.w3.org/TR/2004/REC-voicexml20-20040316/
13. Z. Zhang, "A Flexible New Technique for Camera Calibration", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, n. 11, pp. 1330-1334, 2000.