Tekkotsu: Cognitive Robotics on the Sony AIBO

Ethan J. Tira-Thompson, Neil S. Halelamien,
Jordan J. Wales, and David S. Touretzky
{ejt,nsh,jjw,dst}@cs.cmu.edu
Computer Science Department, Carnegie Mellon University
Pittsburgh, PA 15213

Introduction

Tekkotsu (the name means "framework" in Japanese) is an application development framework for the Sony AIBO robot that endeavors to provide an intuitive set of primitives for perception, manipulation, and control. The framework also addresses some of the human-robot interaction problems faced by mobile robot application developers. See `www.Tekkotsu.org`. The contributions of this work lie in three areas:

1. Providing a set of intuitive, high level primitives for robotics that draw on insights from cognitive science.
2. Making robot programming accessible to a wider audience by teaching students to use these primitives to compose complex behaviors.
3. Offering a robotic platform that cognitive modeling systems could use to interface with the physical world.

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 connections between them (Paivio, 1986). On this view, cognitive problem solving invokes transformations on items within these representational systems; more demanding tasks require translation between the two. In Tekkotsu, visual routines (Ullman, 1984) are provided for operating on pixelized iconic representations, extraction operators produce symbolic descriptions (lines and ellipses), and rendering operators convert from the symbolic form back to the iconic.

Applications can use these primitives to parse the visual world. Our current touchstone task is getting AIBO to play tic-tac-toe on a real board, as shown in Figures 1-3. We use color image segmentation to locate the board and game pieces, and symbolic operations on these lines and ellipses to determine which board positions are occupied.

"Sketches" computed by visual routines, and symbolic objects (lines and ellipses) extracted from these sketches, are automatically organized into a derivation tree that can be viewed with a graphical browser running on a PC, allowing the programmer to directly examine the robot's "mental imagery". In addition, the robot can

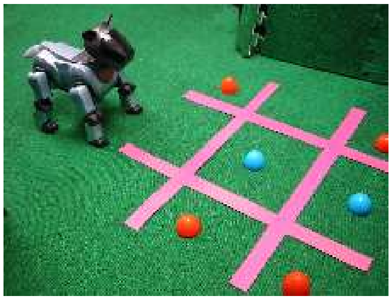

Figure 1: AIBO surveying the game.

Figure 2: AIBO's view of the board.

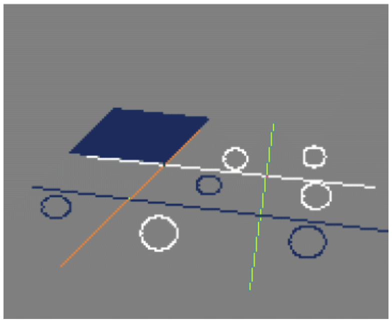

Figure 3: Visual parse of the board, with next move indicated by shading.

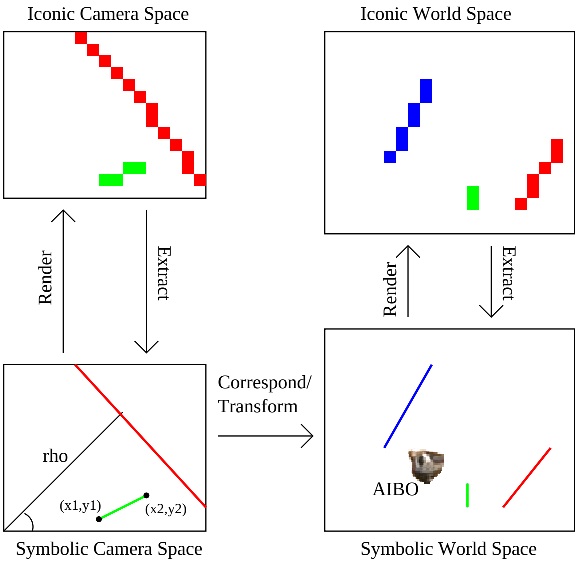

Figure 4: Updating the world map by matching objects in camera space against existing world map structures.

construct sketches specifically for display to the user, as in Figure 3(c) where it indicates its intended move.

Tekkotsu maintains a world map using the same dual coding strategy, as shown in Figure 4. The world map is a persistent allocentric representation built from successive egocentric camera images. Symbolic objects in camera space are matched against their counterparts in the world map, which is updated with each new image frame. The world map is also rendered iconically, allowing visual routines to operate on the map for path planning.

Manipulation

We are currently 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, steady pressure, rolling swipe); (iii) intended motion (along a path, perpendicular to surface, etc.); (iv) type of visual monitoring employed; and (v) motion constraints (okay or not okay to hit other objects, keep away from the wall, etc.) We anticipate this approach will make it intuitive for users to develop new behaviors for manipulating novel objects.

Control

Tekkotsu provides a state machine formalism for describing robot behaviors. State nodes and transition arcs are both implemented as computational processes. Communication takes place via an event-based mechanism. For example, a transition may subscibe to a particular stream of sensor events, and when such an event occurs, the transition will deactivate one state and activate another.

The state machine is hierarchical: a state may contain another state machine inside it. And transitions can "fork", allowing multiple states to become active at once. Tekkotsu does not presently incorporate any long term or working memory model, other than its world map representation. Such functionality could be provided by interfacing with widely used cognitive modeling systems such as SOAR or ACT-R. (Tekkotsu can communicate over wireless with processes running on a PC.)

Attention

A longer range goal for Tekkotsu is to integrate perception with manipulation and navigation. For example, it should be possible to construct a path by "visualizing" it on the world map, and then have the robot traverse that path. Similarly, when manipulating an object, parameters such as "point of contact" or "direction to move along" might be expressed as markers in a sketch produced by a visual routine.

A major issue that remains to be addressed is control of attention. The robot's camera has a limited field of view and the head has a limited range of motion. In order to monitor the object it is manipulating, verify its own location, check for nearby obstacles, and keep track of any humans it's interacting with, the robot must divide its attention and interleave a variety of head and body motions with the other actions it is performing. Intelligent scheduling of these motions will require an explicit representation of the robot's short term planned behavior, so that, for example, it does not stop and look backwards just before reaching a goal location where the entire body will need to turn around anyway. An attentional model that managed the robot's limited resources and arbitrated among competing demands would relieve application developers of a significant burden, and perhaps suggest approaches animals may be using in similar circumstances.

Acknowledgments

This work was funded in part by a grant from the Sony Corporation, and by an REU supplement award to NSF grant IIS-9978403.

References

Paivio, A (1986). *Mental representations: a dual-coding approach.* New York: Oxford University Press.

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

`www.tekkotsu.org`