Better Vision Through Experimental Manipulation

Giorgio Metta[*,**]
[*]LIRA-Lab, DIST
University of Genova
Viale F. Causa, 13
16145 Genova, Italy

Paul Fitzpatrick[**]
[**]MIT AI Lab
200 Technology Square
Cambridge, MA 02139 US

Abstract

Experimentation is crucial to human progress at all scales, from society as a whole to a young infant in its cradle. It allows us to elicit learning episodes suited to our own needs and limitations. This paper develops active strategies for a robot to acquire visual experience through simple experimental manipulation. The experiments are oriented towards determining what parts of the environment are physically coherent – that is, which parts will move together, and which are more or less independent. We argue that following causal chains of events out from the robot's body into the environment allows for a very natural developmental progression of visual competence, and relate this idea to results in neuroscience.

1. Introduction

Much of computer vision is passive in nature, with the emphasis on watching the world but not participating in it. There are advantages to moving beyond this to exploit dynamic regularities of the environment (Ballard, 1991). A robot has the potential to examine its world using causality, by performing probing actions and learning from the response. Tracing chains of causality from motor action to perception (and back again) is important both to understand how the brain deals with sensorimotor coordination and to implement those same functions in an artificial system, such as a humanoid robot. And, as a practical matter, the ability to perform "controlled experiments", such as tapping an object lightly, is crucial to getting to grips with an otherwise complex and uncertain world.

Figure 1 shows three levels of causal complexity we would like the robot to probe. The simplest causal chain that the robot experiences is the perception of its own actions. The temporal aspect is immediate: visual information is tightly synchronized to motor commands. We use this strong correlation to identify parts of the robot body – specifically, the end-point of the arm.

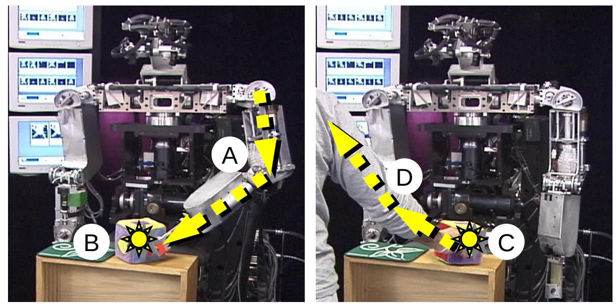


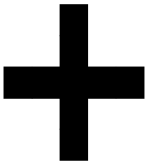
Figure 1: On the left, the robot establishes a causal connection between commanded motion and its own manipulator (A), and then probes its manipulator's effect on an object (B). The object then serves as a literal "point of contact" (C) to link robot manipulation with human manipulation (on the right, D), as is required for a mirror-neuron-like representation.

Once this causal connection is established, we can go further and use it to actively explore the boundaries of objects. In this case, there is one more step in the causal chain, and the temporal nature of the response may be delayed since initiating a reaching movement doesn't immediately elicit consequences in the environment.

In this paper, we propose that such causal probing can be arranged in a developmental sequence leading to a manipulation-driven representation of objects. We present results for two important steps along the way, and describe how we plan to proceed. We argue that following this causal chain outwards will allow us to approach the representational power of "mirror neurons" (Fadiga et al., 2000), where a connection is made between our own actions and the actions of another.

2. The elusive object

Sensory information is intrinsically ambiguous, and very distant from the world of well-defined objects in which humans believe they live. What criterion should be applied to distinguish one object from another? How can perception support such a phenomenon as figure-ground segmentation? Consider



a cross

0	0	1	0	0
0	0	1	0	0
1	1	1	1	1
0	0	1	0	0
0	0	1	0	0

a binary cross

4	12	17	4	15
9	21	3	10	25
5	23	11	37	13
8	18	7	42	6
27	46	31	32	50

?

Figure 2: Three examples of crosses, following (Manzotti and Tagliasco, 2001). The human ability to segment objects is not general-purpose, and improves with experience.

the example in Figure 2. It is immediately clear that the drawing on the left is a cross, perhaps because we already have a criterion, which allows segmenting on the basis of the intensity difference. It is slightly less clear that the zeros and ones on the middle panel are still a cross. What can we say about the array on the right? If we are not told, and we do not have the criterion to perform the figure-ground segmentation, we might think this is just a random collection of numbers. But if we are told that the criterion is "prime numbers vs. non-prime" then a cross can still be identified.

While we have to be inventive to come up with a segmentation problem that tests a human, we don't have to go far at all to find something that baffles our robots. Figure 3 shows a robot's-eye view of a cube sitting on a table. Simple enough, but many rules of thumb used in segmentation fail in this particular case. And even an experienced human observer, diagnosing the cube as a separate object based on its shadow and subtle differences in the surface texture of the cube and table, could in fact be mistaken – perhaps some malicious researcher is up to mischief. The only way to find out for sure is to take action, and start poking and prodding. As early as 1734, Berkeley observed that:

> ...objects can only be known by touch. Vision is subject to illusions, which arise from the distance-size problem... (Berkeley, 1972)

In this paper, we provide support for a more nuanced proposition: that in the presence of manipulation, vision becomes more powerful, and many of its illusions fade away.

Objects and actions

The example of the cross composed of prime numbers is a novel (albeit unlikely) type of segmentation in our experience as adult humans. We might imagine that in our infancy, we had to initially form a set of criteria to solve the object identification/segmentation problem in more mundane circumstances. We ask the question of whether we can discover these criteria during ontogenesis.

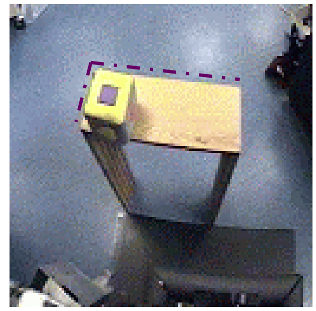


Figure 3: A cube on a table. The edges of the table and cube happen to be aligned (dashed line), the colors of the cube and table are not well separated, and the cube has a potentially confusing surface pattern.

Humans and a small number of other primates are unique in their ability to manipulate their environment using tools. Our capacities are mirrored in the brain by the size of the cortex controlling them. Neuroscience has shown that our brains possess large cortical areas devoted to the control of manipulation – not surprising, given that encephalization is believed to have evolved for the purpose of adaptively controlling action (Maturana and Varela, 1998).

A useful conceptual schema holds that visual information follows two distinct pathways in the brain, namely, the dorsal and the ventral (Ungerleider and Mishkin, 1982, Milner and Goodale, 1995). The dorsal pathway controls action directly and pragmatically; conversely, the ventral takes care of more conceptual skills such as object recognition. Of course it is important to remember, when making this dichotomy, that the two pathways are not completely segregated but rather complement each other and interact in different ways (Jeannerod, 1997).

Objects are thought to maintain a double "identity" depending on whether they are used in perceptual or in motor tasks. The concept of size, for example, might be represented multiple times in different brain areas. Observation of agnosic patients (Jeannerod, 1997) shows an even more complicated relationship than the simple dorsal/ventral dichotomy would suggest. Althought some patients could not grasp generic objects (e.g. cylinders), they could correctly preshape the hand to grasp known objects (e.g. a lipstick); interpreted in terms of the two-pathway system, this implies that the ventral representation of the object can supply the dorsal system with size information. What we consciously perceive as "size" is rather a collection of different percepts interacting in a complicated way, and under pathological circumstances they can be separated from each other. One of the "identities" of objects is thus connected to motor performance.

That such pathways develop and are not completely innate is suggested by the results of (Kovacs, 2000). She has shown that perceptual grouping is slow to develop and continues to improve well beyond early childhood (14 years). Long-range contour integration was tested and this work elucidated how this ability develops to enable extended spatial grouping. These results further suggest that the development of action might precede that of categorization: it is well enstablished that by 4 months of age infants can process complex motion stimuli, depth, and color. Roughly at the same age reaching becomes more consistent. That is, action comes first supported by the pragmatic use of diverse sensory modalities; perception conversely is a long developing process. More studies are needed though to ascertain how the dorsal pathway (action) influences the ventral (perception) both in situations like those already mentioned, and during ontogenesis.

The dorsal stream connects the parietal lobe to the premotor cortex, which project heavily onto the primary motor cortex to eventually control movements. For many years the premotor cortex was considered just another big motor area. New studies (Jeannerod, 1997) have demonstrated that this is not the case. In fact, researchers have identified neurons in the area F5 of the frontal cortex (Fadiga et al., 2000) that are activated in two situations: *i)* when acting onto an object (e.g. grasping), and *ii)* when looking at the same object (visual response). Their firing pattern was quite specific, building a link between the size of the object and the applied grasp type (e.g. a small object requires a precision grip).

These neurons were called canonical. This was quite an astonishing discovery because area F5 was believed to be only a motor area. A possible interpretation is that the brain stores a representation of objects in motor terms, and uses these representations to generate an appropriate response to objects (the concept of Gibsonian affordances translated in neural terms (Gibson, 1977)).

The gap from object manipulation to hand gesture production/recognition is small. In fact F5 contains another class of neurons called mirror neurons. A mirror neuron responds in two situations: *i)* when executing a manipulative gesture, and *ii)* when observing somebody else executing the same action. These neurons provide a link between the observation of somebody else's and our own actions. Beside the recognition of manipulative actions, they seem to support imitative behaviors. An intriguing theory proposed by Rizzolatti and Arbib (Rizzolatti and Arbib, 1998) associates mirror neurons to language.

Another important class of neurons in premotor cortex is found in area F4 (Fogassi et al., 1996). While F5 is more concerned with the distal muscles (i.e. the hand), F4 controls more proximal muscles (i.e. reaching). A subset of neurons in F4 has a somatosensory, visual and motor receptive field. The visual receptive field (RF) extends in 3D from a given body part, for example, the forearm. The somatosensory RF is usually in register with the visual one. Finally, motor information is integrated into the representation by maintaining the RF anchored to the correspondent body part (the forearm in this example) irrespective of the relative position of the head and arm.

A working hypothesis

In the light of these results, we see at least two reasons why intelligence needs to be embodied. First, if robots are to tell us something about the functioning of our brains, we have to study its development in the proper setting, that is, with the robot acting in the environment. As we have seen action is fundamental to a set of highly cognitive skills including imitation and language. Perceptual tasks are also be influenced by action. Second, to build better robots, adaptive to their environment, there is probably no alternative but to build them following to some extent biological principles. The same constraints encountered by biological agents during ontogenesis are encountered by the robot during its simulated development.

Certainly, vision and action are intertwined at a very basic level. While an experienced adult can interpret visual scenes perfectly well without acting upon them, linking action and perception seems crucial to the developmental process that leads to that competence. We can construct a working hypothesis: that action is required to object recognition in cases where an agent has to develop categorization autonomously. Of course in standard supervised learning action is not required since the trainer does the job of pre-segmenting the data by hand. In an ecological context, some other mechanism has to be provided. Ultimately this mechanism is the body itself that through action (under some suitable developmental rule) generates informative percepts.

Neurons in area F4 are thought to provide a body map useful for generating arm, head, and trunk movements. Our robot learns autonomously a crude version of this body map by fusing vision and proprioception. As a step towards establishing the kind of visuomotor representations seen in F5, we then develop a mechanism for using reaching actions to visually probe the connectivity and physical extent of objects without any prior knowledge of the appearance of the objects (or indeed of the arm itself).

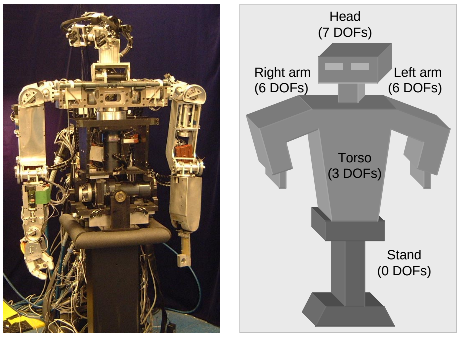


Figure 4: Degrees of freedom (DOFs) of the robot Cog. The arms terminate either in a primitive "flipper" or a four-fingered hand. The head, torso, and arms together contain 22 degrees of freedom.

3. The experimental platform

This work is implemented on the robot Cog, an upper torso humanoid (Brooks et al., 1999). Cog has two arms, each of which has six degrees of freedom. The joints are driven by series elastic actuators (Williamson, 1999). The arm is not designed to enact trajectories with high fidelity. For that a very stiff arm is preferable. Rather, it is designed to perform well when interacting with a poorly characterized environment, where collisions are frequent and informative events. Cog runs an attentional system consisting of a set of pre-attentive filters sensitive to motion, color, and binocular disparity. The different filters generate information on the likelihood that something interesting is happening in a certain region of the image. A voting mechanism is used to "decide" what to attend and track next. The pre-attentive filters are implemented on a space-variant imaging system, which mimics the distribution of photoreceptors in the human retina. The attentional system uses vision and non-visual sensors (e.g. inertial) to generate a range of oculomotor behaviors. Examples are saccades, smooth pursuit, vergence, and the vestibulo-ocular reflex (VOR).

4. Perceiving direct effects of action

Motion of the arm may generate optic flow directly through the changing projection of the arm itself, or indirectly through an object that the arm is in contact with. While the relationship between the optic flow and the physical motion is likely to be complex, the correlation in time of the two events should be exceedingly precise. This time-correlation can be used as a "signature" to identify parts of the scene that are being influenced by the robot's motion, even in the presence of other distracting motion sources. In this section, we show how this tight correlation can be used to localize the arm in the image without any prior information about visual appearance.

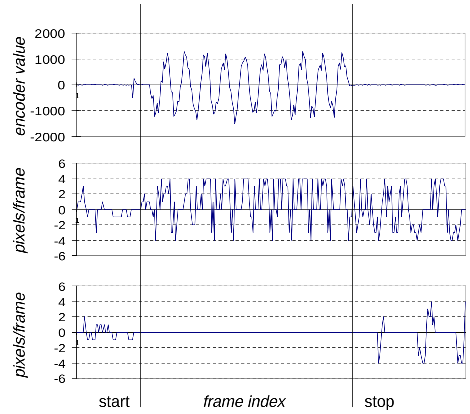


Figure 5: An example of the correlation between optic flow and arm movement. The traces show the movement of the wrist joint (upper plot) and optic flow sampled on the arm (middle plot) and away from it (lower plot). As the arm generates a repetitive movement, the oscillation is clearly visible in the middle plot and absent in the lower. Before and after the movement the head is free to saccade, generating the other spikes seen in the optic flow.

Reaching out

The first step towards manipulation is to reach objects within the workspace. If we assume targets are chosen visually, then ideally we need to also locate the end-effector visually to generate an error signal for closed-loop control. Some element of open-loop control is necessary since the end-point may not always be in the field of view (for example, when it is in its the resting position), and the overall reaching operation can be made faster with a feed-forward contribution to the control.

The simplest possible open loop control would map directly from a fixation point to the arm motor commands needed to reach that point (Metta et al., 1999) using a stereotyped trajectory, perhaps using postural primitives (Mussa-Ivaldi and Giszter, 1992). If we can fixate the end-effector, then it is possible to learn this map by exploring different combinations of direction of gaze vs. arm position (Marjanović et al., 1996, Metta et al., 1999). So locating the end-effector visually is key both to closed-loop control, and to training up a feed-forward path. We shall demonstrate that this localization can be performed without knowledge of the arm's appearance, and without assuming that the arm is the only moving object in the scene.

Localizing the arm visually

The robot is not a passive observer of its arm, but rather the initiator of its movement. This

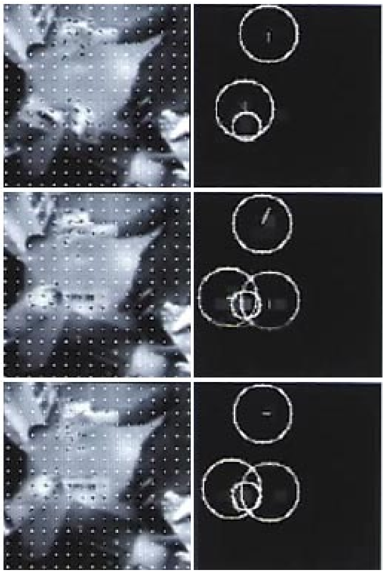


Figure 6: Detecting the arm/gripper through motion correlation. The robot's point of view and the optic flow generated are shown on the left. On the right are the results of correlation. Large circles represent the results of applying a region growing procedure to the optic flow. Here the flow corresponds to the robot's arm and the experimenter's hand in the background. The small circle marks the point of maximum correlation, identifying the regions that correspond to the robot's own arm.

can be used to distinguish the arm from parts of the environment that are more weakly affected by the robot. The arm of a robot was detected in (Marjanović et al., 1996) by simply waving it and assuming it was the only moving object in the scene. We take a similar approach here, but use a more stringent test of looking for optic flow that is correlated with the motor commands to the arm. This allows unrelated movement to be ignored. Even if a capricious engineer were to replace the robot's arm with one of a very different appearance, and then stand around waving the old arm, this detection method will not be fooled.

The actual relationship between arm movements and the optic flow they generate is complex. Since the robot is in control of the arm, it can choose to move it in a way that bypasses this complexity. In particular, if the arm rapidly reverses direction, the optic flow at that instant will change in sign, giving a tight, clean temporal correlation. Since our optic flow processing is coarse (a 16×16 grid over a 128×128 image at 15 Hz), we simply repeat this reversal a number of times to get a strong correlation signal during training. With each reversal the probability of correlating with unrelated motion in the environment goes down.

Figure 5 shows an example of this procedure in

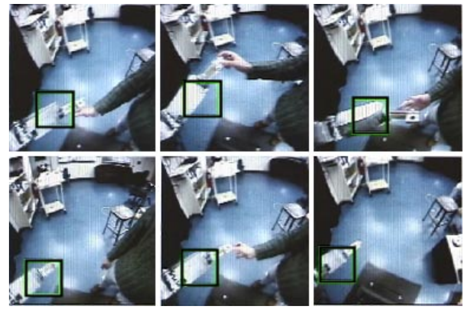


Figure 7: Predicting the location of the arm in the image as the head and arm change position. The rectangle represents the predicted position of the arm using the map learned during a twenty-minute training run. The predicted position just needs to be sufficiently accurate to initialize a visual search for the exact position of the end-effector.

operation, comparing the velocity of the arm's wrist with the optic flow at two positions in the image plane. A trace taken from a position away from the arm shows no correlation, while conversely the flow at a position on the wrist is strongly different from zero over the same period of time. Figure 6 shows examples of detection of the arm and rejection of a distractor.

Localizing the arm using proprioception

The localization method for the arm described so far relies on a relatively long "signature" movement that would slow down reaching. This can be overcome by training up a function to estimate the location of the arm in the image plane from proprioceptive information (joint angles) during an exploratory phase, and using that to constrain arm localization during actual operation. Figure 7 shows the resulting behavior after about twenty minutes of real-time learning.

5. Perceiving indirect effects of action

We have assumed that the target of a reaching operation is chosen visually. As discussed in the introduction, visual segmentation is not easy, so we should not expect a target selected in this way to be a correctly segmented. For the example scene in Figure 3 (a cube sitting on a table), the small inner square on the cube's surface pattern might be selected as a target. The robot can certainly reach towards this target, but grasping it would prove difficult without a correct estimate of the object's physical extent. In this section, we develop a procedure for refining the segmentation using the same idea of correlated motion used earlier to detect the arm.

When the arm enters into contact with an object, one of several outcomes are possible. If the object

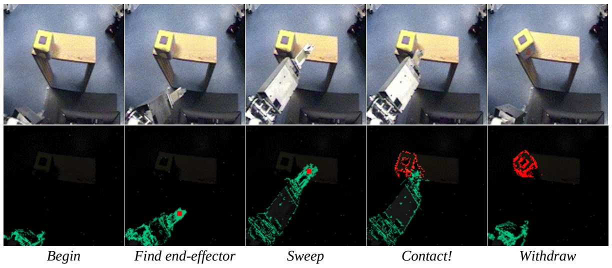


Figure 8: The upper sequence shows an arm extending into a workspace, tapping an object, and retracting. This is an exploratory mechanism for finding the boundaries of objects, and essentially requires the arm to collide with objects under normal operation, rather than as an occasional accident. The lower sequence shows the shape identified from the tap using simple image differencing and flipper tracking.

is large, heavy, or otherwise unyielding, motion of the arm may simply be resisted without any visible effect. Such objects can simply be ignored, since the robot will not be able to manipulate them. But if the object is smaller, it is likely to move a little in response to the nudge of the arm. This movement will be temporally correlated with the time of impact, and will be connected spatially to the end-effector – constraints that are not available in passive scenarios (Birchfield, 1999). If the object is reasonably rigid, and the movement has some component in parallel to the image plane, the result is likely to be a flow field whose extent coincides with the physical boundaries of the object.

Figure 8 shows how a "poking" movement can be used to refine a target. During a poke operation, the arm begins by extending outwards from the resting position. The end-effector (or "flipper") is localized as the arm sweeps rapidly outwards, using the heuristic that it lies at the highest point of the region of optic flow swept out by the arm in the image (the head orientation and reaching trajectory are controlled so that this is true). The arm is driven outward into the neighborhood of the target which we wish to define, stopping if an unexpected obstruction is reached. If no obstruction is met, the flipper makes a gentle sweep of the area around the target. This minimizes the opportunity for the motion of the arm itself to cause confusion; the motion of the flipper is bounded around the endpoint whose location we know from tracking during the extension phase, and can be subtracted easily. Flow not connected to the end-effector can be ignored as a distractor. Figure 9 shows more detailed results, including examples of the actual segmented region assigned to the object. In the absence of strong texture there may be little motion signature in the interior of the object, so we recruit a maximum-flow algorithm due

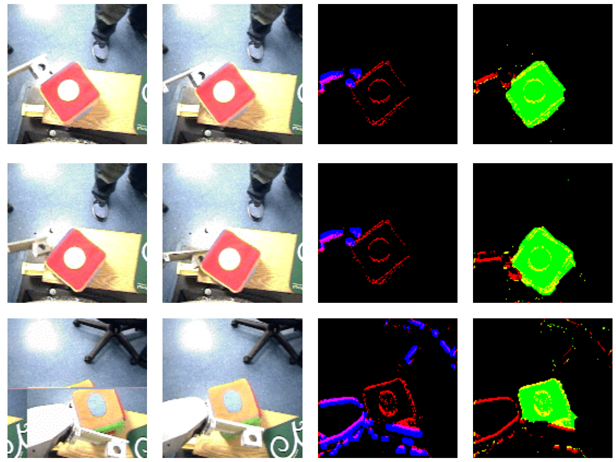


Figure 9: Cog batting a cube around. The top two rows show the flipper poking the object repeatedly from the side, turning it slightly. The third row shows Cog batting an object away. The images in the first column are frames prior to a collision. The second column shows the actual impact. The third column shows the motion signal at the point of contact. The bright regions in the images in the final column show the segmentations produced for the object.

to (Boykov and Kolmogorov, 2001) to fill in such regions efficiently.

The poking operation gives clear results for a rigid object that is free to move. What happens for non-rigid objects and objects that are attached to other objects? Here the results of poking are likely to be more complicated to interpret – but in a sense this is a good sign, since it is in just such cases that the idea of an object becomes less well-defined. Poking has the potential to offer an operational theory of "objecthood" that is more tractable than a vision-only approach might give, and which cleaves better to the true nature of physical assemblages. The idea of a physical object is rarely completely coherent, since it depends on where you draw its boundary and that may well be task-dependent. Poking allows us to determine the boundary around a mass that moves together when disturbed, which is exactly what we need to know for manipulation. As an operational definition of object, this has the attractive property of breaking down into ambiguity in the right circumstances – such as for large interconnected messes, floppy formless ones, liquids, and so on.

6. Developing mirror neurons?

Poking moves us one step outwards on a causal chain away from the robot and into the world, and gives a simple experimental procedure for segmenting objects. There are many possible elaborations of this method (some are mentioned in the conclusions), all

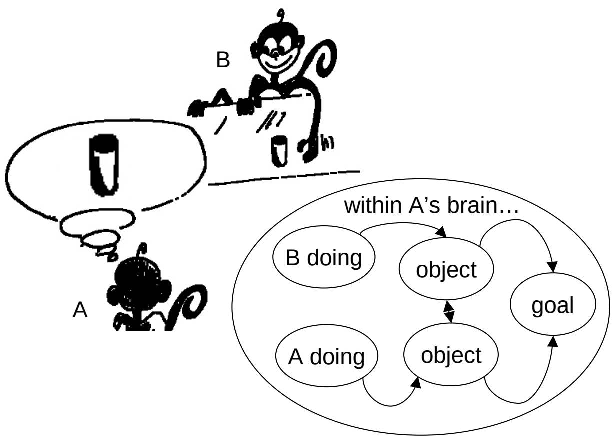


Figure 10: Mirror neurons and causality: from the observer's point of view (A), understanding B's action means mapping it onto the observer's own motor repertoire. If the causal chain leading to the goal is already in place (lower branch of the graph) then the acquisition of a mirror neuron for this particular action/object is a matter of building and linking the upper part of the chain to the lower one. There are various opportunities to reinforce this link either at the object level, at the goal level or both.

of which lead to a vision system that is tuned to acquiring data about an object by seeing it manipulated by the robot. An interesting question then is whether the system could extract useful information from seeing an object manipulated by someone else. In the case of poking, the robot needs to be able to estimate the moment of contact and to track the arm sufficiently well to distinguish it from the object being poked. We are interested in how the robot might learn to do this. One approach is to chain outwards from an object the robot has poked. If someone else moves the object, we can reverse the logic used in poking – where the motion of the manipulator identified the object – and identify a foreign manipulator through its effect on the object. Poking is an ideal testbed for future work on this, since it is much simpler than full-blown object manipulation and would only require a very simple model of the foreign manipulator to work.

There is considerable precedent in the biological literature for a strong connection between viewing object manipulation performed by either oneself or another (Wohlscläger and Bekkering, 2002). Also the role of object in the understanding of action performed by others has been investigated (Woodward, 1998). In a series of experiments Woodward and colleagues elucidated the contribution that seeing an object makes for 5, 6, and 9 month old infants. They provided evidence that the object and the goal-directness of the action represent an important component in the understanding of the intentions of others.

At the neural level, we already mentioned the presence of neurons in F5 that have a very specific response when an object is either fixated or manipulated (canonical neurons). Grossly simplifying, we might think of canonical neurons as an association table of grasp/manipulation (action) types with object (vision) types.

F5 also contains mirror neurons. These neurons, as we described before, respond when either watching somebody else performing a manipulative action or when actually manipulating an object. They can be thought of as a second-level association map which links together the observation of a manipulative action performed by somebody else with the neural representation of one's own action.

The question of whether a mirror-like representation can be autonomously developed by the robot (or a human for that matter) can then be answered. The association map can be constructed by identifying when the goal and the object are the same irrespective of who is the actor. Actions that lead to the same consequences are thus part of the same equivalence class. This is exactly what mirror neurons represent.

Figure 10 shows this causal chain in action. There are a series of interesting behaviors that can be realized based on mirror neurons. Mimicry is an obvious application, since it requires just this type of mapping between other and self in terms of motor actions. Another important application is the prediction of future behavior from current actions, or even inverting the causal relation to find the action that most likely will get to the desired consequence.

7. Discussion and Conclusions

In this paper, we showed how causality can be probed at different levels by the robot. Initially the environment was the body of the robot itself, then later a carefully circumscribed interaction with the outside world. This is reminiscent of Piaget's distinction between primary and secondary circular reactions (Ginsburg and Opper, 1978). Objects are central to interacting with the ouside world. We raised the issue of how an agent can autonomously acquire a working definition of objects.

The number of papers written on techniques for visual segmentation is vast. Methods for characterizing the shape of an object through tactile information are also being developed, such as shape from probing (Paulos, 1999) or pushing (Moll and Erdmann, 2001). But while it has long been known that motor strategies can aid vision (Ballard, 1991), work on active vision has focused almost exclusively on moving cameras. There is much to be gained by bringing a manipulator into the equation, as we have shown in this paper.

Acknowledgements

Many people have contributed to developing the Cog platform (Brooks et al., 1999). This work benefited from discussions with Charles Kemp and Giulio Sandini. Funds for this project were provided by DARPA (contract number DABT 63-00-C-10102), and by the Nippon Telegraph and Telephone Corporation as part of the NTT/MIT Collaboration Agreement.

References

Ballard, D. H. (1991). Animate vision. *Artificial Intelligence*, 48(1):57–86.

Berkeley, G. (1972). *A new theory of vision and other writings*. Dent, London. First published in 1734.

Birchfield, S. (1999). *Depth and Motion Discontinuities*. PhD thesis, Dept. of Electrical Engineering, Stanford University.

Boykov, Y. and Kolmogorov, V. (2001). An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 359–374.

Brooks, R. A., Breazeal, C., Marjanovic, M., and Scassellati, B. (1999). The Cog project: Building a humanoid robot. *Lecture Notes in Computer Science*, 1562:52–87.

Fadiga, L., Fogassi, L., Gallese, V., and Rizzolatti, G. (2000). Visuomotor neurons: ambiguity of the discharge of 'motor' perception? *International Journal of Psychophysiology*, 35:165–177.

Fogassi, L., Gallese, V., Fadiga, L., Luppino, G., Matelli, M., and Rizzolatti, G. (1996). Coding of peripersonal space in inferior premotor cortex (area F4). *Journal of Neurophysiology*, pages 141–157.

Gibson, J. J. (1977). The theory of affordances. In Shaw, R. and Bransford, J., (Eds.), *Perceiving, acting and knowing: toward an ecological psychology*, pages 67–82. Hillsdale NJ: Lawrence Erlbaum Associates Publishers.

Ginsburg, H. and Opper, S. (1978). *Piaget's theory of intellectual development*. Prentice-Hall, Englewood Cliffs, NJ. 2nd edition.

Jeannerod, M. (1997). *The Cognitive Neuroscience of Action*. Blackwell Publishers Inc., Cambridge Massachusetts and Oxford UK.

Kovacs, I. (2000). Human development of perceptual organization. *Vision Research*, 40(10-12):1301–1310.

Manzotti, R. and Tagliasco, V. (2001). *Coscienza e realtà: una teoria della coscienza per costruttori di menti e cervelli*. il Mulino.

Marjanović, M. J., Scassellati, B., and Williamson, M. M. (1996). Self-taught visually-guided pointing for a humanoid robot. In *From Animals to Animats: Proceedings of 1996 Society of Adaptive Behavior*, pages 35–44, Cape Cod, Massachusetts.

Maturana, R. and Varela, F. (1998). *The tree of knowledge, the biological roots of human understanding*. Boston & London. Shambhala Publications, revised edition.

Metta, G., Sandini, G., and Konczak, J. (1999). A developmental approach to visually-guided reaching in artificial systems. *Neural Networks*, 12:1413–1427.

Milner, A. D. and Goodale, M. A. (1995). *The visual brain in action*, volume 27. Oxford University Press.

Moll, M. and Erdmann, M. A. (2001). Reconstructing shape from motion using tactile sensors. In *Proc. 2001 IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems*, Maui, HI.

Mussa-Ivaldi, F. A. and Giszter, S. F. (1992). Vector field approximation: a computational paradigm for motor control and learning. *Biological Cybernetics*, 67:491–500.

Paulos, E. (1999). Fast construction of near optimal probing strategies. Master's thesis, University of California, Berkeley.

Rizzolatti, G. and Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, 21:188–194.

Ungerleider, L. G. and Mishkin, M. (1982). Two cortical visual systems. In *Analysis of visual behavior*, pages 549–586. MIT Press, Cambridge, Massachusetts.

Williamson, M. (1999). *Robot Arm Control Exploiting Natural Dynamics*. PhD thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

Wohlschläger, A. and Bekkering, H. (2002). Is human imitation based on a mirror-neurone system? Some behavioural evidence. *Experimental Brain Research*, 143:335–341.

Woodward, A. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, 69:1–34.