

The Whole World in Your Hand: Active and Interactive Segmentation

Artur Arsenio* Paul Fitzpatrick*
*MIT AI Lab
Cambridge, Massachusetts, USA

Charles C. Kemp* Giorgio Metta**¹
**Lira Lab, DIST, University of Genova
Genova, Italy

Abstract

Object segmentation is a fundamental problem in computer vision and a powerful resource for development. This paper presents three embodied approaches to the visual segmentation of objects. Each approach to segmentation is aided by the presence of a hand or arm in the proximity of the object to be segmented. The first approach is suitable for a robotic system, where the robot can use its arm to evoke object motion. The second method operates on a wearable system, viewing the world from a human’s perspective, with instrumentation to help detect and segment objects that are held in the wearer’s hand. The third method operates when observing a human teacher, locating periodic motion (finger/arm/object waving or tapping) and using it as a seed for segmentation. We show that object segmentation can serve as a key resource for development by demonstrating methods that exploit high-quality object segmentations to develop both low-level vision capabilities (specialized feature detectors) and high-level vision capabilities (object recognition and localization).

1. Introduction

Both the machine vision community and cognitive science researchers recognize objects as a powerful abstraction for intelligent systems. Likewise, those who study cognitive development have a long history of analyzing the detailed maturation of object related competencies in infants and children. But despite the acknowledged importance of objects to human cognition and visual perception, our robots continue to be challenged by the everyday objects that surround them. Fundamentally, robots must be able to perceive objects in order to learn about them, manipulate them, and develop the important set of intellectual capabilities that rely on them. In this paper, we demonstrate three embodied methods that allow

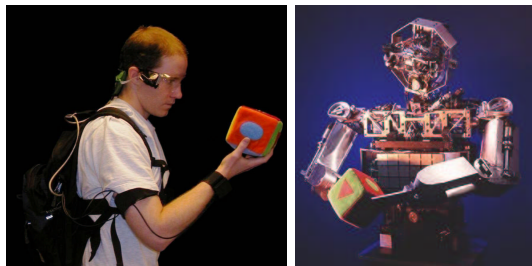


Figure 1: The platforms.

machines to visually perceive the extent of manipulable objects. Furthermore, we show that the object segmentations that result from these methods can serve as a powerful foundation for the development of more general object perception.

The presence of a body changes the nature of perception. The body provides constraint on interpretation, opportunities for experimentation, and a medium for communication. Hands in particular are very revealing, since they interact directly and flexibly with objects. In this paper, we demonstrate several methods for simplifying visual processing by being attentive to hands, either of humans or robots. This is an important cue also in primates, as was shown by Perret and colleagues (Perrett et al., 1990), who located areas in the brain specific to the processing of the visual appearance of the hand (one’s own or observed). Our first argument is that in a wide range of situations, there are many cues available that can be used to make object segmentation an easy task. This is important because object segmentation or figure/ground separation is a long-standing problem in computer vision, and has proven difficult to achieve reliably on passive systems. The segmentation methods we present are particularly well suited to segmenting manipulable objects, which by definition are potentially useful components of the world and therefore worthy of special attention. We look at three situations in which active or interactive cues simplify segmentation:

¹authors ordered alphabetically

(i) Active segmentation for a robot viewing its own actions. A robot arm probes an area, seeking to trigger object motion so that the robot can identify the boundaries of the object through that motion.

(ii) Active segmentation for a wearable system viewing its wearer’s actions. The system monitors human action, issues requests, and uses active sensing to detect grasped objects held up to view.

(iii) Demonstration-based segmentation for a robot viewing a human’s actions. Segmentation is achieved by detecting and interpreting natural human showing behavior such as finger tapping, arm waving, or object shaking.

Our second argument is that visual object segmentation can serve as a powerful foundation for the development of useful object related competencies in epigenetic systems. We support this by demonstrating that when segmentation is available, several other important vision problems can be dealt with successfully – object recognition, object localization, edge detection, etc.

2. Object perception

What our retinas register when we look at the world and what we actually believe we *see* are notoriously different (Johnson, 2002). How does the brain make the leap from sensing photons to perceiving objects? The development of object perception in human infants is an active and important area of research (Johnson, 2003). A central question is that of segmentation or ‘object unity’ – how a particular collection of surface fragments become bound into a single object representation. In our work we focus on identifying or engineering special situations when object unity is simple to achieve, and show how to exploit such situations as opportunities for development, so that object unity judgements can be made in novel situations. There is evidence that a similar process occurs in infants. Spelke and others have shown that the coherent motion of an object is a cue that young infants can use to unite surface fragments into a single object (Jusczyk et al., 1999). Needham gives evidence that even a brief exposure to independent motion of two objects can influence an infant’s perception of object boundaries in later presentations (Needham, 2001). The ability to achieve object unity does not appear fully-formed in the neonate, but develops over time (Johnson, 2002). In this paper, we explore analogues of this developmental step, and demonstrate that the ability to perceive the boundaries of objects in special, constrained situations can in

fact be automatically generalized to other situations. Elsewhere, we have used this ability as the basis for learning about and exploiting an object affordance (Metta and Fitzpatrick, 2003), and to learn about activities by tracking actions taken on familiar objects (Fitzpatrick, 2003).

Switching our attention from theoretical to practical considerations, decades of experience in computer vision have shown that object segmentation on unstructured, non-static, noisy and low resolution images is a hard problem. The techniques this paper describes for object segmentation deal with different combinations of the following situations, many of which are classically challenging:

- ▷ Segmentation of an object with colors or textures that are similar to the background.
- ▷ Segmentation of an object among multiple moving objects in a scene.
- ▷ Segmentation of fixed or heavy objects in a scene, such as a table or a sofa.
- ▷ Segmentation of objects printed or drawn in a book or in a frame, which cannot be moved relative to other objects on the same page.
- ▷ Insensitivity to luminosity variations.
- ▷ Fast operation (near real-time).
- ▷ Low resolution images.

The next three sections document three basic active and interactive approaches to segmentation, and then the remainder of the paper shows how to use object segmentation to develop object localization, recognition, and other perceptual abilities.

3. Segmentation on a robot

The idea of using action to aid perception is the basis of the field of “active perception” in robotics and computer vision (Ballard, 1991, Sandini et al., 1993). The most well-known instance of active perception is active vision. The term “active vision” has become essentially synonymous with moving cameras, but it need not be. Work on the robot Cog (pictured in Figure 1) has explored the idea of manipulation-aided vision, based on the observation that robots have the opportunity to examine the world using causality, by performing probing actions and learning from the response. In conjunction with a developmental framework, this could allow the robot’s experience to expand outward from its sensors into its environment, from its own arm to the objects it encounters, and from those objects outwards to other actors that encounter those same objects.

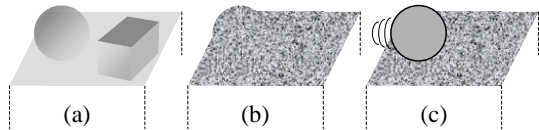


Figure 2: Cartoon motivation for active segmentation. Human vision is excellent at figure/ground separation (top left), but machine vision is not (center). Coherent motion is a powerful cue (right) and the robot can invoke it by simply reaching out and poking around.

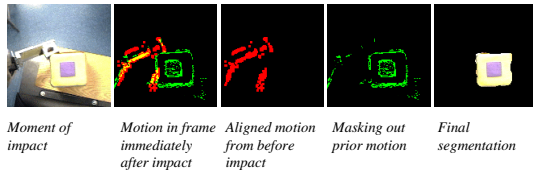


Figure 3: This images show the processing steps involved in poking. The moment of impact between the robot arm and an object, if it occurs, is easily detected – and then the total motion after contact, when compared to the motion before contact and grouped using a minimum cut approach (Boykov and Kolmogorov, 2001) gives a very good indication of the object boundary.

Object segmentation is a first step in this progression. To enable it, Cog was given a simple “poking” behavior, whereby it selects locations in its environment, and sweeps through them with its arm (Metta and Fitzpatrick, 2003). If an object is within the area swept, then the motion generated by the impact of the arm with that object greatly simplifies segmenting that object from its background, and obtaining a reasonable estimate of its boundary (see Figure 3). The image processing involved relies only on the ability to fixate the robot’s gaze in the direction of its arm. This coordination can be achieved either as a hard-wired primitive or through learning. Within this context, it is possible to collect good views of the objects the robot pokes, and the robot’s own arm.

This choice of activity has many benefits. (i) The motion generated by the impact of the arm with a rigid object greatly simplifies segmenting that object from its background, and obtaining a reasonable estimate of its boundary (see Figure 3). (ii) The poking activity also leads to object-specific consequences, since different objects respond to poking in different ways. For example, a toy car will tend to roll forward, while a bottle will roll along its side. (iii) The basic operation involved, striking objects, can be performed by either the robot or its human companion, creating a controlled point of comparison between robot and human action.

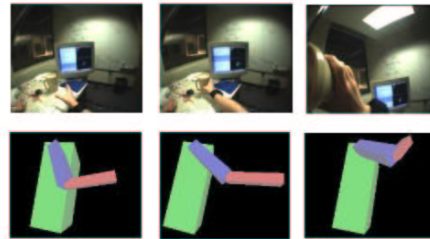


Figure 4: The wearable system monitors the wearer’s point of view (top row) while simultaneously tracking the wearer’s arm (bottom row).



Figure 5: The wearable system currently achieves segmentation by active sensing. When the wearer brings an object up into view (first column), an oscillating light source is activated (second column). The difference between images (third column) is used to compute a mask (fourth column) and segment out the grasped object and the hand from the background via a simple threshold.(fifth column).

4. Segmentation on a wearable

Wearable computing systems have the potential to measure most of the sensory input and physical output of a person as he or she goes through everyday activities. A wearable system that controls a human’s actions while making these measurements could take advantage of the wearer’s embodiment and expertise in order to develop more sophisticated perceptual processing.

One of the authors is designing a system named Duo that consists of a wearable creature and a cooperative human (Kemp, 2002). The wearable component of Duo serves as a high-level controller that requests actions from the human through speech, while the human serves as an innate and highly sophisticated infrastructure for Duo. From a developmental perspective the human is analogous to a very sophisticated set of innate abilities that Duo can use to bootstrap development. In order for Duo to take full advantage of these abilities, Duo must learn to better interpret human actions and their consequences, and learn to appropriately request human actions.

The wearable side of Duo currently consists of a head-mounted camera, 4 absolute orientation sensors, an LED array, and headphones. The wide angle lens and position of the head-mounted cam-

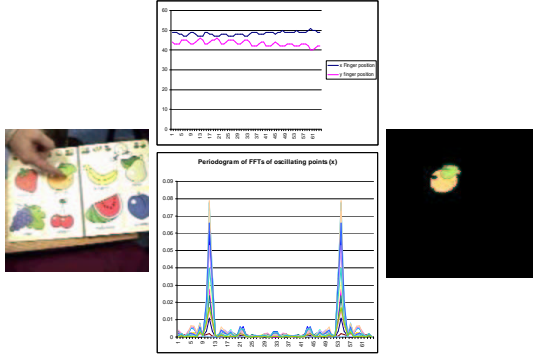


Figure 6: Segmentation based on finger tapping (left). This periodic motion can be detected through a windowed FFT on the trajectory of points tracked using optic flow, and the points implicated in the motion used to seed a color segmentation. The segmentation is applied to a frame with the hand absent, grabbed when there is no motion.

era help Duo to view the workspace of the dominant arm. The 4 absolute orientation sensors are affixed to the lower arm, upper arm, torso and head of the human, so that Duo may estimate the kinematic configuration of the person’s head and dominant arm. The wearable system makes spoken requests through the headphones and uses the LED array to aid vision (see Figure 4).

Currently, when Duo detects that the arm has reached for an object and picked the object up, Duo asks to see the object better. When a cooperative person brings the object close to his head for inspection, Duo recognizes the proximity of the object using the arm kinematics, and turns on a flashing array of white LEDs. The illumination clearly differentiates between foreground and background since illumination rapidly declines as a function of depth. By simply subtracting the illuminated and non-illuminated images from one another and applying a constant threshold, Duo is able to segment the object of interest and the hand (see Figure 5). While the human is holding the object close to the head, Duo kinematically monitors head motion and requests that the person keep his head still if the motion goes above a threshold. Minimizing head motion improves the success of the simple segmentation algorithm and reduces the need for motion compensation prior to subtracting the images.

5. Segmentation by demonstration

The two segmentation scenarios described so far operate on first-person perspectives of the world – the robot watching its own motion, or a wearable watching its wearer’s motion. Now we de-

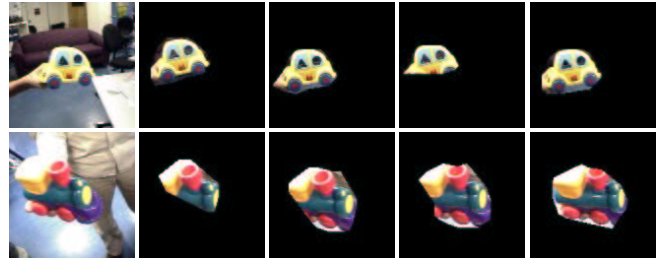


Figure 7: Periodic motion can also be used to segment an object held by the teacher, if they shake it.

velop a method that is suitable for segmenting objects based on external cues. We assume the presence of a cooperative human or “teacher” who is willing to present objects according to a protocol based on periodic motion – waving the object, tapping it with one’s finger, etc. (Arsenio, 2002).

5.1 Periodicity detection

For events created by human teachers, such as tapping an object or waving their hand in front of the robot, the periodic motion can be used to help segment it. Such events are detected through two measurements: a motion mask derived by comparing successive images from the camera and placing a non-convex polygon around any motion found, and a skin-tone mask derived by a simple skin color detector. A grid of points are initialized and tracked in the moving region. Tracking is implemented through the computation of the optical flow using the Lucas-Kanade pyramidal algorithm. Their trajectory is evaluated using a windowed FFT (WFFT), with the window size on the order of 2 seconds. If a strong periodicity is found, the points implicated are used as seeds for color segmentation. Otherwise the window size is halved and the procedure is tried again for each half. A periodogram is determined for all signals from the energy of the WFFTs over the spectrum of frequencies. These periodograms are then processed to determine whether they are usable for segmentation. A periodogram is *rejected* if one of the following four conditions holds: i) there is more than one energy peak above 50% of the maximum peak; ii) there are more than three energy peaks above 10% of the maximum peak value; iii) the DC component corresponds to the maximum energy; iv) peaks in the signal spectrum are diffuse rather than sharp. This is equivalent to passing the signals through a collection of band-pass filters. Once we can detect periodic motion and isolate it spatially, we can use identify waving actions and use them to guide segmentation.

5.2 *Waving the hand/arm/finger*

This method has the potential to segment objects that cannot be moved independently, such as objects painted in a book (see Figure 6), or heavy, stationary objects such as a table or sofa. Events of this nature are detected whenever the majority of the periodic signals arise from points whose color is consistent with skin-tone. The algorithm assumes that skin-tone points moving periodically are probably projected points from the arm, hand and/or fingers. An affine flow-model is applied to the optical flow data at each frame, and used to determine the arm/hand/finger trajectory over the temporal sequence. Points from these trajectories are collected together, and mapped onto a reference image taken before the waving began (this image is continuously updated until motion is detected). A standard color segmentation (Comaniciu and Meer, 1997) algorithm is applied to this reference image, and points taken from waving are used to select and group a set of segmented regions into what is probably the full object. This is done by merging the regions of the color segmented image whose pixel values are close to the seed pixel values, and which are connected with the seed pixels.

5.3 *Waving the object*

Multiple moving objects create ambiguous segmentations from motion, while difficult figure/ground separation makes segmentation harder. The strategy described in this section filters out undesirable moving objects, while providing the full object segmentation from motion. Whenever a teacher waves an object in front of the robot, or sets an oscillating object in motion, the periodic motion of the object is used to segment it (see Figure 7). This technique is triggered whenever the majority of periodic points are generic in appearance, rather than drawn from the hand or finger. The set of periodic points tracked over time are sparse, and hence an algorithm is required to group them into a meaningful template of the object of interest. An affine flow model is estimated by a least squares minimization criterion from the optical flow data. The estimated model plus covariance matrices are used to recruit other points within the Mahalanobis distance. Finally, a non-convex approximation algorithm is applied to all periodic, non skin-colored points to segment the object. Note that this approach is robust to humans or other objects moving in the background – they are ignored as long as their motion is non-periodic.

6. Building on segmentation

We see object segmentation as the first step on a developmental trajectory towards a robust, well-adapted vision system. It is a key opportunity for many kinds of visual learning:

Learning about low-level features: The segmented views of objects can be pooled to train detectors for basic visual features – for example, edge orientation. Once an object boundary is known, the appearance of the edge between the object and the background can be sampled, and each sample labeled with the orientation of the boundary in its neighborhood.

Learning to recognize objects: High-quality segmented views of objects can serve as extremely useful training data for object detection and recognition systems, since they unambiguously label the visual features that are associated with an object. Often these visual features can be used to detect, track, and recognize the object in new contexts where the segmentation methods presented here are not applicable.

Learning about object behavior: Once objects can be located and segmented from the background, they can be tracked to learn about their dynamic properties (Metta and Fitzpatrick, 2003).

6.1 *Learning about low-level features*

Object segmentation identifies the boundaries around an object. By examining the appearance of this boundary over many objects, it is possible to build up a model of the appearance of edges. This is an empirically grounded alternative to the many analytic approaches such as (Freeman and Adelson, 1991). Figure 8 shows examples of the kind of edge samples gathered using active segmentation on the robot Cog. The results show that the most frequent edge appearances are “ideal” straight, noise-free edges, as might be expected. Line-like edges also occur, although with lower probability, along with a diversity of other more complicated edges (zig-zags, dashed edges, and so on). Although these samples are collected for object boundaries, they can be used to estimate orientation throughout an image, giving a general-purpose orientation detector that works in situations outside the one for which it is explicitly trained (Fitzpatrick, 2003).

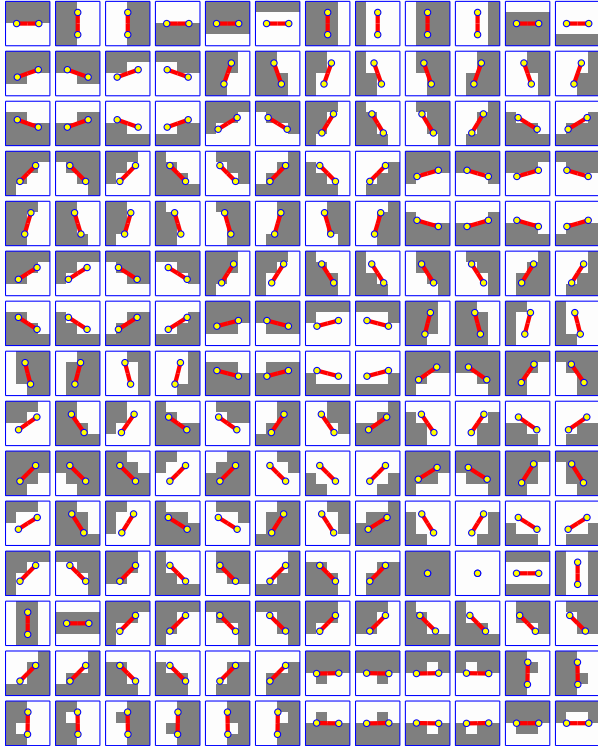


Figure 8: The empirical appearance of edges. Each 4×4 grid represents the possible appearance of an edge, quantized to just two luminance levels. The line centered in the grid is the average orientation that patch was observed on object boundaries during segmentation. Shown are the most frequent appearances observed in about 500 object segmentations.

6.2 Learning to recognize objects

With any of the active segmentation behaviors introduced here, the system can familiarize itself with the appearance of nearby objects in a special, constrained situation. It is then possible to learn to locate and recognize those objects whenever they are present, even when the special cues used for active segmentation are not available. The segmented views can be grouped by their appearance and used to train up an object recognition module, which can then find them against background clutter (see Figure 9).

Object recognition is performed using geometric hashing (Wolfson and Rigoutsos, 1997), based on pairs of oriented regions found using the detector developed in Section 6.1. The orientation filter is applied to images, and a simple region growing algorithm divides the image into sets of contiguous pixels with coherent orientation. For realtime operation, adaptive thresholding on the minimum size of such regions is applied, so that the number of regions is bounded, independent of scene com-

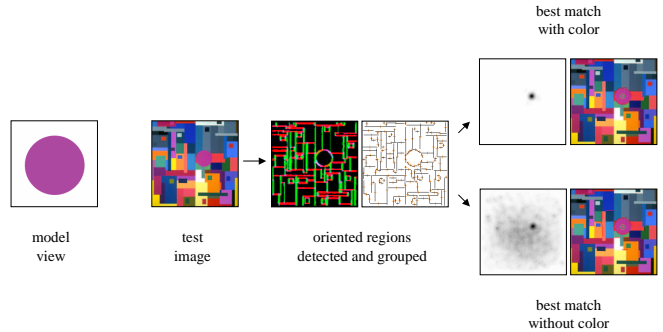


Figure 9: A simple example of object localization: finding a circle buried inside a Mondrian. Given a model view (left) of the desired object free from any background clutter, a cluttered view of the object (second from left) can be searched for the specific feature combinations seen in the model (center), and the target identified amidst the clutter (right). The features we used combined geometric and color information across pairs of oriented regions (Fitzpatrick, 2003).

plexity. In “model” (training) views, every pair of regions belonging to the object is considered exhaustively, and entered into a hash table, indexed by relative angle, relative position, and the color at sample points between the regions (if inside the object boundary). When searching for the object, every pair of regions in the current view is compared with the hash table and matches are accumulated as evidence for the presence of the object. As a simple example of how this all works, consider the test case shown in Figure 9. The system is presented with a model view of the circle, and the test image. For simplicity, the model view in this case is a centered view of the object by itself, so no segmentation is required. The processing on the model and test image is the same; first the orientation filter is applied, and then regions of coherent orientation are detected. For the circle, these regions will be small fragments around its perimeter. For the straight edges in the test image, these regions will be long. So finding the circle reduces to locating a region where there are edge fragments at diverse angles to each other, and with the distance between them generally large with respect to their own size. Even without using color, this is quite sufficient for a good localization in this case. The perimeter of the circle can be estimated by looking at the edges that contribute to the peak in match strength. The algorithm works equally well on an image of many circles with one square, and has been applied to many kinds of objects (letters, compound geometric shapes, natural objects such as a bottle or toy car).

The matching process also allows the boundary

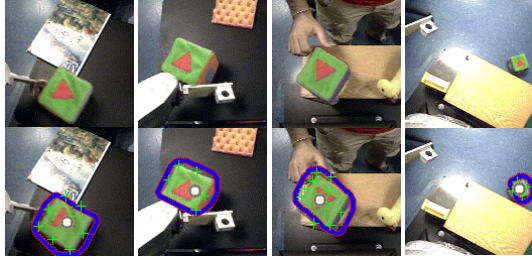


Figure 10: A cube being recognized, localized, and segmented in real images. The image in the first column is one taken when the robot Cog was poking an object, and was used (along with others) to train the recognition system. The image in the remain columns are test images. The border superimposed on the images in the bottom row represents the border of the object produced automatically. Note the scale and orientation invariance demonstrated in the final image.

of the object in the image to be recovered. Figure 10 shows examples of an object (a cube) being located and segmented automatically, without using any of the special segmentation contexts discussed in this paper, except for initial training. Testing on a set of 400 images of four objects poked by the robot, with half the images used for training, and half for testing, gives a recognition error rate of 2%, with a median localization error of 4.2 pixels in a 128×128 image (as determined by comparing against the center of the segmented region given by active segmentation). By segmenting the image by grouping the regions implicated in locating the object, and filling in, a median of 83.5% of the object is recovered, and 14.5% of the background is mistakenly included (again, determined by comparison with the results of active segmentation).

In geometric hashing, the procedure applied to an image at recognition time is essentially identical to the procedure applied at training time. We can make use of that fact to integrate training into a fully online system, allowing behavior such as that shown in Figure 11, where a previously unknown object can be segmented through active segmentation and then immediately localized and recognized in future interaction.

6.3 Learning about object behavior

Once individual objects can be recognized, properties that are more subtle than physical appearance can be learned and associated with that object. For a robot, the affordances offered by an object are important to know (Gibson, 1977). In previous work, Cog was given the ability to characterize the tendency of an object to roll when

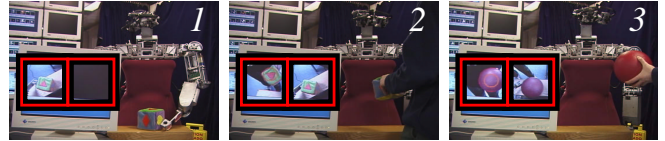


Figure 11: This figure shows stills from a short interaction with Cog. The area highlighted with squares show the state of the robot – the left box gives the view from the robot’s camera, the right shows an image it associates with the current view. In the first frame, the robot is looking at a cube, which it does not recognize. It pokes the cube, segments it, and then it can recognize the cube in future (frame two) and distinguish it from other objects it has poked such as the ball (frame three).

struck, and was able to use that information to invoke rolling behavior in objects such as a toy car (Metta and Fitzpatrick, 2003).

7. Discussion and conclusions

In one view of developmental research the goal is to identify a minimal set of hypotheses that can be used to bootstrap the system towards a higher level of competency. In the field of visuo-motor control some authors (Metta et al., 1999, Marjanović et al., 1996) used this approach, initializing a robotic system with simple behaviors and then developing more complicated ones through robot-environment interaction. In this paper we have shown that object segmentation based on minimal and generic assumptions represents a productive basis for such work. Related work (Metta and Fitzpatrick, 2003) has shown that behavior dependent on robot-object interaction and mimicry can be based substantially on object segmentation alone. This work also relates to a branch of developmental research that probes very young human infant behavior in search of the building blocks of cognition (Spelke, 2000). It has been observed that very young infants a few hours after birth already possess a bias in recognizing faces, human voices, smell, and in exploring the environment (relatively sophisticated haptic exploration strategies have been documented). Also a crude form of object recognition seems to be in place, to the level of distinguishing roundness or spikiness of objects both haptically and visually, for instance. In this paper we examined yet another possible candidate: object segmentation. We did not venture into the definition of the developmental rules that might help the robot in building complex behaviors by means of this primitive, but showed that in principle a system can build on top of object segmentation. We

also showed that both higher level abilities such as recognition or lower level vision (edge orientation estimation) can benefit from this approach. In the future the developmental mechanism allowing the combination of these hypothetical building blocks into complex behaviors will be the subject of investigation.

Acknowledgements

We would like to thank our anonymous reviewers for their very constructive feedback. Funds for this project were provided by DARPA as part of the “Natural Tasking of Robots Based on Human Interaction Cues” project under contract number DABT 63-00-C-10102, and by the Nippon Telegraph and Telephone Corporation as part of the NTT/MIT Collaboration Agreement. Artur Arsenio was supported by a Portuguese grant PRAXIS XXI BD/15851/98.

References

- Arsenio, A. (2002). Boosting vision through embodiment and situatedness. In *MIT AI Laboratory Research Abstracts*.
- Ballard, D. H. (1991). Animate vision. *Artificial Intelligence*, 48(1):57–86.
- 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.
- Comaniciu, D. and Meer, P. (1997). Robust analysis of feature spaces: Color image segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico.
- Fitzpatrick, P. (2003). *From First Contact to Close Encounters: A developmentally deep perceptual system for a humanoid robot*. PhD thesis, Massachusetts Institute of Technology, Department of Electrical Engineering Computer Science, Cambridge, MA.
- Freeman, W. T. and Adelson, E. H. (1991). The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9):891–906.
- 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.
- Johnson, S. P. (2002). Development of object perception. In Nadel, L. and Goldstone, R., (Eds.), *Encyclopedia of cognitive science*, volume 3, pages 392–399. Macmillan, London.
- Johnson, S. P. (2003). Theories of development of the object concept. In Bremner, J. G. and Slater, A. M., (Eds.), *Theories of infant development*. Blackwell, Cambridge, MA. In press.
- Jusczyk, P. W., Johnson, S. P., Spelke, E. S., and Kennedy, L. J. (1999). Synchronous change and perception of object unity: evidence from adults and infants. *Cognition*, 71:257–288.
- Kemp, C. C. (2002). Humans as robots. In *MIT AI Laboratory Research Abstracts*.
- 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.
- Metta, G. and Fitzpatrick, P. (2003). Better vision through manipulation. *Adaptive Behavior*. In press.
- Metta, G., Sandini, G., and Konczak, J. (1999). A developmental approach to visually-guided reaching in artificial systems. *Neural Networks*, 12:1413–1427.
- Needham, A. (2001). Object recognition and object segregation in 4.5-month-old infants. *Journal of Experimental Child Psychology*, 78(1):3–22.
- Perrett, D. I., Mistlin, A. J., Harries, M. H., and Chitty, A. J. (1990). Understanding the visual appearance and consequence of hand action. In *Vision and action: the control of grasping*, pages 163–180. Ablex, Norwood, NJ.
- Sandini, G., Gandolfo, F., Grosso, E., and Tistarelli, M. (1993). Vision during action. In Aloimonos, Y., (Ed.), *Active Perception*, pages 151–190. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Spelke, E. S. (2000). Core knowledge. *American Psychologist*, 55:1233–1243.
- Wolfson, H. and Rigoutsos, I. (1997). Geometric hashing: an overview. *IEEE Computational Science and Engineering*, 4:10–21.