Objects, Actions and Physical Interactions

Angel P. Del Pobil, Enric Cervera and Eris Chinellato

Robotic Intelligence Laboratory,
Universitat Jaume I, E-12071 Castelló, Spain
{pobil, ecervera, eris}@icc.uji.es

Abstract

We deal with a problem that can be considered as part of the symbol grounding problem and is related to anchoring, but for a more general case that includes symbols that do not denote directly physical objects but rather the physical sensorimotor interactions between a robot body and physical objects. The fundamental role of this kind of symbols for robotic intelligence can be derived from the evolutionary importance of those symbols for living organisms as supported by current neurophysiology. We provide a detailed example of this approach in the context of a manipulation task: the peg-in-hole insertion problem.

Introduction

This paper addresses a problem that can be considered as part of the symbol grounding problem. It is closely related to anchoring, but for a more general case that includes symbols that do not denote directly physical objects but rather the physical sensorimotor interactions between a robot body and physical objects. The fundamental role of these kind of symbols in living organisms is supported by neurobiology and we describe how neural networks can provide a suitable method for mapping such complex perceptual signals to that kind of symbols in a particular robot manipulation scenario.

The symbol grounding problem is a classical challenge for AI (Harnad 90). The symbols in a symbol system are systematically interpretable as meaning something; however, in a traditional 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. And yet, when we reason, 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. Though this is a more general problem than anchoring, a solution to this paradox may lie in Robotic Intelligence (RI) (Harnad 95, del Pobil 98): in an RI system the symbols should be grounded in the system own capacity to interact robotically with what its symbols are about. 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 his/her symbol system must necessarily differ from those of a child with intact vision, because his/her interaction with the world is handicapped.

Whereas strictly speaking anchoring focuses on perceivable physical objects (Coradeschi and Saffiotti 2003), and symbol grounding includes more abstract symbols like 'justice' or 'beauty', we claim that there are still symbols that are fundamental for robotic intelligence but do not refer directly to physical objects but rather to the physical sensorimotor interactions between the robot itself and the objects in the world. This kind of symbols would be more related to action, and they seem to have appeared in the evolutionary landscape long before vision as 'sight' (Goodale and Westwood 2004). This distinction between 'vision for action' as opposed to 'vision for perception' is supported by neurophysiological findings as we discuss in section 2. In section 3 we comment on a particular implementation for robotic grasping, in which each symbols refers to a particular interaction between the robot hand and a physical object.

A possible answer to the question of how to ground/anchor such symbols 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 95); in section 4 we provide a detailed example of this approach in the context of another manipulation task: the peg-in-hole insertion problem.

Lessons from Neurophysiology: Object-Oriented and Action-Oriented Vision

Motor primitives are a type of basic behaviors common to robots and humans. Such primitives show different levels of complexity, and compose hierarchically to form a behavior vocabulary. Complex movements and action sequences are composed in a almost linguistic way from this motor vocabulary. The question is, can such motor primitives be considered as symbols and thus extend the symbol grounding problem to a larger universe?

From neuroscience, two of the most important discoveries of the last 20 years support the idea of extending the symbol concept. The first of these findings is that of mirror neurons. This type of neurons can be found in a purely motor area in monkeys (F5) but show responsiveness to the observation of actions performed by others. Mirror neurons are normally related with one particular action (most studies focused on reaching, pushing, grasping) and activate only in two conditions: when the subject perform that specific action and when the subject observe someone else (human or monkey, but not something else, e.g. a robot) perform that action. Therefore, these neurons seem to codify actions in a semantic way (Rizzolatti and Arbib 98). This idea is reinforced by the fact that Broca's area in humans, traditionally related to language production, is the most likely correspondent of F5. Taking a step farther, more and more researchers are now arguing that language evolved from a neural motor system involved in action recognition (Keysers et al. 2003). This would confirm that action understanding, recognition and mental imagery of actions do not differ too much from object recognition or imaging, and that complex cognitive processes emerge from simple behaviors which firstly evolved in order to give the organism skills for better interacting in its environment.

The second fundamental argument backing the idea of symbolic actions is the distinction between the two main visual pathways going from primary visual cortex V1 to cortical association areas. It has been observed that visual processes related to specific actions in primates are different from visual processes which are not explicitly oriented to interaction of the subject with the environment. Looking at an object with the purpose of interacting with it (e.g. reaching, hitting, pushing, grasping, …) activates a dorsal neural pathway which is not active when actions are not involved. This activation seems to represent a “potential action”.

In fact, there are two visual pathways going from primary visual cortex to different association areas, the posterior parietal cortex (PPC) and the inferior temporal cortex (IT) (see Fig. 1). The traditional distinction talks about ventral "what" and dorsal "where/how" visual pathways. This distinction has been confirmed but has also evolved to the extent of identifying the ventral pathway with object visual recognition and coding and the dorsal stream with action recognition and coding (Milner and Goodale 95). The claim is that of a duplex nature of the visual system, in which perceptual information streams are directed toward different cortical areas according to the purpose of the observation (Goodale and Westwood 2004).

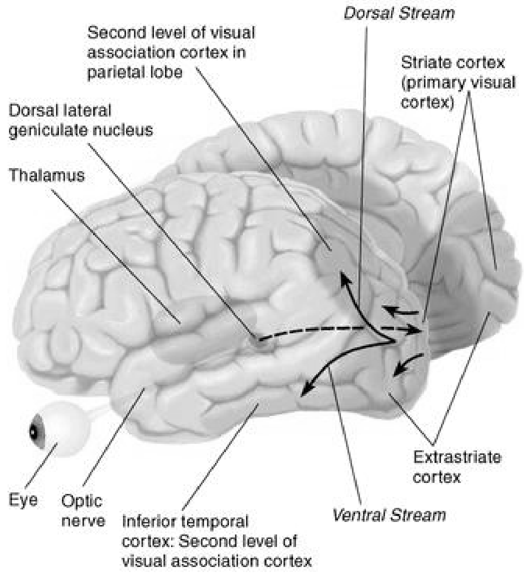
Dorsal Stream
Second level of visual association cortex in parietal lobe
Striate cortex (primary visual cortex)
Dorsal lateral geniculate nucleus
Thalamus
Extrastriate cortex
Eye
Optic nerve
Inferior temporal cortex: Second level of visual association cortex
Ventral Stream

Figure 1. Dorsal and ventral visual pathways; from http://homepage.psy.utexas.edu

Researches focused on PPC (target of the dorsal visual stream) suggest that even in this area actions are coded not only in a pragmatic way, but also in a semantic one. In fact, neurons are found which discharge both when the subject is performing an object-oriented action, and when observing the same object with the goal of acting on it (Fadiga et al. 2000). Studies on people having their ventral stream impaired show that there is no need to be able to name or even recognize an object in order to properly code an action to perform on it.

Grasping in primates and robots: hand-object interactions as symbols

A field in which the above-described motor primitive codification can be especially observed is that of reaching and grasping. As previously explained, prototypes of simple actions (both cognitive and practical ones) are coded in primate brains, and the way they connect generates higher level behaviors. In the case of grasping and reaching, areas F4 and F5 of the inferior premotor cortex of macaque monkeys are believed to contain a vocabulary of motor actions of different complexity, duration, significance (e.g. preshape the fingers for a precision grip) (Rizzolatti and Arbib 98). Such actions are selected and combined in different ways according to the task (e.g. push or grasp), the object shape and size, the timing of the action and other aspects

Area F5 is strictly connected with the anterior intraparietal area (AIP) of posterior parietal cortex, which is one of the main targets of the dorsal visual stream, being the area commonly related with grasping actions. As for other zones of PPC, some neurons of AIP are found to be active when grasping some particular objects, but also when looking at them with the purpose of grasping (and only in those cases) (Fadiga et al. 2000). Some other neurons of the same zone are sensitive to the size or orientation of the objects, and to their affordances: the intrinsic visual features which codify the ways in which they can be grasped. Therefore, AIP area encodes the 3D features of objects in a way that is suitable to guide the movements for grasping them, movements that are stored in the premotor cortex. Similar results seem to be valid for humans as well (Culham 2004).

Summarizing, area AIP codifies visual information in a grasp oriented way, storing object affordances and communicating them to area F5, which contains the motion primitives used to compose the required grasping action. Hence, we have a behavioral situation in which humans (and other primates) use symbols to interact with objects, but such symbols do not codify the objects themselves. Instead of this, they codify action-oriented visual features, or even association conditions between objects and distal subject effectors.

Hence, a grasp codifies a relation between the hand and the object, or between a tool and the object, and if this is true for humans, the application to robotics sounds straightforward. An example of this is in the grasping approach developed by Morales et al. (Morales et al. 2004), in which visual information is used in a grasp oriented manner, in a way similar to that of the dorsal stream in primates. According to this view, there is no need to model, recognize, classify an object in order to grasp it. The symbols derived from visual information identify grasps, i.e. particular physical interactions between the robot hand and the target object.

Physical Interactions as Symbols, the Peg-in-Hole Example

This section describes our approach to extract symbolic information from sensor data. The symbols in the symbol system refer to different physical contact states. Our study is based on simulations of the two-dimensional peg-in-hole insertion task. In Fig. 2 the geometry of our model is shown. It is a rectangular peg of width W_p and height H_p, to be inserted in a chamferless hole of width W_h and infinite depth in an x-y plane. Let O - xy be a coordinate system attached to the hole. The coordinates of the peg (x_p,y_p) are given by its leftdown corner, and the orientation by the angle α with X axis. In our simulations we only consider positive angles in the interval $]0,\pi/2[$, since negative angles only add symmetry to the problem. Thus, the peg is described with coordinates (x_p,y_p,α), which define the configuration space. In order to identify contacts, we should measure the forces between the peg and hole. Wrist-mounted force sensors are used to measure the external forces and torque applied to the peg as it interacts with the environment.

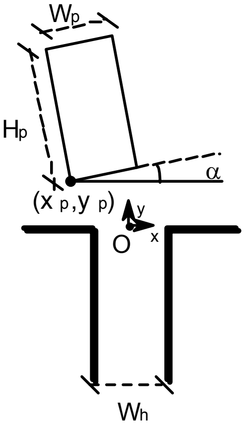

Figure 2. Peg and hole geometry

A schematic view of a gripper grasping a peg and our simplified model is depicted in Fig. 3. This sensor gives us the measures of force (fx,fy) along two axes of a coordinate system attached to it and a torque signal m with respect to O. It is worth noting that differently from the standard anchoring problem, the symbols to be anchored do not denote physical objects but rather physical interactions between the robot —the peg can be considered as a prolongation of the robot gripper— and the environment; the correct identification of these interactions is of fundamental importance for the adequate execution of the task: in this particular case the insertion of the peg into the hole.

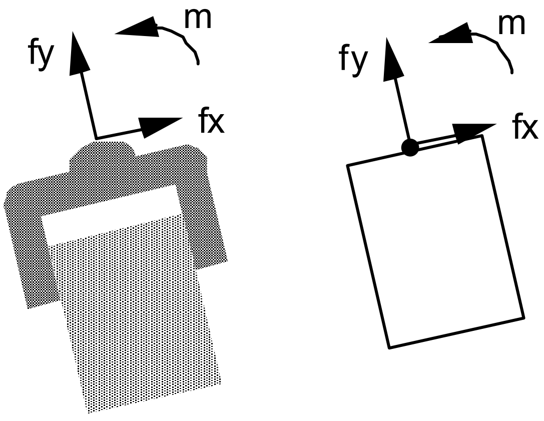

Figure 3. Schematic view of the gripper and task model

We are interested in identifying contact states with the help of the force sensor. In Fig. 4, all these contact states, including no contact, are shown. The no-contact state shows the weight force, and the others only show the reaction forces, but weight is considered too. A quasi-static model is used, i.e. inertia forces are neglected. In order to identify the contact states, only the direction of forces is relevant. We choose an arbitrary modulus, namely the unit vector. The Coulomb friction model is used, and the static and dynamic friction coefficients are considered to be equal. Obviously the sensor does not provide us directly with the symbolic contact state, rather it gives us three raw signals of force and torque. The problem is then to map these signals to the contact state appropriately. Fig. 4 shows that this is a non-trivial problem, due to the variability of the forces and the superposition of the peg weight and one or several reaction forces.

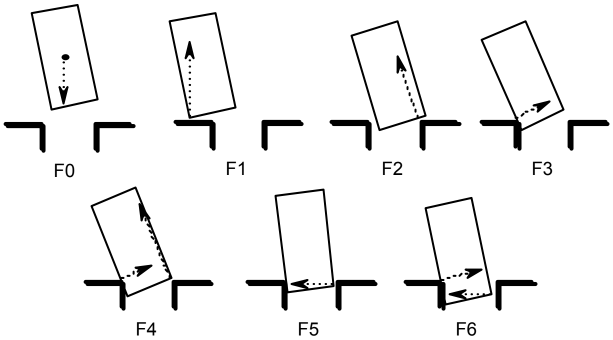

Figure 4. Set of contact states

In our approach, we use the Self-Organizing Map, which is a type of unsupervised neural networks, introduced by Kohonen. Its theory and applications are thoroughly explained in (Kohonen 95). There is an input layer which is fully connected to all the units of the network, and each unit has an output. Unlike other multilayer neural networks, the neighboring neurons of the self-organizing map co-operate during training, providing a more powerful system than a usual multilayer neural network.

Unsupervised networks, unlike supervised ones, do not need output information during training, they are trained with a set of input data alone. Their main advantage is flexibility: we are not constraining the network to learn some a-priori states. The network self-organizes discovering regularities or correlations in the input data. Later on, when the training is over, we test the network response on those a-priori states. Hopefully, identification of the states will be possible if the network's response is different for each state. If this does not occur, some states are ambiguous, i.e., they cannot be identified with only the information provided in the input signals. This is an important result which supervised networks are unable to show, and allows the designer to rearrange the input information to overcome those ambiguities. Another interesting advantage is that the network can discover new states that were not considered a-priori. If this occurs, there will be some network response that we cannot associate to any of the known states. Studying the input values which caused that response will allow to identify the unforeseen situation.

In our experiments, the appropriate friction forces are chosen from the friction cones with a uniform random probability. The peg angle is kept positive or zero. Since we are mainly interested in the influence of the task parameters, and not in those of the network, we will keep the network dimensions constant (a lattice of 15x10 hexagonally connected units) as well as the training parameters (learning rate, neighborhood, etc.). We investigate the performance of the network for several combinations of clearance and friction (μ) parameters.

The experiments consist of three phases. Each phase involves an independent set of samples which are randomly generated. The same number of samples for each contact state is chosen. Any parameter subject to uncertainty is considered to have a uniform probability density function. In the same way, any random choice is equally probable. Each sample consists of the two force components, normalized to unit modulus, and the appropriate torque value.

The three phases are the following:

1) Training. The weights are randomly initialized. The neural network is trained with a set of 1200 random input samples. This process is split into two iterations. The first one (ordering phase) is 3000 training steps long, and the initial parameters are learning_rate=0.02 and radius=12. The second iteration (tuning phase) has 60000 training steps, and initially learning_rate=0.001 and radius=5. The learning rate parameter decreases linearly to zero during training. The radius parameter decreases linearly to one during training. A detailed explanation of the training process is given in (Kohonen 95).
2) Calibration. A set of 600 samples is used. The network response is analyzed and state labels are associated with the network units. A unit will be associated to a contact state if that unit's response is greater with input data of that state than with data of any other state. This can be easily calculated by counting how many times a neuron is selected as the closest to the input samples of the different states. The state that has more 'hits' is selected for that unit's label. A second label (of the state with the second number of hits) will also be used during visualization, in order to highlight the overlapping among states in the map.
3) Testing. The set consists of 600 samples. The performance of the network is tested with an independent set of data. For each sample, the most responsive unit is selected, i.e. the one whose weights are closer to the input signals.

The contact state is simply given by that unit's label. An uncertain response occurs when that unit is unlabeled. In order to solve this problem, we will introduce another method for calculating the network's output.

Results for a preliminary experiment are shown in Table 1. Two states, F2 and F5, are perfectly identified. Other two, F1 and F4, are almost perfectly identified. F1 is correctly identified in the 94% of the cases, it is erroneously identified as F3 in the 5% and it is unclassified in the remaining 1%. Meanwhile, F4 is properly classified in the 97% of the cases, but it is unknown in the remaining 3%. The other two states, F3 and F6, are more ambiguous, and the proper classification percentages are smaller. The average network performance is very good, a 88% success, and we must take into account that only force information has been used.

State	Right	Wrong	Unknown
F1	94	5	1
F2	100	-	-
F3	59	34	7
F4	97	-	3
F5	100	-	-
F6	78	21	1
Total %	88	10	2

Table 1. Classification percentages; $\mu = 0.2$, Clearance: 1%

This neural network can easily be visualized by using the so called u-matrix visualization (Ultsch 93) and consists in visualizing the distances between reference vectors of neighboring map units using gray levels. The farther the distance, the darker the representation. In this way, one can identify clusters, or groups of neurons with similar response, which should be desirable to belong to the same contact state. The map we obtained with the u-matrix visualization is represented in Fig. 5.

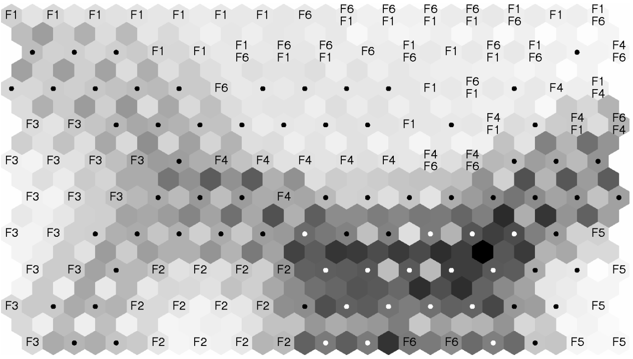

Figure 5. U-matrix representation of the neural network; task parameters: $\mu = 0.2$, Clearance = 1%

It is possible to observe a big white region on the top, with units labeled with states F1, F3 and F6, and three smaller light regions isolated by darker zones, i.e. long distances. This regions are labeled with F2, F4 and F5. This representation reflects the state ambiguities, which are also presented in the table. Some units are labeled twice to show this problem, that occurs with states F1, F3 and F6. This means that those units not only are selected for the first state, but sometimes they are also selected for another state. Unlabeled neurons are displayed as a dot.

After establishing the correspondence between force sensor data and symbols denoting contacts, the peg-in-hole insertion problem can be solved in a number of ways. Fig. 6 shows a perception-based plan in which the increments inside the nodes denote the action to be performed. Additional details can be found in (Cervera and del Pobil 2000). Several simulations were performed to show that the plan exhibits a good behavior without the need for information about the position and orientation of the peg. The approach has been further applied to solve the peg-in-whole problem with a real robot and sensor (Cervera and del Pobil 2002). In this case the plan relating symbolic contact states and motor actions was built by means of reinforcement learning.

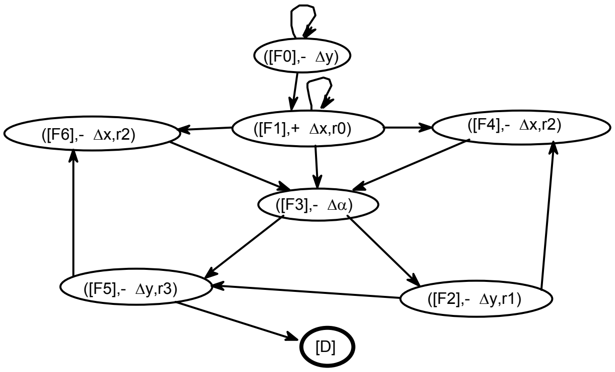

Figure 6. Perception-based motion plan

Conclusion

We have discussed the existence of symbols that are fundamental for robotic intelligence and do not refer directly to physical objects but rather to physical sensorimotor interactions between the robot itself and the objects in the world. This kind of symbols are more related to actions and they seem to have appeared in the evolutionary landscape long before vision as 'sight'. We commented on a particular implementation for robotic grasping and provided a detailed example of our approach in the context of the peg-in-hole insertion problem.

Further work in this direction may contribute to a better understanding of the symbol grounding and the anchoring problems, and to make progress towards the achievement of true robotic intelligence.

Acknowledgements

This paper describes research done in the Robotic Intelligence Laboratory. Support for this laboratory is provided in part by the CICYT under project DPI2001-3801, by the Generalitat Valenciana under projects GRUPOS03/153, CTIDIA/2002/195 and GV01-244, by Fundacio Caixa-Castello under project P1-1B2001-28, and by a scholarship of the FPI Program of the Spanish Department of Education and Science. The neural network simulations have been done with the software package SOMPAK, which was developed at the Helsinki University of Technology.

References

Cervera, E.; del Pobil, A.P. 2000. A Qualitative-Connectionist Approach to Robotic Spatial Planning. *Spatial Cognition and Computation* 2(1):51-76.

Cervera, E.; del Pobil, A.P. 2002. Sensor-Based Learning for Practical Planning of Fine Motions in Robotics. *International Journal of Information Sciences* 145(1-2):147-168.

Coradeschi, S.; Saffiotti, A. 2003. An Introduction to the Anchoring Problem". *Robotics and Autonomous Systems, Special issue on perceptual anchoring* 43(2-3):85-96.

Culham, J. 2004. Human brain imaging reveals a parietal area specialized for grasping. In *Functional Neuroimaging of Visual Cognition: Attention and Performance XX*, 417-438. Kanwisher, N.; Duncan, J. eds. Oxford University Press.

del Pobil, A.P. 1998. The Grand Challenge is Called: Robotic Intelligence. In *Tasks and Methods in Applied Artificial Intelligence*, LNCS 1416, 15-24. del Pobil, A.P.; Mira, J.; Ali, M. eds. Springer, Berlin.

Fadiga, L.; Fogassi, L.; Gallese, V.; Rizzolatti, G. 2000. Visuomotor neurons: ambiguity of the discharge or 'motor' perception? *Int. J. Psychophysiology*, 35:165-177.

Goodale, M.A.; Westwood, D.A. 2004. An evolving view of duplex vision: separate but interacting cortical pathways for perception and action. *Current Opinion in Neurobiology*, 14(2):203-211.

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

Harnad, S. 1995. Grounding Symbolic Capacity in Robotic Capacity. In *The Artificial Life route to Artificial Intelligence*, 277-286. Steels, L.; Brooks, R.A. eds. Lawrence Erlbaum.

Keysers, C.; Kohler, E.; Umiltà, M.A.; Fogassi, L.; Nanetti, L.; Gallese, V. 2003. Audio-visual mirror neurones and action recognition. *Experimental Brain Research*, 153:628-636.

Kohonen, T. 1995. *Self-Organizing Maps*. Springer Series in Information Sciences, 30. Springer, Berlin.

Milner, A.D.; Goodale, M.A. 1995. *The visual brain in action.* Oxford University Press.

Morales, A.; Chinellato, E.; Fagg, A.H.; del Pobil, A.P. 2004. Using experience for assessing grasp reliability. *Intl. J. of Humanoid Robotics*. In Press.

Rizzolatti, G.; Arbib, M.A. 1998. Language within our grasp. *Trends in Neurosciences*, 21:188-194.

Ultsch, A. 1993. Self-Organized feature maps for monitoring and knowledge acquisition of a chemical process. In *Proc. Int. Conf. Artificial Neural Nets* (ICANN93), 864-867.