Teleassistance: Contextual Guidance for Autonomous Manipulation*

Polly K. Pook and Dana H. Ballard
Computer Science Department
University of Rochester
Rochester, NY 14627-0226 USA
pook|dana@cs.rochester.edu

Abstract

We present *teleassistance*, a two-tiered control structure for robotic manipulation that combines the advantages of autonomy and teleoperation. At the top level, a teleoperator provides global, *deictic* references via a natural sign language. Each sign indicates the next action to perform and a relative and hand-centered coordinate frame in which to perform it. For example, the teleoperator may point to an object for reaching, or preshape the hand for grasping. At the lower level autonomous servo routines run within the reference frames provided. Teleassistance offers two benefits. First, the servo routines can position the robot in relative coordinates and interpret feedback within a constrained context. This significantly simplifies the computational load of the autonomous routines and requires only a sparse model of the task. Second, the operator's actions are symbolic, conveying intent without requiring the person to literally control the robot. This helps to alleviate many of the problems inherent to teleoperation, including poor mappings between operator and robot physiology, reliance on a broad communication bandwidth, and the potential for robot damage when solely under remote control. To demonstrate the concept, a Utah/MIT hand mounted on a Puma 760 arm opens a door.

Introduction

Autonomous servo control and teleoperation have complementary advantages and disadvantages as robotic control schemes. Servo control of robotic manipulation has benefited from the development of compliant manipulators with rich position and force sensing capabilities. When situated in a local context, fast distributed servo feedback enables a robot to react quickly and appropriately [Brooks 1986][Connell 1989]. Servo control currently suffers, however, from a poor ability to perceive and plan according to global state. Teleoperation addresses this weakness by putting a person in the loop, one who can provide more global guidance. But teleoperation eliminates local servo control, resulting in slow jerky movement that is quite tedious for the teleoperator. Local reactivity is lost. What is needed is a two-layer control strategy: a high level to select a context in which low-level behaviors situate

Consider the advantages of each control mechanism for the hypothetical task of planing a board. A teleoperator can more easily set up the task by locating and positioning the planer on the board, than an autonomous robot can. Once context is established the servo controller more readily maintains a smooth, sliding contact with the board. It moves along the central axis of the planer, independent of world coordinates, while tightly monitoring force feedback on an orthogonal axis, independent of the orientation of the board. Feedback is interpreted within the context of planing to adjust force and position. The teleoperator provides a high-level reference context that low-level servo behaviors could exploit.

Setting up a reference frame for subsequent relative actions is an example of a *deictic*, or pointing, strategy [Agre & Chapman 1987]. We propose a *teleassisted* controller for robotic manipulation that interfaces high-level deictic strategies with context-sensitive servo behaviors. A person wearing a master teleoperation device provides the deictic references. Rather than literally teleoperate the robot and bear the concomitant problems of delayed and limited feedback, the operator uses a gestural sign language to select successive contexts for low-level autonomous force and position control.

Deictic strategies

Studies suggest that animals use high-level deictic strategies to bind low-level perceptual and motor behaviors to the current context [Kowler & Anton 1987][Ballard et al. 1992]. Visual fixation is an example of a deictic strategy that binds motor behavior to a relative coordinate frame. For example, an object can be grasped by first looking at it and then directing the hand to the center of the image coordinate frame. In depth, the hand can be servoed relative to

*This work was supported by NSF research grant no. IRI-8903582 and by a research grant from the Human Science Frontiers Program.

the horopter using binocular cues. This strategy is invariant to agent movement and task locale.

Like visual fixation, the act of grasping an object provides a framework for interpreting the associated kinesthetic feedback. For example, bi-manual animals often use one hand as a vise and the other for dexterous manipulation. We hypothesize that the vise hand marks a reference frame and that the dexterous motions are made relative to that frame. Similarly, body position creates a reference for motor actions. Studies of human hand-eye coordination have found head movements to often be closely associated with hand movements. [Pelz et al. 1994] found that most subjects square their head with respect to the site of a complex manipulation action and then hold it very still during the manipulation, even when the eyes are looking elsewhere. This fits our hypothesis if we consider head position as a reference frame for relative hand motion. To make this more concrete as a control strategy, we assign a deictic strategy two roles: to make a *temporal* and a *spatial* binding to the world.

Temporal binding

The intended action, such as reaching or grasping, binds the servo controller to the *temporal context.* Without a detailed model of the world, many interpretations of feedback are possible; temporal context can allow the controller to select the correct one. Earlier studies demonstrate the importance of temporal context in reducing the complexity of robot motor tasks. In [Pook & Ballard 1992], a Utah/MIT dexterous manipulator autonomously grasps a spatula, positions it in a pan and flips a plastic egg without relying on precise positioning or force information. Each successive action is controlled by interpreting qualitative changes in force feedback within the current context. For example, identifying force contact can be done by simply noting a significant change in tension on any hand joint, regardless of the particular hand or task configuration. To determine what event is associated with a qualitative change in force, whether it be from grasping the spatula, placing it in the pan or sliding it under the egg, the controller can refer to the current context. In deterministic sequential tasks, this context is implicit in the control program. This approach has parallels with the behavioral approach of [Salganicoff & Bajcsy 1991] but relies on local feedback.

Spatial binding

A *spatial binding* defines a relative coordinate frame for successive perceptual and motor behaviors. Deictic bindings avoid world-centered geometry that varies with robot movement. For fixation the reference is gaze-centered. To open a door, for instance, looking at the doorknob defines a relative servo target. [Crisman & Cleary 1994] demonstrate the computational advantage of target-centered frames for mobile robot navigation.

Pointing and preshaping the hand create hand-centered spatial frames. Pointing defines a relative axis for subsequent motion. In the case of preshaping, the relative frame attaches within the *opposition space* [Arbib, Iberall, & Lyons 1985] of the fingers. With adequate dexterity and compliance, simply flexing the fingers toward the origin of that frame coupled with a force control loop suffices to form a stable grasp. Since the motor action is bound to the local context, the same grasping action can be applied to different objects, a spatula, a mug, a doorknob, by changing the preshape.

A teleassisted control strategy

In teleassistance, the operator, wearing a master device, guides the robot with natural gestures that correspond to the task at hand. These gestures form a sign language. The signs are deictic in that they define a spatial and temporal context for subsequent robot behaviors. A sign is recognized by monitoring the operator's finger positions and matching them to stored models. On recognition of a hand sign, the controller calls the servo behaviors, with the current spatial context as a parameter. In this way, a person motions the robot through a complex and possibly non-deterministic task.

Each sign can be mapped to a high-level state in a non-deterministic finite state machine (FSM), as shown in Figure 1. At top are the deictic states, marked with dashed lines. From each high-level state, control is passed to the appropriate low-level servo routines, shown at bottom, that perform the intended action. The two state classes are separated in the figure for clarity. In practice, all states belong to a single FSM for the task. The topology encodes the task context.

We will illustrate teleassistance through the task of opening a door. A simple scheme for door-opening is to reach for the handle, then grasp and turn it. Under teleassistance, the task might be performed as follows. The operator steers the robot to the door handle by pointing and shapes the hand and wrist for the handle type. The robot in turn moves along the given direction until it bumps into the door, copies the operator's hand shape, moves to contact the handle, then grasps and turns it. These behaviors are described in detail in the next section.

The non-determinism of the FSM affords the operator some flexibility. The operator may choose to compose task actions arbitrarily when order is not crucial. Or the operator may wish to take emergency action when an error is detected. The FSM topology can be designed to support such deviations. In our simple examples, the operator can rest or stop at intermediate points in the task. It is worth noting that the topology must be pre-specified for each task and so this flexibility is restricted.

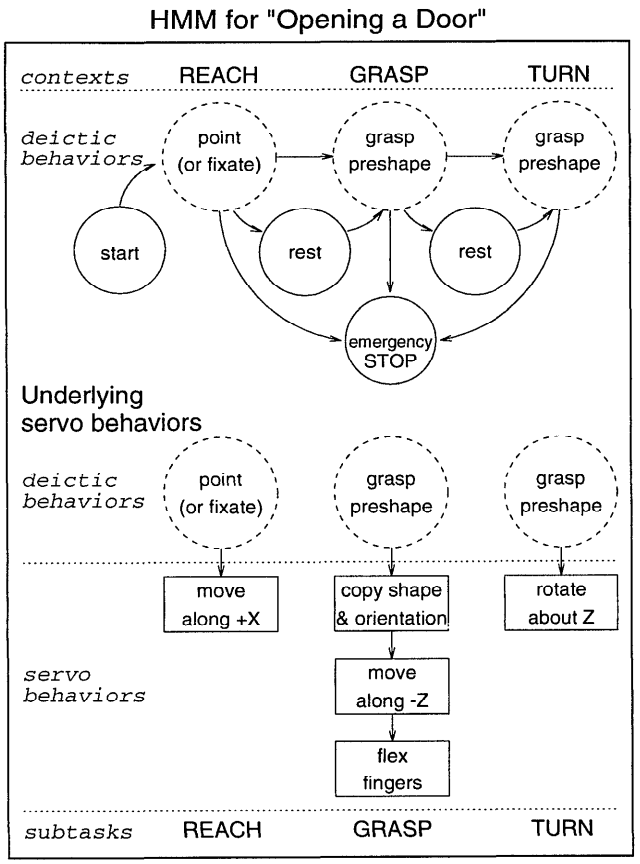

Figure 1: A FSM for opening a door. The flow of control among high-level states is shown at *top*. Each deictic gesture, marked with a dashed line, defines a relative coordinate frame and a task context (REACH, GRASP, TURN). From a deictic state, control is passed to appropriate low-level robot behaviors that perform the sub-task (shown at *bottom*). Each low-level action moves within its relative frame, servoing on guard conditions such as force contact.

Example: Opening a door

To illustrate the two-tiered control strategy, a robot manipulator opens a small door equipped with a lever handle. The door is placed arbitrarily within the working space of the hand. Each high-level deictic routine, ties a temporal context (REACH, GRASP, TURN) and a relative coordinate frame to subsequent low-level actions. In this example, all coordinate frames are hand-centered.

The low-level servo routines are a sequence of guarded moves. The direction of each move (of the fingers and/or the arm) is implicit in the context of the task and is made relative to the current frame. Feedback consists of finger position error, i.e., the difference between the actual and commanded finger joint angles. A change in the position error is interpreted within the current context of REACHING, GRASPING, etc., to control the action appropriately. Each routine is described in greater detail below.

Lab Setup

Hardware The manipulator is a sixteen degree-of-freedom Utah/MIT hand mounted on a six degree-of-freedom PUMA 760 arm. The hand has four fingers with four joints apiece. A pair of pneumatically driven agonist-antagonist tendons actuates each joint. Hall effect sensors monitor each joint angle. The hand is both dexterous and compliant so its control strategy can ignore many inessential variations in the task configuration, such as the precise shape and orientation of a door handle [Pook & Ballard 1992].

The teleoperator wears an EXOS Dexterous Hand Master (TM) that measures four joint angles on each of four fingers. Additionally, an Ascension Bird (TM) polhemus sensor mounted on the back of the teleoperator's hand measures the position and orientation of the operator's arm and wrist in lab coordinates.

Coordinate frame mapping The teleoperator and the robot each have their own relative coordinate systems with a mapping between them. For arm motion, only the orientation of the teleoperator's polhemus sensor is mapped to the corresponding robot frame. For finger motion, only the joint angles are mapped. In either case, translation is ignored.

The sign language The sign language for this task consists of six signs: point, preshape, rest, emergency stop, speedup, and slow down. Previously, we empirically derived a range of permissible joint angles for the operator's hand and arm for each sign. To recognize a sign on-line, the program monitors the joint angles on the EXOS and polhemus devices and identifies a match when it occurs. We are currently working on learning new signs automatically, rather than through empirical determination. See the last section, *Future Work*, for details.

The program

REACHING for the door. The teleoperator commences action by pointing the robot arm toward the door, as shown in the upper left of Figure 2. This provides a temporal launching point for the program and defines a spatial coordinate frame centered on the back of the robot hand. While the operator points, the PUMA moves in the direction of the pointing axis, independently of world coordinates. Thus the robot reach is made relative to a deictic axis which the teleoperator can easily adjust.

The autonomous move monitors the guard condition of a change in position error on any of the Utah/MIT finger joints. In the context of REACHING, the change in robot finger position is interpreted as contact with a non-compliant surface and the reach halts and backs off a small distance from the point of contact. So, when the hand bumps into the door, the arm stops.

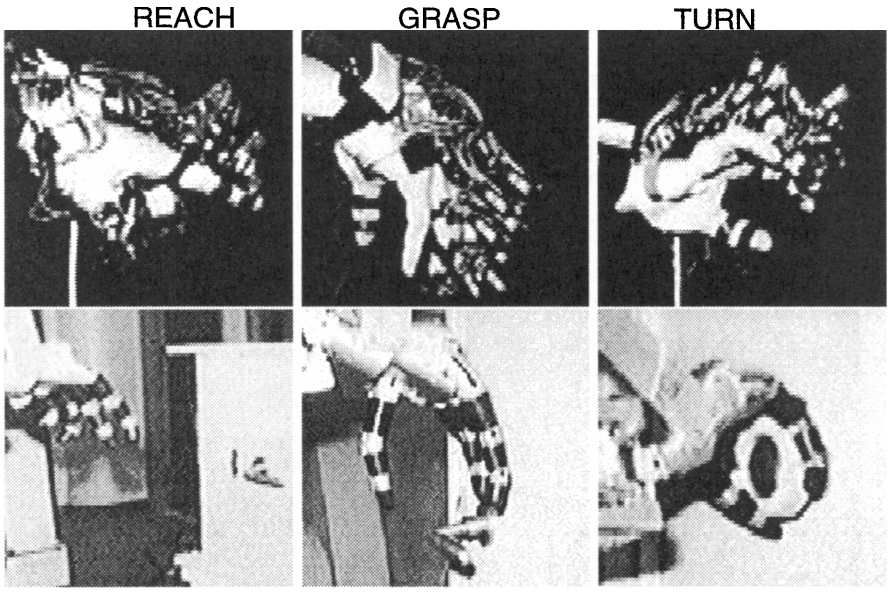

Figure 2: At top are the teleoperator reference poses to REACH, GRASP and TURN the door handle. For illustrative purposes both a lever and a knob are shown, although only the lever is used. Snapshots of subsequent robot behavior are shown at bottom. In REACHING, the context of a pointing index finger signifies that only the x-direction of the polhemus sensor is important. In GRASPING the overall shape of the three fingers and the thumb define a preshape for the particular handle. In TURNING, the grasp shape reveals the handle type and, by extension, its pivot point. Notably, the robot hand itself could provide the deictic reference for TURNING, by noting its current pose.

The FSM has self-loops in each deictic behavior so that the operator can successively point in new directions to accurately place the robot hand. In a typical trial, the operator points the robot arm toward the door. When the robot reaches the door by sensing contact, the operator repeatedly points anew to position the hand over the door handle. Each movement halts when the operator changes the pointing axis, stops pointing, or the hand detects contact. A new pointing axis initiates a new movement. When satisfied, the teleoperator adopts a grasp preshape. The FSM recognizes the new sign and shifts context to that of GRASPING.

GRASPING the door handle. A grasp preshape defines a new spatial frame centered on the palm of the hand. The robot mimics the preshape using a linear, joint to joint mapping between the EXOS and the robot hand [Speeter 1993]. The middle column of Figure 2 shows the preshape for turning the door lever. Subsequent servo routines find the handle and grasp it.

In the new spatial context, the negative Z-axis points out of the palm. To find the handle, the robot hand moves along this axis until it senses contact. The advantage of a deictic framework can be seen by comparing this action for different door handles. For the door *lever*, the hand-centered Z-axis points downward. For a *knob*, the Z-axis is horizontal. The move is independent of world coordinates, however, because it is made relative to the hand frame.

A change in position error, resulting from contact with the doorhandle, stops the motion.[1] In either case, the controller interprets the error as contact and so, within the current context, proceeds to grasp.

To grasp the door handle, the robot flexes the finger joints until they are all either fully flexed or maintaining a position error. The preshape set the fingers in opposition to one another as needed to grasp the particular door handle. Because the grasp is performed within the context of that preshape this strategy applies to either door handle.

TURNING the handle. The robot pose (or the human preshape) defines whether the context is one of grasping a knob or a lever. The rightmost column of Figure 2 shows the hand grasping the door knob. In this case the pivot point of the handle is set to the center of the palm (i.e., the origin of the current deictic frame). If the hand shape corresponds to a wrap grasp, then the pivot is about a point alongside the grip, i.e., offset along the current Y-axis which corresponds to the axis of the lever. A new spatial frame is attached to the pivot point and the arm rotates about the new Z-axis. Such a rotation is much simpler than computing an equivalent trajectory in lab space.

[1] The hand must be positioned fairly accurately over the lever, such that one or more of the finger links makes contact with it. A better solution is to have a contact sensor on the palm or a force sensor in the wrist.

The arm continues its rotation until a position error is sensed. In this context, the controller interprets the error to mean the handle's mechanical stop has been reached.

Results

Four different operators each performed ten trials of the task under teleoperation control and ten trials under teleassistance. For each controller, the operators trained until they were comfortable: 10 to 15 minutes for teleoperation and 2 to 10 minutes for teleassistance. Twice during each 10-trial set, the door was moved to a new position selected arbitrarily within the workspace.

The results are shown in Table 1, along with the results of a completely autonomous controller. Teleassistance required about the same amount of time as the autonomous controller, and both are, on average, 33% faster than teleoperation. The time is occupied differently by each controller, however. 60% of the time spent under teleassistance is in fact under autonomous control. Thus, the operator actually spent only 8 seconds, on average, controlling the robot via hand signs. The teleoperator, in contrast, spent more than treble the time (29 sec.) in literal master-slave control of the robot.

The teleoperator, however, is able to avoid failures that other two controllers cannot. In the case of teleassistance, all failures were due to not turning the door handle far enough to free the catch. The servo controller did not detect the mechanical stop correctly. The autonomnous controller failed due to its reliance on hard-coded position information. Thus, changing the doors position twice during the trials resulted in the controller failing (not finding the door or the handle) two-thirds of the time. This figure is not generalizable as it depends entirely on the degree of certainty in the task. In a structured, static environment, the autonomous controller would work very well. In an unstructured world, however, teleassistance is preferable.

Conclusion and related work

The idea of combining traditional teleoperation and autonomous servo controllers has been suggested in various forms by several researchers. [Sayers, Paul, & Mintz 1992] at the University of Pennsylvania Grasp Lab allows the teleoperator to select among a menu of geometric targets. The teleoperation apparatus (the master) is then constrained to motions along the selected geometry. [Yokohoji et al. 1993] suggest building a manual control box which the teleoperator can pre-set to a desired strategy as needed: full teleoperation, full autonomy, or one of two combinations . [Kuniyoshi, Inaba, & Inoue 1992], [Ikeuchi & Suehiro 1992], and [Kang & Ikeuchi 1994] visually recognize teleoperator motions to guide the robot.

Our approach is to make the transfer of control between teleoperator and robot implicit in the context of the task. A task specific FSM encodes the context in its topology. The teleoperator, rather than being responsible for direct control of the robot, guides the robot with gestures and supplies the spatial and temporal frameworks in which to situate the autonomous servo behaviors[2]. The behavior can then operate in a relativistic constrained domain and perform with computational efficiency. This method of providing for local interpretation of feedback compares with the qualitative vision strategies made possible with gaze control vision systems [Ballard 1991][Ballard & Brown 1992] . Given the task context, these methods can rely on a very sparse model of the world. Such strategies are common to animal systems [Bower 1982][Twitchell 1970] but are a notable departure from traditional robot control schema that rely on precise force and position models.

Future work: acquiring and recognizing new hand signs

The transfer of control between high and low-level behaviors, implicit in the task, requires explicit recognition of the sign language used by the operator. Currently, we provide an *a priori* empirical model of each hand sign. This method has at least two drawbacks. One is that the models must be robust to different operators and to the non-linearities of the EXOS device. We have handled this by defining the models quite loosely. However, as more signs are added to the lexicon, overlap will occur and ambiguity will arise. A more accurate, but still robust, method is needed. Secondly, it would be useful to learn signs dynamically, allowing for operator preference.

Earlier studies suggest a remedy to these drawbacks. These studies demonstrate the ability to learn and recognize manipulation primitives from temporal features in the robot state [Pook & Ballard 1993]. (see also [Hannaford & Lee 1991]). In this experiment, several teleoperators performed samples of each primitive while we recorded the finger joint tensions and velocities. By applying Learning Vector Quantization (LVQ) [Kohonen 1990] to these recordings we could create canonical patterns for each primitive. Using these learned patterns we could segment a compound manipulation task, flipping an egg, into its recognizable primitives, although with numerous errors. However these errors, in the form of ambiguities and spurious misclassifications, could be eliminated by performing the segmentation within the context of the task. The context was encoded in the topology of a hidden markov model (HMM), just as the context of opening a door is encoded in an FSM. An HMM is a probabilistic finite state machine that models a markov process; i.e., it accommodates natural variation and is sensitive only to the current state[see Rabiner & Juang 1986].

[2]If other deictic inputs are available, such as a computer vision system with gaze control, they could replace the teleoperator where desired.

Controller	Avg. Time (sec.)	Mean Time (sec.)	% Time under Operator control	% Time under Auto control	Failure Rate
Teleoperation	29	24.5	100%	0%	0%
Teleassistance	20	19.0	40%	60%	13%
Autonomy	20	20.0	0%	100%	66%*

*see text

Table 1: Results for the task of opening a door under the three controllers.

The probabilistic contextual constraint provided by the HMM makes matching considerably easier.

This method of pattern-matching within a known context was robust to the vagaries of different operators and changes in the task configuration. We are now investigating the use of this method to dynamically and automatically learn the deictic sign language. Preliminary results are promising and address the drawbacks of our present empirical scheme.

Acknowledgments

The authors wish to thank Ray Frank, Tim Becker and Luid Bukys for their technical assistance and forbearance and John Lloyd and Vincent Hayward for the RCCL Puma controller.

References

[1] P. E. Agre and D. Chapman. Pengi: An implementation of a theory of activity. In *Proceedings of the Sixth National Conference on Artificial Intelligence*, pages 268–272. Morgan Kaufmann, Los Altos, CA, 1987.

[2] M. Arbib, T. Iberall, and D. Lyons. Coordinated control programs for movements of the hand. Technical report, COINS Dept. of Comp. and Inf. Science, University of Massachussetts, 1985.

[3] D. H. Ballard. Animate vision. *Artificial Intelligence*, pages 57–86, February 1991.

[4] D. H. Ballard and C. M. Brown. Principles of animate vision. *CVGIP: Image Understading*, July 1992.

[5] D.H. Ballard, M.M. Hayhoe, F. Li, and S.D. Whitehead. Hand-eye coordination during sequential tasks. *Proc. of the Phil. Trans. Royal Soc. of London*, 1992.

[6] T. G. R. Bower. *Development in Infancy.* New York: W.H. Freeman and Co., 1982.

[7] R. Brooks. A layered intelligent control system for a mobile robot. *IEEE Journal of Robotics and Automation*, pages 14–23, April 1986.

[8] J. Connell. A colony architecture for an artificial creature. Tech. Report 1151, MIT AI Lab, 1989.

[9] J. Crisman and M. Cleary. Deictic primitives for general purpose navigation. *Proc. of the AIAA Conf. on Intelligent Robots in Factory, Field, Space, and Service (CIRFFSS)*, March 1994.

[10] B. Hannaford and P. Lee. Hidden markov model analysis of force/torque information in telemanipulation. *International Journal of Robotics Research*, pages 528–538, October 1991.

[11] K. Ikeuchi and T. Suehiro. Towards an assembly plan from observation. *Proc. of the IEEE International Conference on Robotics and Automation*, May 1992.

[12] S.B. Kang and K. Ikeuchi. Grasp recognition and manipulative motion characterization from human hand motion sequences. *Proc. of the IEEE International Conference on Robotics and Automation*, May 1994.

[13] T. Kohonen. Improved versions of learning vector quantization. *Proc. of the International Joint Conference on Neural Networks*, June 1990.

[14] E. Kowler and S. Anton. Reading twisted text: Implications for the role of saccades. *Vision Research*, pages 27:45–60, 1987.

[15] Y. Kuniyoshi, M. Inaba, and H. Inoue. Seeing, understanding and doing human task. *Proc. of the IEEE Int'l Conf. on Robotics & Automation*, May 1992.

[16] J. Pelz, M. M. Hayhoe, D. H. Ballard, and A. Forsberg. Separate motor commands for eye and head. *Submitted, Investigative Ophthalmology and Visual Science, Supplement 1994*, 1993.

[17] P. K. Pook and D. H. Ballard. Sensing qualitative events to control manipulation. *Proceedings of the SPIE Sensor Fusion V Conference*, November 1992.

[18] P. K. Pook and D. H. Ballard. Recognizing teleoperated manipulations. *Proc. of the IEEE International Conference on Robotics and Automation*, May 1993.

[19] L.R. Rabiner and B.H. Juang. An introduction to hidden markov models. *IEEE ASSP Magazine*, Jan. 1986.

[20] M. Salganicoff and R. Bajcsy. Sensorimotor learning using active perception in continuous domains. *AAAI Fall Symposium Series: Sensory Aspects of Robotic Intelligence*, November 1991.

[21] C. Sayers, R. Paul, and M. Mintz. Operator interaction and teleprogramming for subsea manipulation. *4th IARP Workshop on Underwater Robotics*, 1992.

[22] T. H. Speeter. Transforming human hnad motion for telemanipulation. *Presence: Teleoperators and Virtual Environments*, 1992.

[23] W. Twitchell. Mechanisms of motor development. In *Reflex Mechanisms and the Development of Prehension.* New York: Academic Press, 1970.

[24] Y. Yokohoji, A. Ogawa, H. Hasanuma, and T. Yoshikawa. Operation modes for cooperating with autonomous functions in intelligent teleoperation systems. *Proc. of the IEEE International Conference on Robotics and Automation*, May 1993.