Transactions on Information and Communications Technologies vol 8, © 1995 WIT Press, www.witpress.com, ISSN 1743-3517

The role of perception and action in intelligent systems

A. Pasqual del Pobil
Department of Computer Science, Jaume I University, Penyeta Roja Campus, E-12071 Castellón, Spain

Abstract

Robotics plays an important role in the applications of Artificial Intelligence in Engineering, since it deals with the interaction of an intelligent system with the world by means of perception and action. However, Robotics has been traditionaly considered as just a mere application area of AI. In this paper, the role of perception and action in current AI systems is analyzed and some recent fundamental results concerning the use of Robotic Intelligence for applications of AI in Engineering are discussed, in order to build artificial systems that behave in an intelligent way in the real world. Finally, two implemented systems based on this methodology are described to show the relevance of the proposed strategy to Engineering applications, one using visual perception, and another based on force/torque sensing.

1 Motivation

In the first page of the second edition of his celebrated textbook P.H. Winston defines Artificial Intelligence as "the study of ideas that enable computers to be intelligent" [Winston 84]. In 1984 J.M. Brady gave his famous definition of Robotics Science as "the intelligent connection of perception to action" [Brady 84]. Curiously enough, in the third, revised and expanded edition of his textbook, Winston changes his previous definition of Artificial Intelligence to state that it is "the study of the computations that make it possible to perceive, reason, and act" [Winston 1992]. The resemblance of these last two definitions for Robotics Science and Artificial Intelligence seems to suggest that these two different sciences should be moving in the same direction, or even more, they should converge towards the same goal. This coincidence is by no means casual, and it serves us to motivate this paper: Robotic Intelligence (RI) must play a fundamental role in Artificial Intelligence (AI) so that it is properly oriented and founded.

We propose that the objectives and methodologies of current AI systems should shift towards what we call *Robotic Intelligence* as opposed to pure abstract, symbolic AI. In the rest of this paper we will support this assertion.

A remark must be made now regarding what we consider to be the ultimate and short-term goals of AI. We accept the first of Winston's

Transactions on Information and Communications Technologies vol 8, © 1995 WIT Press, www.witpress.com, ISSN 1743-3517

definition, as stated above, or Minsky's when he says that "AI is the science of making machines do things that would require intelligence if they were made by men"; but we want that these artificial systems behave in an intelligent way *in the real world.*

2 Extending the Turing Test

Since it was formulated in 1950, the so-called Turing Test (T2) was generally accepted as the right, definitive test to discriminate an intelligent computer from a non-intelligent one [Turing 50]. Basically, in the test there is a person communicating via a teleprinter with a computer and another person, but ignoring what line is connected to the person and what to the computer. If the person cannot tell how the lines are connected after a dialogue through both lines, then the computer is said to have passed T2 and it can be rated as intelligent. It has to be noted that T2 is implicitly assuming that intelligence is the ability to reason and to communicate by language.

In 1980 John Searle put forward a thought experiment with the intention of showing that computers cannot really understand what they do [Searle 80]. Since then, it has raised much controversy among the AI community, and discussions still keep on and on as to its correctness. In essence, Searle's Chinese Room argument assumes that in T2 the language used for communicating is Chinese instead of English, and the computer is replaced by a person, Searle himself, called the operator, locked in the Chinese Room. The operator understands no Chinese at all, but he is provided with a set of instructions in English to manipulate Chinese written symbols in such a way that, following these instructions, the operator is able to produce a set of symbols as the output to a given input set of Chinese symbols. Now, if this operator is able to pass T2 in Chinese, we should conclude that he understands Chinese, while the operator, being Searle himself, does not know a single word of this language, on the contrary, all he has done is following the instructions for manipulating meaningless symbols. An immediate consequence of this argument is that T2 is not the definitive test for intelligence: a computer passing it understands no more what it is doing than Searle understands Chinese.

Stevan Harnad has proposed the Total Turing Test (T3) as an extension of T2 that is not invalidated by Searle's argument [Harnad 89]. In T3 the computer is replaced by a robot, and the person carrying out the test is not communicating through a teleprinter but actually seeing the candidate robot and a real person, while both are operating directly on the world. If after a certain amount of time (as long as desired, even lifelong) the observer is not able to say which is the robot and which the person, then the robot has passed the test and it can be rated as intelligent. The key point is that now, in addition to reasoning and communicating by language, the candidate must exhibit all robotic capacities a person has, including the ability to see, grasp, manipulate, move, hear, recognize, etc., in a way that is indistinguishable from those of a person. What really matters in T3 is that robotic capacity has been integrated as an inseparable part of the definitive test for intelligence. If we want an intelligent system to pass T3, it must be endowed with *Robotic* Intelligence.

3 The Symbol Grounding Problem

Classical AI is based in the use of pure symbol systems, i.e., following the traditional distinction between a symbolic level (the software) and its implementation in a particular computer (the hardware). The "Symbol Grounding Problem" was yet another challenge to pure symbolic AI [Harnad 90]. The symbols in a symbol system are systematically interpretable as meaning something; however, in a typical AI system, that interpretation is not intrinsic to the system, it is always given by an external interpreter (e.g., the designer of the system). Neither the symbol system in itself nor the computer, as an implementation of the symbol system, can ground their symbols in something other than more symbols. The operator in the Chinese Room will never be able to understand Chinese because it is somebody else who knows the interpretation of Chinese symbols, the one who designed the instructions for manipulating them. And yet, when we think, unlike computers, we use symbol systems that need no external interpreter to have meanings. The meanings of our thoughts are intrinsic, the connection between our thoughts and their meanings is direct and causal, it cannot be mediated by an interpreter, otherwise it would lead to an infinite regress if we assume that they are interpretable by someone else.

Again, the solution to this paradox is in Robotic Intelligence systems instead of pure symbolic AI systems [Harnad 93]. In an RI system, with T3-level performance, the symbols are grounded in the system's own capacity to interact robotically with what its symbols are about [del Pobil, Escrig, Jaen 94]. Such an RI system should be able to perceive, manipulate, recognize, classify, modify, ..., and reason about the real-world objects and situations that it encounters. In this way, its symbols would be grounded in the same sense that a person's symbols are grounded, because it is precisely those objects and situations that their symbols are about. If we think of a symbol that corresponds to a word, we ground it when we first learn our mother tongue through interaction with the outer world, because we cannot obviously ground it in more words. In this respect, for a blind child the meanings of its symbol system must necessarily differ from those of a normal child, because its interaction with the world is severely handicapped.

A possible answer to the question of how to ground basic spatial concepts is the use of connectionism. Neural nets can be a feasible mechanism for learning the invariants in the analog sensory projection on which categorization is based [Harnad 93], [Martin, del Pobil 94], [Cervera, del Pobil, Marta and Serna, 95]. This is further discussed in section 6.

4 The Right Level of Competence

Another point in favour of Robotic Intelligence concerns the appropriate level of competence an AI system should exhibit. Sometimes there exists a lack of balance between AI systems and natural systems in some aspects of their competence. For example, there are chess-playing systems that are able to reach grand master level of competence, only being defeated by a few persons in the world; or expert systems that show an expert competence in, say, diagnosing infectious diseases. And, on the other hand, there is no existing system that surpasses the competence of a cockroach in moving around with a goal in an unstructured world. This enormous distance tends to be always between pure abstract intellectual tasks at one end, and robotic tasks, at the other, i.e., those that involve sensorimotor interaction with the real world. In the case of human level competence, not to speak of cockroaches, the gap between these two levels of competence is still larger. Our simplest, everyday, common-sense robotic capacities are very far from what robots can currently do [del Pobil, Serna 1995]: our artificial grand master would inevitably *die* in case of fire just because it would not be able to find the exit door and turn the

Transactions on Information and Communications Technologies vol 8, © 1995 WIT Press, www.witpress.com, ISSN 1743-3517

handle to open it (turning door handles is one of the tasks that current robots cannot do in a general case).

5 The Role of Perception in Reasoning

Even accepting that perception and action as robotic capacities must play a fundamental role in any intelligent system, it is usually assumed that sensorimotor abilities can be isolated from the rest of the system and just be implemented as input/output modules that interface to the main processing --or pure abstract reasoning-- unit. This is the Turing vision of AI that is usually assumed as a start point.

This *modularity hypothesis* cannot be justified by any evidence whatsoever, neither from neurophysiology, nor from biology or cognitive psychology. It is just a traditional problem reduction methodology that is systematically used to tackle hard problems. However, in natural living systems, cognitive abilities are intimately and inseparably tied to perception and (maybe to a smaller extent) action capacities. Psychophysical evidence suggests this fact and, moreover, that the representation of the world used by an intelligent system is directly dependent on its perception skills [Brooks 91].

6 Implemented Engineering Applications

An immediate consequence of the previous discussion is that reasoning models or representations should include certain reference to perception; what we call *perception-based reasoning* [Cervera and del Pobil 95]. We are going to show two implemented applications based on the described strategies, that are relevant to Engineering problems.

6.1 Perception-based qualitative spatial reasoning

We advocate that a coherent approach to spatial reasoning should include al least a description of how it is to be integrated with perception. Moreover, perception must provide a bottom-up grounding for the spatial concepts that are represented in the system by means of symbols.

In this approach, the representation of spatial knowledge as described in [del Pobil, Escrig and Jaén, 93] is based on a spherical reference system that is directly related to the spherical perspective model for visual perception. In addition, this model offers some interesting advantages when compared with orthographic and planar perspective models [Penna and Chen, 90]. To capture the salient geometric features of objects that are used below to define ideal meanings for perceptually salient relations, an spherical representation for objects is used [del Pobil and Serna, 94a]. This representation has been successfully applied to some problems in robotics, as collision detection [del Pobil, Serna and Llovet, 92], find-path [del Pobil and Serna, 92] and motion planning [del Pobil and Serna, 92]. It has been recently extended [del Pobil, Martínez and Calvo, 94]. This provides a direct connection between our spatial reasoning model and a low-level description from a vision system. Indeed, it is possible to compute a volume description of objects as generalized cones from sparse, imperfect 3-D data, such as may be obtained from stereo vision [Rao and Nevaitia, 90].

According to the usual notation for spherical coordinates, a point in space will be given by the three coordinates (ρ, θ, ϕ). An spherical three-dimensional interval will be given by $[\rho \pm \Delta\rho]$, $[\theta \pm \Delta\theta]$, $[\phi \pm \Delta\phi]$.

Transactions on Information and Communications Technologies vol 8, © 1995 WIT Press, www.witpress.com, ISSN 1743-3517

As a first approximation, an object will be perceived as the spherical interval that completely encloses the object stereo image. If the spherical representation is used, it is easy to see that for each sphere we would have such an interval with

$$\Delta\rho \approx R, \qquad \Delta\theta \approx R/\rho \sin \phi, \qquad \Delta\phi \approx R/\rho,$$

where (ρ, θ, ϕ) denotes the center of the sphere.

We consider several possible spatial relations between objects corresponding to commonsense spatial concepts as given by ideal meanings of some English prepositions. Our model is adequate to deal with prepositions such as *over, under, between, beyond, near, far*; and the egocentric meanings of *in_front_of, behind, to_the_left_of,* and *to_the_right_of*. By way of illustration, let us analyze in detail the case of the relation *under*.

According to Herskovits [86], the ideal meaning of *under* can be defined as being the *partial inclusion of a geometrical construct in the lower space defined by some surface, line, or point.* For the first representation of an object as given by the smallest enclosing sphere, this lower space would be the volume of the cylinder that prolongs from the sphere in the downward direction. However, from the point of view of the observer's perception and the corresponding spherical vision model, that ideal meaning amounts to:

i. Both objects must share the same longitude, i.e., the same θ-coordinate.

ii. Both objects project on the horizontal plane at the same distance from the viewer.

These two conditions assure that one spatial entity is *under* the other, but to establish that it is precisely object *A* that is under object *B* we must add another condition:

iii. The latitude of *B* is less than that of *A*.

These equalities must obviously be understood as qualitative equalities. To put it in a formal way, given two objects *A* and *B* perceived as

$$[\rho_A \pm \Delta\rho_A],\ [\theta_A \pm \Delta\theta_A],\ [\phi_A \pm \Delta\phi_A], \text{ and}$$
$$[\rho_B \pm \Delta\rho_B],\ [\theta_B \pm \Delta\theta_B],\ [\phi_B \pm \Delta\phi_B],$$

the qualitative relation *under* will hold between them if and only if:

i. $[\theta_A \pm \Delta\theta_A] \cap [\theta_B \pm \Delta\theta_B] \neq \varnothing$

ii. $[\rho_A \sin\phi_A \pm R_A] \cap [\rho_B \sin\phi_B \pm R_B] \neq \varnothing$

iii. $[\phi_B - \Delta\phi_B] < [\phi_A + \Delta\phi_A]$

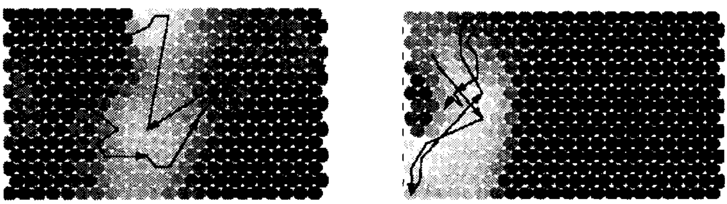
6.2 Robotic Fine Manipulation Based on Perception

The second engineering application based on the above strategies deals with robotic fine manipulation involving contact. An off-line planning system by itself can hardly cope with position uncertainties that always arise in real-world conditions, rather perception-based reasoning is called for. In Fig. 1 one of this tasks is shown, a robot manipulator is inserting a tool into a pallet lying on a robot vehicle. Visual perception is not enough to manage this kind of tasks since accuracy of the order of millimeters is required, then a force/torque sensor attached to the robot wrist is necessary. The sensor provides six signals for a certain task. A direct interpretation of this signals is impossible in a complex real task involving contact. The aproach is described in length in [Cervera, del Pobil, Marta and Serna, 1995].

Transactions on Information and Communications Technologies vol 8, © 1995 WIT Press, www.witpress.com, ISSN 1743-3517

Figure 1. A task requiring perception-based reasoning

The perception-based methodology uses a hybrid approach to connect sensor signals with the high-level qualitative task planner. A subsymbolic system based on neural networks is applied to project the data from the sensor space onto a planar self-organizing map. Fig. 2 shows an example of the result of the task in Fig. 1 for a correct case and for an incorrect one due to an error of just 4 mm. The lighter *neurons* correspond to an activation after learning a given task. The paths correspond to the winner neurons in each sequence. It can be clearly seen how the two situations differ. With this, regions on the map are assigned to particular qualitative contact states that are used by the planner. It must be noted that this qualitative states are not known a priori, since Kohonen's maps use an unsupervised learning scheme.



Correct task Incorrect task

Figure 2. Projection of perception space onto a neural map

Transactions on Information and Communications Technologies vol 8, © 1995 WIT Press, www.witpress.com, ISSN 1743-3517

Conclusion

A fundamental discussion regarding the interaction of an intelligent system with the world by means of perception and action has been presented. The role of perception and action in current AI systems has been analyzed and some recent fundamental results concerning the use of Robotic Intelligence for applications of AI in Engineering were discussed, in order to build artificial systems that behave in an intelligent way in the real world. Finally, two implemented systems based on this methodology have been described to show the relevance of the proposed strategy to Engineering applications. First a framework for a qualitative approach to commonsense reasoning about space has been further developed by considering the adequate process to build fundamental spatial relations from bottom-up, starting from the data provided by a visual perception system with the spherical perspective model. Then, an approach to a realistic robotic task involving manipulation with uncertainty has been presented that is based on coupling a subsymbolic and symbolic systems. Reasoning was based on force and torque sensing the second example.

Traditional AI research has made an excessive use of *toy models* to show the interest of certain theories. Engineering applications take place typically in the real world under real conditions, for which these toy models are of no use at all, and for which perception and action is often necessary The main contribution of this paper is to clarify the right methodology with which this kind of engineering AI applications should be approached.

References

Cervera, E., del Pobil, A.P. (1995) "Perception-Based Qualitative Reasoning in Manipulation with Uncertainty", technical report, Dept. of Computer Science, Universitat Jaume I.

Cervera, E., del Pobil, A.P., Marta, E., Serna, M.A. (1995) "Unsupervised Learning for Error Detection in Task Planning", technical report, Dept. of Computer Science, Universitat Jaume I.

del Pobil, A.P., Escrig, M.T., Jaen, J.A., (1993) "An Attempt Towards a General Representation Paradigm for Spatial Reasoning", *Proc. IEEE Intern. Conf. on Systems, Man and Cybernetics*, vol. 1, pp. 215-220, Le Touquet, France.

del Pobil, A.P., Martínez, B., Calvo, M.A., (1994) "An Extended Spherical Representation for Physical Objects", *IEEE Intern. Conf. on Systems, Man and Cybernetics*, San Antonio, Texas.

del Pobil, A.P., Serna, M.A. (1995) *Spatial Representation and Motion Planning*, Springer-Verlag.

del Pobil, A.P., Serna, M.A., (1994a) "A New Object Representation for Robotics and Artificial Intelligence Applications", *International Journal of Robotics & Automation*, vol. 9, no. 1, pp. 11-21.

del Pobil, A.P., Serna, M.A., (1994b) "A Simple Algorithm for Intelligent Manipulator Collision-Free Motion", *Journal of Applied Intelligence*, vol.4, pp. 83-102.

del Pobil, A.P., Serna, M.A., (1992) "Solving the Find-Path Problem by a Simple Object Model", *Proc. 10th European Conference on Artificial Intelligence*, pp. 656-660, Vienna, Austria.

del Pobil, A.P., Serna, M.A.,Llovet, J., (1992.) "A New Representation for Collision Avoidance and Detection", *Proc. IEEE International Conference on Robotics and Automation*, pp. 246-251, Nice, France

Harnad, S. (1989) Minds, Machines and Searle. Journal of Theoretical and Experimental Artificial Intelligence 1: 5-25.

Harnad, S. (1990) The Symbol Grounding Problem. Physica D 42: 335-346.

Harnad, S. (1993) Grounding Symbolic Capacity in Robotic Capacity, in L. Steels, R.A. Brooks (eds.) The Artificial Life route to Artificial Intelligence. Lawrence Erlbaum.

Herskovits, A., (1986) *Language and Spatial Cognition*, Cambridge University Press, Cambridge, U.K..

Martin, P., del Pobil, A.P., (1994) "Application of Artificial Neural Networks to the Robot Path Planning Problem", in *Applications of Artificial Intelligence in Engineering IX* edited by G. Rzevski, R.A. Adey and D.W. Russell, Computational Mechanics Publications, Boston, pp. 73-80.

Musto, D., Konolige, K. (1993) "Reasoning about Perception", AICOM 6: 207-212.

Penna, M.A., Chen, S., (1990) "Spherical Analysis in Image Understanding" in *Advances in Spatial Reasoning*, edited by S. Chen, vol. 2, pp. 253-283, Ablex Publishing Corp., Norwood, NJ.

Rao, K., Nevaitia, R., (1990) "Computing Volume Descriptions from Sparse 3-D Data" in *Advances in Spatial Reasoning*, edited by S. Chen, vol. 2, pp. 65-96, Ablex Publishing Corp., Norwood, NJ.

Searle, J. R. (1980) Minds, brains and programs.Behavioral and Brain Sciences 3: 417-424.

Turing, A.M. (1950) "Computing Machinery and Intelligence", *Mind* LIX, no.2236, 433-60.

Acknowledgements. This paper describes research done at the Robotic Intelligence Research Group of Jaume I University. Support for the research is provided in part by the Comision Interministerial de Ciencia y Tecnologia under project TAP92-0391-C02-01, and in part by the Fundació Caixa de Castello under grants A-36-IN and B-41-IN.

The author is grateful to Stevan Harnad for his comments on an earlier draft of this paper.