Extracting User Intent in Mixed Initiative Teleoperator Control

Andrew H. Fagg,* Michael Rosenstein, Robert Platt, Jr., and Roderic A. Grupen
University of Massachusetts, Amherst, Massachusetts, 01003, USA

I. Introduction

User fatigue is common with robot teleoperation interfaces. Mixed initiative control approaches attempt to reduce this fatigue by allowing control responsibility to be shared between the user and an intelligent control system. A critical challenge is how the user can communicate her intentions to the control system in an intuitive manner as possible. In the context of control of a humanoid robot, we propose an interface that uses the movement currently commanded by the user to assess the intended outcome. Specifically, given the observation of the motion of the teleoperated robot for a given period of time, we would like to automatically generate an abstract explanation of that movement. Such an explanation should facilitate the execution of the same movement under the same or similar conditions in the future.

How do we translate these observations of teleoperator behavior into a deep representation of the teleoperator's intentions? Neurophysiological evidence suggests that in primates, the mechanisms for the recognition of the actions of other agents are intertwined with the mechanisms for execution of the same actions. For example, Rizzolatti *et al.* (1988)[7] identified neurons within the ventral premotor cortex of monkey that fired during execution of specific grasping movements. Although this area is traditionally thought of as a motor execution area, Rizzolatti *et al.* (1996)[8] showed that neurons in a subarea were active not only when the monkey executed certain grasping actions, but also when the monkey observed others making similar movements. These and other results suggest that generators of action could also facilitate the recognition of motor actions taken by another entity (in our case, the teleoperator).

The foci of this study are teleoperated pick-and-place tasks using the UMass Torso robot.[6] This robot consists of an articulated, stereo biSight head; two 7-DOF Whole Arm Manipulators (WAMs); two 3-fingered hands (each finger is equipped with a six-axis force/torque sensor); and a quadraphonic audio input system. The teleoperator interface consists of a red/blue stereo display and a P5 Essential Reality glove that senses the position and orientation of the user's hand, as well as the flexion of the user's fingers.

II. Control Projection for Estimating Teleoperator Intent

A. A Basis Set of Controllers

In order to function in stochastic environments, control primitives must be robust to uncertainty and noise. This includes weak or non-functional actuators and noisy or incorrect data concerning the external environment. Closed-loop control is a well-studied mechanism for rejecting noise in these environments. In this approach, an error function formalizing the control objective is continuously evaluated and used to update control signals in a way that reduces error. Closed-loop control laws are robust to stochastic environmental

*Now at the University of Oklahoma.

perturbations because in the controllability region, the control law drives the system back toward the basin of attraction.

The *control basis* describes a large set of controllers defined by the Cartesian product of sensory abstractions (that define metric spaces), potential functions defined over those space, and the sets of controllable resources.[4–6] A particular parameterized controller selects one element from each of these three sets. During execution, the controller produces control actions that follow the negative gradient of the potential function. For example, a controller that is responsible for moving the hand of a robot to a particular location in the global coordinate frame could be parameterized with sensors that report the Cartesian location of the goal and of the hand, a potential function that measures distance to the goal, and the joints of the robot's left arm. The action of the controller is to move the joints of the arm so as to reduce the distance to the specified target. Note that there may be multiple discrete goals or even continuous manifolds of goals in the robot's configuration space.

B. Generating a Set of Context-Specific Hypothesis Controllers

The set of objects within the workspace of the robot *afford* a discrete set of actions that can be executed by the robot. Here, we use the term "affordance" to mean a mapping from some form of object description to the different ways in which the object can be acted upon.[1–3] In other words, the affordance translates object properties into a specific parameterization of a set of controllers (including a description of the goals for the various controllers). It is important to note that this mapping is not necessarily unique. For example, an object may be grasped in one of many different ways (each of which is captured as a separate parameterized controller).

In our implementation, a vision system extracts coarse object models that capture approximate pose, shape, size, and color distribution of the objects. A set of heuristically generated affordances translate these object descriptions into controller parameterizations that correspond to reach-to-grasp motions. Any one of these resulting controllers could be activated in order to grasp an object within the current workspace.

The *control projection approach* uses the structure provided by the set of available controllers as a framework in which to interpret the control actions provided by the teleoperator. In this context, a controller becomes a passive entity – the action of the controller does not affect the robot, but instead is compared against the control actions taken by the teleoperator. In general, when there is some form of agreement between these two control actions over some non-trivial period of time, the controller is considered to be an explanation of the actions taken by the teleoperator.

III. Predicting Intent for Mixed Initiative Control

The control projection approach can be used as a mechanism for predicting the intended goal of a movement that is being commanded by a teleoperator. Such a prediction (when correct) allows a control system to intervene on behalf of the teleoperator to complete the initiated movement. This is particularly useful when the movement requires a significant degree of precision because it reduces the responsibility of the user in the fine guidance of the movement and it gives the user the opportunity to rest for a short period of time.

In our implementation, both the actions of the hypothesis controller and the teleoperator can be described in terms of a joint velocity vector. The instantaneous degree of match between the two is measured as a function of the angle between these two joint space vectors. When the degree of match is high over a period of samples, we consider the likelihood of the hypothesis controller explaining the given motion to also be high. The relative likelihood for each hypothesis is communicated to the user by superimposing a circle on the 3D teleoperator display whose location corresponds to that of the matching object. The diameter of the circle encodes the relative likelihood of the hypothesis that the user is currently reaching for that object. The hypothesis with the maximum likelihood may be selected at any time by the user through a simple hand gesture. Once selected, control is handed from the user to the automated system (which may be interrupted with a subsequent gesture). With this style of interface, the user is able to visually confirm that the hypothesized movement is the correct one prior to relinquishing control.

IV. Extracting Sequences of Submovements

In this class of tasks, the teleoperator demonstrates a sequence of pick-and-place movements. The goal is for the system to extract a high-level representation of the sequence that enables the control system to reproduce the sequence in similar conditions (including different objects and workspace locations). This problem is particularly challenging because it is not always clear when one movement ends and the next begins.

In our implementation, a subsequence of action observations is explained by a single hypothesis controller when the following conditions are met. First, each observation step in the subsequence closes the distance to the goal (the system makes progress with respect to the potential function). Second, the distance to the controller's goal at the end of the subsequence is small (the system is close to a well – or a goal – in the potential function). Third, the end of the sequence is punctuated by a tactile event (as an object is grasped or released by the robot's hand).

Preliminary results show that sequences of pick-and-place operations can be reliably extracted from the demonstration. Because each hypothesis controller is paired with a description of the original object, this sequence of operations can be described as a combination of the object properties and the way in which the object was grasped. Furthermore, when presented with a similar situation, the same sequence of actions can be executed automatically (following a role assignment step). This execution is performed using the same set of controllers used in the recognition phase, relativized to the new locations of the participating objects.

Acknowledgments

Preparation of this manuscript was supported by grants from NSF (#CDA 9703217), NASA/RICS (University of Houston #215), NASA (#NAG-9-1379, and DARPA-MARS subcontract #NAG-9-1445), and DARPA-MARS (#DABT63-99-1-0004). Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Aeronautics and Space Administration, the National Science Foundation, or the Defense Advanced Research Projects Agency.

References

[1] J. Gibson and E. Gibson. Perceptual learning: Differentiation or enrichment? *Psychological Review*, 62:32–51, 1955.

[2] J. J. Gibson. The optical expansion-pattern in aerial location. *American Journal of Psychology*, 68:480–484, 1955.

[3] J. J. Gibson. *The Senses Considered as Perceptual Systems*. Allen and Unwin, 1966.

[4] M. Huber. *A Hybrid Control Architecture for Adaptive Robot Control*. PhD thesis, Department of Computer Science, University of Massachusetts, Amherst, 2001.

[5] Manfred Huber and Roderic A. Grupen. A hybrid discrete event dynamic systems approach to robot control. Technical Report 96-43, University of Massachusetts Amherst, Department of Computer Science, October 1996.

[6] R. Platt, Jr., O. Brock, A. H. Fagg, D. Karupiah, M. Rosenstein, J. Coelho, M. Huber, J. Piater, D. Wheeler, and R. A. Grupen. A framework for humanoid control and intelligence. In *the Proceedings of Humanoids 2003*, 2003.

[7] G. Rizzolatti, R. Camarda, L. Fogassi, M. Gentilucci, G. Luppino, and M. Matelli. Functional organization of inferior area 6 in the macaque monkey II. area F5 and the control of distal movements. *Experimental Brain Research*, 71:491–507, 1988.

[8] G. Rizzolatti, L. Fadiga, V. Gallese, and L. Fogassi. Premotor cortex and the recogntion of motor actions. *Cognitive Brain Research*, 3:131–141, 1996.