A Cognitive Robot Architecture Based on 3D Simulator of Robot and Environment

Antonio Chella[1] and Irene Macaluso[1]

[1] Dipartimento di Ingegneria Informatica Università di Palermo
viale delle Scienze ed.6, 90128 Palermo, Italy

Abstract. The paper proposes a robot architecture based on a comparison between the effective robot sensations and the expected sensations generated by a 3D robot/environment simulator. The robot perceptions are generated by the simulator driven by this comparison process. The architecture is operating in "Cicerobot" a museum robot offering guided tours at the Archaeological Museum of Agrigento, Italy.

1 Introduction

An autonomous robot operating in real and unstructured environments has to interact with a dynamic world populated with objects, people, and in general, other agents: people and agents may change their position and identity during time, while objects may be moved or dropped. In order to work properly, the robot should have a *perception* of its environment and it should create links between its sensory inputs and motor actions.

Humphrey [12] makes a clear distinction between *sensations* and *perceptions*: sensations are active responses generated by the body in reaction to external stimuli. They refers to the subject, they are about "what is happening to me". Perceptions are mental representations related to something outside the subject. They are about "what is happening out there". Sensations and perceptions are two separate channels; a possible interaction between the two channels is that the perception channel may be recoded in terms of sensations, i.e., the mental representation of the vase may be recoded in terms of sensations and compared with the effective stimuli from the outside, in order to catch and avoid perceptual errors. This process is similar to the "echoing back to source" strategy for error detection and correction.

Gärdenfors [6] discusses the role of *simulators* related to sensations and perceptions. He claims that sensations are immediate sensory impressions, while perceptions are built on *simulators* of the external world. A simulator receives as input the sensations coming from the external world, it fills the gaps and it may also add new information in order to generate perceptions. The perception of an object is therefore more rich and expressive than the corresponding sensations. In Gärdenfors terms, "perceptions are sensations that are *reinforced* with simulations". The role of

Chella A. and Macaluso I. (2007).
A Cognitive Robot Architecture Based on 3D Simulator of Robot and Environment.
In *Robot Vision*, pages 67-76
DOI: 10.5220/0002066700670076

simulators in motor control has been extensively analyzed from the neuroscience point of view, see [23] for a review.

Grush [10,11], proposes several cognitive architectures based on simulators ("emulators" in Grush terms). The basic architecture is made up by a feedback loop connecting the controller, the plant to be controlled and a simulator of the plant. The loop is *pseudo-closed* in the sense that the feedback signal is not directly generated by the plant, but by the simulator of the plant, which parallels the plant and it receives as input the efferent copy of the control signal sent to the plant. In this case, the sensations are generated by the system as the output of the simulator.

A more advanced architecture proposed by Grush and inspired to the work of Gerdes and Happee [8] takes into account the basic schema of the Kalman filter. In this case, the residual correction generated by the difference between the effective plant output and the emulator output (the sensations) are sent to the plant simulator via to the Kalman gain. In turns, the simulator send its inner variables as feedback to the controller. In this case the sensations are output of the simulator process and they are more or less of the same type of sensory inputs, while the perceptions are the inner variables of the simulator. The simulator inner variables are therefore more expressive that rough sensations and they may contain also information not directly perceived by the system, as the occurring forces in the perceived scene, or the object-centered parameters and in general the variables used in causal reasoning [7].

Grush [9] discusses the adoption of neural networks to learn the operations of the simulators, and Oztop et al. [19] propose sophisticated learning techniques of simulators based on inferences of the theory of mind of others.

An early implementation of a robot architecture based on simulators is due to Mel [17]. He proposed a simulated robot moving in an environment populated with simple 3D objects. The robot is controlled by a neural network that learns the aspects of the objects and their relationships with the corresponding motor commands. It becomes able to simulate and to generate expectations about the expected object views according to the motor commands, i.e., the robot learns to simulate the external environment. A successive system proposed by Mel is MURPHY [18] in which a neural network controls a robot arm. The system is also able to perform off-line planning of the movements by means of the internal simulator of the environment.

Other early implementation of robots operating with internal simulators of the external environment are MetaToto [21], and the *internalized plan* architecture [20]. In both systems, a robot builds an inner model on the environment to be reactively explored by simulated sensorimotor actions to generate action plans.

An effective robot able to build an internal model of the environment has been proposed by Holland and Goodman [13]. The system is based on a neural network that controls a Khepera minirobot that is able to simulate actions and perceptions and to anticipate perceptual activities is a simplified table environment. Holland [14] speculates on the relationships between embodiment, internal models and consciousness.

2 Robot Architecture

The robot architecture proposed in this paper is based on an internal simulator of the robot and the environment world that takes into account the previously discussed distinction between sensations and perceptions (Fig. 1). The *Robot* block is the robot itself and it is equipped with motors and a video camera. It is modeled as a block that receives in input the motor commands M and it sends in output the robot *sensations* S, i.e., the scene acquired by the robot video camera. The *Controller* block controls the actuators of the robot and it sends the motor commands M to the robot. The robot moves according to M and its output is the 2D pixel matrix S corresponding to the scene image acquired by the robot video camera.

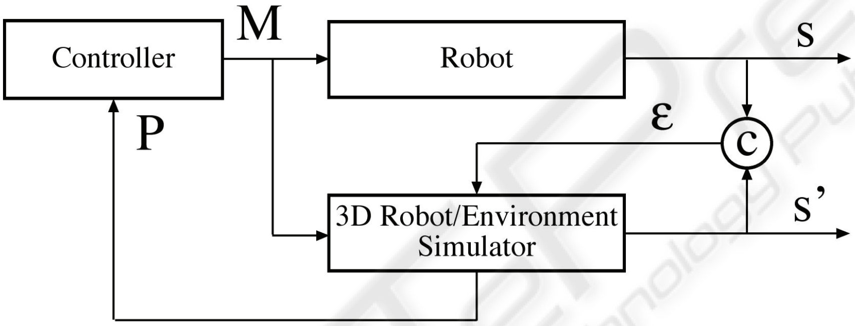

Fig. 1. The robot architecture.

At the same time, an *efferent* copy of the motor commands is sent to the *3D Robot/Environment Simulator*. The simulator is a 3D reconstruction of the robot environment with the robot itself. It is an *object-centered* representation of the world in the sense of Marr [16]. The simulator receives as input the controller motor command M and it simulates the corresponding motion of the robot in the 3D simulated environment. The output S' of the simulator is the 2D image obtained as a projection of the simulated scene acquired by the simulated robot. In this sense S is the effective image scene acquired by the robot and S' is the expected image scene acquired by the robot according to the simulator. Both images are *viewer-centered*, in Marr's terms.

The acquired and the expected image scenes are compared by the *comparator* block c and the resulting error ε is sent back to the simulator to align the simulated robot with the real robot. At the same time, the simulator send back all the relevant 3D information P about the robot position and its environment to the controller, in order to adjust the motor plans, as described below

It should be noted that S and S' are 2D image scenes; they are *modal* information that may be considered *sensations* as they refer to "what is happening to the robot", i.e., they are responses to the robot visual stimuli referred to the robot itself. Instead, P is *amodal* information that may be considered the robot *perception*, as it is a set of 3D information referred to "what is happening out there".

The image matrix S’ is the 2D recoding of 3D perception used by the simulator correct and align the simulated robot, while P is the *amodal* interpretation of robot sensations by means of the 3D simulator. The 2D reconstruction S’ of the scene is built by the robot as a projection of the 3D entities in the simulator and from the data coming from robot sensors. This reconstruction constitutes the *phenomenal* experience of the robot, i.e., what the robot sees at a given instant. This kind of seeing is an active process, since it is a reconstruction of the inner percept in ego coordinates, but it is also driven by the external flow of information. It is the place in which a global consistency is checked between the internal model and the visual data coming from the sensors. The robot acquires evidence for what it perceives, and at the same time it interprets visual information according to its internal model. Any discrepancy asks for a readjustment of its internal model. Furthermore, through this 2D image, the robot has an immediate representation of the scene as it appears in front of it, very useful for rapid decision and reactive behaviour.

3 Planning and Imagination

The proposed framework may be extended to allow the robot to deliberate its own sequences of actions. In this perspective, planning may be performed by taking advantage from the representations in the 3D Robot/Environment Simulator. Note that we are not claiming that all kinds of planning must be performed within a simulator, but the forms of planning that are more directly related to perceptual information can take great advantage from visual awareness in the described architecture.

The signal P describes the perception of a *situation* of the world out there at time *t*, i.e., it describes the perception of the current *situation* of the world. The simulator, by means of its simulation engine (see below), is able to generate expectations of P at time *t+1*, i.e., it is able to simulate the robot *action* related with motor command M generated by the controller and the relationship of the action with the external world.

In facts, the preconditions of an action can be simply verified by geometric inspections in P at time *t*, while in the STRIPS planner [5] the preconditions are verified by means of logical inferences on symbolic assertions. Also the effects of an action are not described by adding or deleting symbolic assertions, as in STRIPS, but they can be easily described by the situation resulting from the expectations of the execution of the action itself in the simulator, i.e., by considering the expected perception P at time *t+1*.

In the proposed architecture, the recognition in a scene of a certain component of a situation described by P may elicit the expectation of the other components of the situation itself. The recognition of a certain situation by means of the perception P at time *t* could also elicit the expectation of a subsequent situation and the generation of the expected perception P at time *t+1*.

We take into account two main sources of expectations. On the one side, expectations are generated on the basis of the structural information stored in a symbolic knowledge base of the simulator. We call *linguistic* such expectations. As soon as a situation is perceived which is the precondition of a certain action, then the symbolic description elicit the expectation of the effect situation, i.e., it generates the expected perception P at time *t+1*.

On the other side, expectations could also be generated by a purely *Hebbian* association between situations. Suppose that the robot has learnt that when it sees somebody pointing on the right, it must turn in that direction. The system learns to associate these situations and to perform the related action. We call *associative* this kind of expectations.

In order to explain the planning by imagination mechanism, let us suppose that the robot has perceived the current situation P_0 e.g., it is in a certain position of a room. Let us suppose that the robot knows that its goal g is to be in a certain position of another room with a certain orientation. A set of expected perceptions $\{P_1, P_2, \ldots\}$ of situations is generated by means of the interaction of both the linguistic and the associative modalities described above. Each P_i in this set can be recognized to be the effect of some action related with a motor command M_j in a set of possible motor commands $\{M_1, M_2, \ldots\}$where each action (and the corresponding motor command) in the set is compatible with the perception P_0 of the current situation.

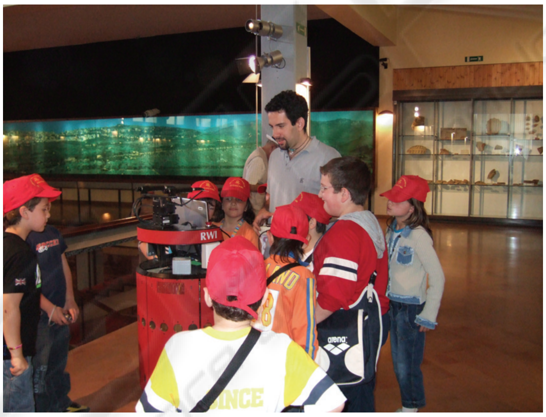

Fig. 2. The robot "Cicerobot" operating at the Archaeological Museum of Agrigento.

The robot chooses a motor command M_j according to some criteria; e.g., is the action whose expected effect has the minimum Euclidean distance from the "goal" g, or, for example, considering the valuation of the expected effect. Once that the action to be performed has been chosen, the robot can imagine to execute it by simulating its effects in the 3D simulator then it may update the situation and restart the mechanism of generation of expectations until the plan is complete and ready to be executed.

On the one side, linguistic expectations are the main source of deliberative robot plans: the imagination of the effect of an action is driven by the description of the action in the simulator KB. This mechanism is similar to the selection of actions in deliberative forward planners. On the other side, associative expectations are at the basis of a more reactive form of planning: in this latter case, perceived situations can "reactively" recall some expected effect of an action.

Both modalities contribute to the full plan that is imagined by the robot when it simulates the plan by means of the simulator. When the robot becomes fully aware of the plan and of its actions, it can generate judgments about its actions and, if necessary, imagine alternative possibilities.

4 Robot at Work

The proposed architecture has been implemented in "Cicerobot", an autonomous robot RWI B21 equipped with sonar, laser rangefinder and a video camera mounted on a pan tilt. The robot has been employed as a museum tour guide operating at the Archaeological Museum of Agrigento, Italy offering guided tours in the "Sala Giove" of the museum (Fig. 2). A first session of experimentations, based on a previous version of our architecture, has been carried out from January to June 2005 and the results are described in [4,15]. The second session, based on the architecture described in this paper, started in March and ended in July 2006.

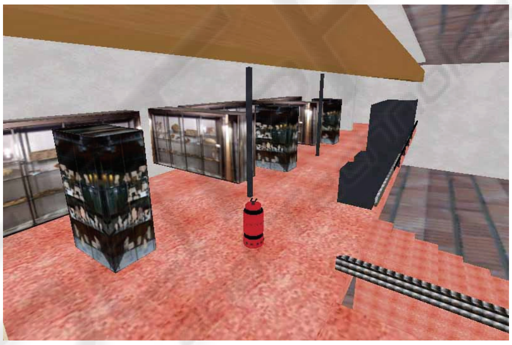

Fig. 3. The 3D robot/environment simulator.

The task of museum guide is considered a significant case study [3] because it concerns perception, self perception, planning and human-robot interactions. The task it is relevant as a test bed for phenomenal consciousness. It can be divided in many subtasks operating in parallel, and at the same time at the best of the robot capabilities. Moreover, the museum is a dynamic and unpredictable environment.

Referring to the architecture in Fig. 1, the controller includes a standard behavior-based architecture (see, e.g., [1]) equipped with all the standard reactive behaviors as the static and dynamic obstacle avoidance, the search of free space, the path following and so on.

Fig. 3 shows the 3D robot/environment simulator. As previously described, the task of the block is to generate the expectations of the interactions between the robot and the environment; it should be noted that the robot also simulates itself in its environment.

In order to keep aligned the simulator with the external environment, the simulator engine is equipped with a stochastic algorithm known as *particle filter* (see, e.g., [22]). In brief, the simulator hypothesize a cloud of expected possible positions of the robot. For each expected position, the corresponding expected image scene S' is generated, as in Fig. 4 (right). The comparator then generates the error measure ε between the expected and the effective image scene S (Fig. 4 left). The error ε weights the expected position under consideration; in a subsequent step, only the winning expected positions that received the higher weigh are taken, while the other ones are dropped. Fig. 5 (left) shows the initial distribution of expected robot position and Fig. 5 (right) shows the small cluster of winning positions.

Now the simulator receives the new motor command M related with the chosen action as described before, and, starting from the winning hypotheses, it generates a new set of hypothesized robot positions. The filter iterates between these two steps until convergence, i.e., until the winning positions converge to a single moving point.

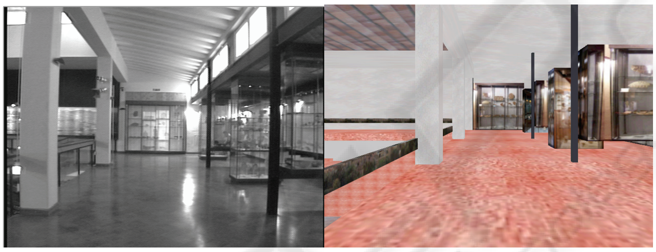

Fig. 4. The 2D image output of the robot video camera (left) and the corresponding image generated by the simulator (right).

Fig. 4 shows the 2D image S as output of the robot video camera (left) and the corresponding image S' generated by the simulator (right) by re-projecting in 2D the 3D information from the current point of view of the robot.

The comparator block **c** compares the two images of Fig. 4 by using elastic templates matching [2]. In the current implementation, features are long vertical edges extracted from the camera image. Spatial relations between edges' midpoints are used to locate each edge in the simulated image and compute the relative distortion between the expected and the effective scene. The relative distortion is a measure of the error ε related to the differences between the expected image scene and the effective scene. As previously stated, this error is sent back to the simulator in order to correct the robot position in the 3D simulator.

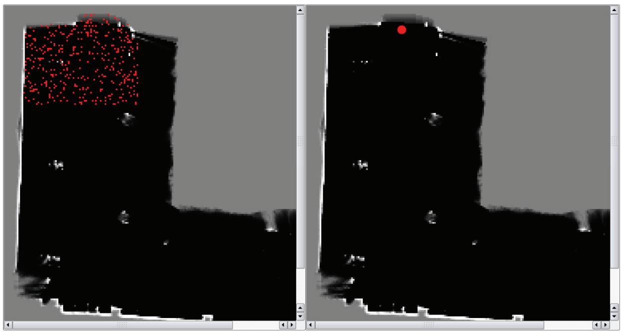

Fig. 5. The operation of the particle filter. The initial distribution of expected robot positions (left), and the cluster of winning expected positions (right).

5 Experimental Results

In order to test the proposed architecture we compared the operations of the robot equipped with the described system with the operations of the robot driven by the odometric information only.

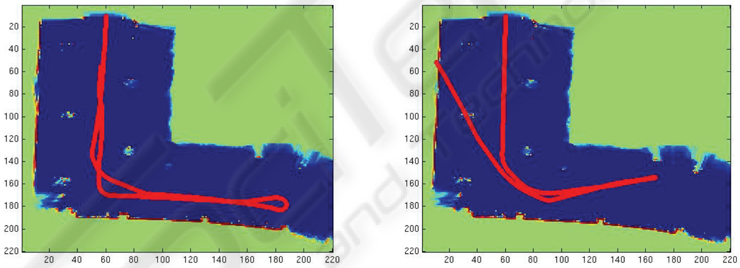

Fig. 6. The operation of the robot equipped with the described architecture (left) and with the reactive controller and the odometric feedback only (right).

Fig. 6 (left) shows the operation of the robot during the tour guide. To compare the operation of the robot, we tested the robot by considering the operation of a reactive robot equipped with the odometric sensor data; results are shown in Fig 6 (right).

Fig. 7 shows the difference of the two trajectories versus time. It should be noticed that at the very beginning the two tours are quite the same, while early in the robot operations, the trajectory related with the odometric data only tends to diverge. During the second part of the tour, when the two robots turn right in order to follow the corridor and to come back to base home, the two trajectories became again quite similar (the valley at nearly 1200) and then they diverge again.

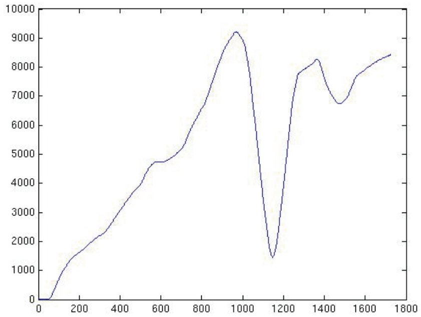

Fig. 7. The figure shows the difference during the tour guide between the two trajectories versus time.

It should be noted that the proposed architecture based on the described account let the robot to operate more precisely and in a more satisfactory way. In facts, the described sensations and perceptions mechanism let the robot to be aware of its position and of its perceived scene. The robot is therefore able to eventually adjust and correct its own subsequent motion actions. Moreover, the robot is able to imagine its future actions and it is therefore able to choose the best motion actions according to the current perceived situation.

Acknowledgements

Authors would like to thank Salvatore Gaglio, Peter Gärdenfors and Riccardo Manzotti for discussions about the proposed architecture. The director and the staff of the Archaeological Museum of Agrigento kindly supported the Cicerobot project. Edoardo Ardizzone, Massimo Cossentino, Antonio Gentile, Rosario Gradino, Ignazio Infantino, Marilia Liotta, Daniele Peri, Lorenzo Riano, Riccardo Rizzo, Giuseppe Scardino contributed in several ways to the implementation and tests of the proposed architecture.

References

1. R.C. Arkin: Behavior-based robotics. MIT Press, Cambridge, MA, 1998.
2. C. Balkenius, L. Kopp: Elastic template matching as a basis for visual landmark recognition and spatial navigation, in Proc. of AISB 1997 workshop on Spatial reasoning in mobile robots and animals, Manchester: Manchester University, UK.
3. 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, pp. 3–55, 1999

4. A. Chella, M. Frixione, S. Gaglio: Planning by imagination in Cicerobot, a robot for museum tours, in: Proc. of AISB 2005 Symposium on Next Generation Approaches to Machine Consciousness, pp. 40-49, University of Hertfordshire, Hatfield, UK.
5. R.E. Fikes and N.J. Nilsson: STRIPS: a new approach to the application of theorem proving to problem solving, Artificial Intelligence, 2, pp. 189-208, 1971.
6. P. Gärdenfors: How Homo Became Sapiens. Oxford, Oxford University Press, 2003.
7. P. Gärdenfors: Emulators as sources of hidden cognitive variables, Behavioral and Brain Sciences, 27, 3, p. 403, 2004.
8. V.G.J. Gerdes and R. Happee: The use of an internal representation in fast goal-directed movements: a modeling approach, Biological Cybernetics, 70, pp. 513-524, 1994.
9. R. Grush: Emulation and cognition. Doctoral dissertation, Department of Cognitive Science and Philosophy. University of California, San Diego, 1995.
10. R. Grush: Wahrnehmung, Vorstellung und die sensomotorische Schleife. (English translation: Perception, imagery, and the sensorimotor loop), in: F. Esken and H.-D. Heckmann (eds.): Bewußtsein und Repräsentation, Verlag Ferdinand Schöningh.
11. R. Grush: The emulator theory of representation: Motor control, imagery and perception, Behavioral and Brain Sciences, 27, 3, pp. 377-442, 2004.
12. N. Humphrey: A History of the Mind. New York, Simon & Schuster, 1992.
13. O. Holland and R. Goodman: Robots with internal models – A route to machine consciousness? Journal of Consciousness Studies, 10, 4-5, pp. 77-109, 2003.
14. O. Holland: The future of embodied artificial intelligence: Machine consciousness?, in: F. Iida et al. (eds.): Embodied artificial intelligence. LNAI 3139, Berlin Heidelberg, Springer-Verlag, pp. 37-53, 2004.
15. I. Macaluso, E. Ardizzone, A. Chella, M. Cossentino, A. Gentile, R. Gradino, I. Infantino, M. Liotta, R. Rizzo and G. Scardino: Experiences with Cicerobot, A museum guide cognitive robot, in: S. Bandini and S. Manzoni (eds.): AI*IA 2005, LNAI 3673, Berlin Heidelberg, Springer-Verlag, pp. 474-482, 2005.
16. D. Marr: Vision. W.H. Freeman, New York, 1982.
17. B.W. Mel: A connectionist learning model for 3-dimensional mental rotation, zoom and pan, in: Proc. of the 8th Ann. Conf. of the Cognitive Science Soc., pp. 562-571, 1986.
18. B.W. Mel: Connectionist robot motion planning: A neurally-inspired approach to visually-guided reaching. Cambridge, MA, Academic Press, 1990.
19. E. Oztop, D. Wolpert, M. Kawato: Mental state inference using visual control parameters, Cognitive Brain Research, 22, pp. 129-151, 2005.
20. D.W. Payton: Internalized plans: A representation for action resources, Robotics and Autonomous Systems, 6, pp. 89-103, 1990.
21. L.A. Stein: Imagination and situated cognition, MIT AI Memo No. 1277, 1991.
22. S. Thrun, W. Burgard, D. Fox: Probabilistic Robotics. MIT Press, Cambridge, MA, 2005.
23. D.M. Wolpert and Z. Ghahramani: Computational principles of movement neuroscience, Nature neuroscience supplement, 3, pp.1212-1217, 2000.