Cognitive Primitives for Mobile Robots

Ethan J. Tira-Thompson, Neil S. Halelamien, Jordan J. Wales, and **David S. Touretzky**
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15123-3891

Abstract

Tekkotsu (see `www.Tekkotsu.org`) is an application development framework for the Sony AIBO mobile robot that endeavors to provide an intuitive set of primitives for perception, manipulation, and attentional control, drawing on insights from cognitive science. The framework also addresses some of the human-robot interaction problems faced by mobile robot application developers.

`www.Tekkotsu.org`

Recent work: Perception

Tekkotsu provides a "dual coding" (iconic and symbolic) representation of images. The term comes from Paivio's "dual coding theory" of mental representations, which posits parallel verbal and non-verbal (i.e., imagistic) systems with extensive referential connectionst between them (Paivio 1986). On this view, simple cognitive problem solving invokes transformations on items in one of these representational systems; more demanding tasks require translation between the two. In Tekkotsu, visual routines (Ullman 1984; Roelfsema, Lamme, & Spekreijse 2000) operate on the iconic representations, extraction operators derive symbolic structures from them, and rendering operators convert from the symbolic form back to the iconic.

Applications can use these primitives to parse the visual world, with intermediate results automatically organized into a derivation tree. Our current touchstone task is playing tic-tac-toe on a real board, as shown in Figure 1.

The system actually maintains two dual coding representations, called "camera space" and "world space". World space is a persistent allocentric representation built up from successive egocentric camera images. Symbolic objects in camera space are matched against the symbolic world map, which is updated with each new image frame. The map is also rendered iconically, allowing visual routines to operate on the image for path planning.

Current work: Manipulation

We are currently working on designing closed-loop manipulation primitives for visually guided pushing or dragging of small objects. Different types of motion are required to push an object along a path, drag or swipe an object away from a barrier, or apply pressure to an object such as a key on a computer keyboard.

A manipulation schema comprises the following elements: (i) point of contact on the object; (ii) type of contact: ballistic impact, steady pressure, rolling swipe, etc.; (iii) type of intended motion: along a defined path, perpendicular to a surface, etc.; (iv) type of visual monitoring employed; (v) motion constraints, e.g., okay or not okay to hit other objects, keep away from the wall, etc.

In the visual monitoring component, we are exploring the notion of "affordances" (Gibson 1977; 1979) by having object perception routines return parameters for possible manipulations rather than mere scene descriptions. We anticipate that this approach will make it easy and intuitive for users to develop new behaviors for manipulating objects.

Future work: Attention

Because the camera has a limited field of view and the neck a limited range of motion, the robot needs to make both head and body motions to overcome its perceptual limitations. To push a tic-tac-toe game piece into position, the robot must divide its attention among looking down to monitor the piece it is moving, looking ahead to see where it's going, and looking around to maintain its position on the world map. Systems that autonomously manage the robot's attention can relieve programmers of a significant burden. We envision a next generation robot control system in which programmers make suggestions about where the robot should devote its attentional resources, but the scheduling and execution of individual head and body motions is done autonomously. Doing this intelligently requires detailed representations of the robot's planned behavior, so that visual search can be smoothly integrated with path following and object manipulation-related head and body motions.

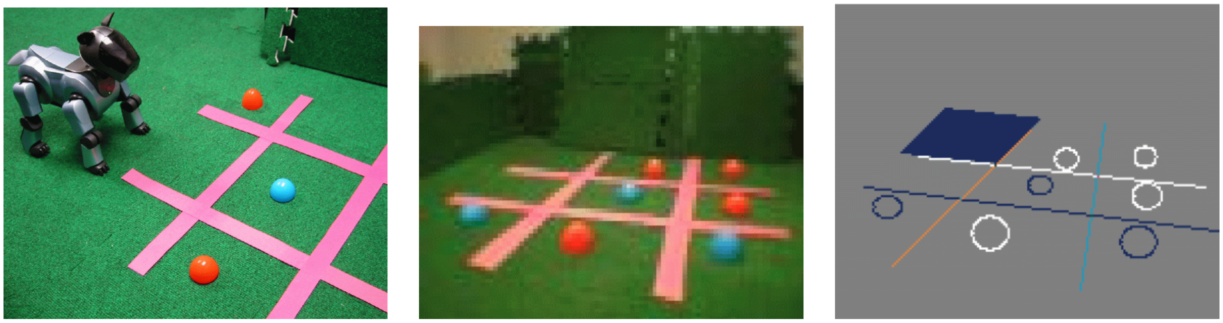

Figure 1: (a) AIBO observing tic-tac-toe board. (b) View through the AIBO's camera. (c) Parsed board image, with the robot's intended move highlighted.

Human-Robot Interaction

Tekkotsu has been used to explore several issues in human-robot interaction. For some purposes it's better to control a robot remotely via a GUI tool or a command line interface, but other situations call for hands-on control, using buttons on the robot's body. Tekkotsu provides a menu system abstraction that supports all three modes of interaction. Buttons on the body can be used to navigate a recursive menu system, with feedback provided by audio cues and the AIBO's LED display. In the future we may add support for a handheld PDA interface as well.

Tekkotsu's philosophy is that robot operations should be fully transparent. "Sketches" computed by visual routines, and symbolic objects extracted from these sketches, are automatically organized into a derivation tree that is displayed using a graphical browser (Figure 2), allowing the programmer to directly examine the robot's parse of the image (Figure 3.) In addition, the robot can construct "sketches" specifically for display to the user, as in Figure 1(c) where it indicates its intended move.

Tekkotsu is in use at a dozen laboratories around the world. Future plans include development of an undergraduate course in Cognitive Robotics to explore how people can make use of cognitive primitives to more efficiently construct new robot applications.

Acknowledgments: Supported by NSF REU supplement to IIS-9978403, and a grant from the Sony corporation.

References

Gibson, J. J. 1977. The theory of affordances. In Shaw, R. E., and Bransford, J., eds., *Perceiving, Acting, and Knowing*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gibson, J. J. 1979. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Paivio, A. 1986. *Mental Representations: A Dual-Coding Approach*. New York: Oxford University Press.

Roelfsema, P. R.; Lamme, V. A. F.; and Spekreijse, H. 2000. The implementation of visual routines. *Vision Research* 40:1385–1411.

Ullman, S. 1984. Visual routines. *Cognition* 18:97–159.

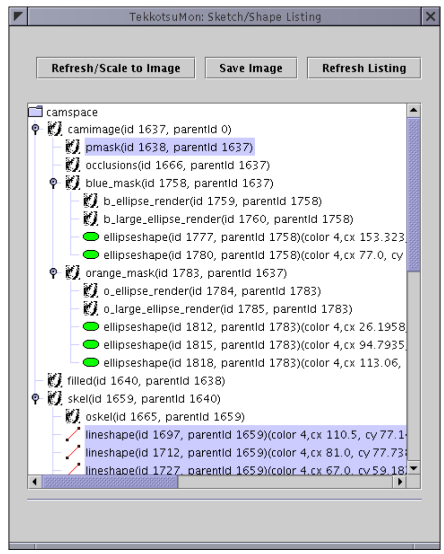

Figure 2: Graphical browser showing the derivation tree for the tic-tac-toe board parse.

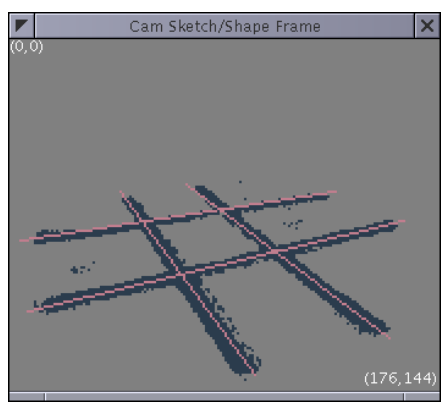

Figure 3: Board pixels and extracted lines as displayed by the graphical browser.