Two Arms are Better than One: A Behavior Based Control System for Assistive Bimanual Manipulation

Aaron Edsinger
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
edsinger@csail.mit.edu

Charles C. Kemp
Health Systems Institute
Georgia Institute of Technology
charlie.kemp@hsi.gatech.edu

Abstract— **Human environments present special challenges for robot manipulation, since they are complex, dynamic, uncontrolled, and difficult to perceive reliably. For tasks that involve two handheld objects, the use of two arms can help overcome these challenges. With bimanual manipulation, a robot can simultaneously control two handheld objects in order to better perceive key features, control the objects with respect to one another, and interact with the user.**

In this paper we present an approach to robot manipulation that emphasizes three design themes for robots that manipulate within human environments: *cooperative manipulation*, *task relevant features*, and *let the body do the thinking*. We have previously illustrated these themes with a behavior-based control system that enables a humanoid robot to help a person place everyday objects on a shelf. This system predominantly manipulates a single object at a time with a single arm. Within this paper, we present an extension to this control system that enables a robot to bimanually perform tasks that involve two handheld objects. In our tests, the robot successfully performs insertions that are akin to common activities such as pouring and stirring using a variety of objects, including a bottle, spoon, box, and cup. The success of this extended system suggests that our approach to robot manipulation can support a broad array of useful applications, and demonstrates several distinct advantages of using two arms.
[1]

I. Introduction

Robots that work alongside people in their homes and workplaces could potentially extend the time an elderly person can live at home, provide physical assistance to a worker on an assembly line, or help with household chores. Human environments present special challenges for robot manipulation, since they are complex, dynamic, uncontrolled, and difficult to perceive reliably. For tasks that involve two handheld objects, the use of two arms can help overcome these challenges. With bimanual manipulation, a robot can simultaneously control two handheld objects in order to better perceive key features, control the objects with respect to one another, and interact with the user.

[1]This work was sponsored by Toyota Motor Corporation: Autonomous Manipulation Capabilities for Partner Robots in the Home.

Addressing the challenges of manipulation in human environments is an active area of research. For example, the ARMAR project is investigating manipulation in human environments and has shown results including the bimanual opening of a jar [21]. Researchers working with the NASA Robonaut [1] have demonstrated a cooperative manipulation task where the robot employs a power drill to tighten lugnuts under human direction. Work at AIST has pursued fetch-and-carry tasks of everyday objects under partial teleoperation[18], while work at Stanford has recently investigated learning to grasp novel, everyday objects [16]. Many groups are also pursuing research on autonomous mobile manipulation in human environments [11], [19].

For most of these projects, the robots do not physically interact with people. They also tend to use detailed models of the world that are difficult to generalize and neglect opportunities for physical interactions with the world that can simplify perception and control. In contrast, our approach to robot manipulation emphasizes three design themes: *cooperative manipulation*, *task relevant features*, and *let the body do the thinking*. We have previously illustrated these themes with a behavior-based control system that enables a humanoid robot to help a person place everyday objects on a shelf [5]. Within this paper we extend this control system to enable a robot to perform tasks bimanually with everyday handheld objects. The success of this extended system suggests that our approach to robot manipulation can support a broad array of useful applications, and demonstrates several distinct advantages of using two arms.

Our work is implemented on the 29 degree-of-freedom humanoid robot, Domo, pictured in Figure 1. Domo is mechanically distinctive in that it incorporates passive compliance and force sensing throughout its body [7]. Its Series Elastic Actuators lower the mechanical impedance of its arms, allowing for safe physical interaction with a person [15], [20]. Working with unmodeled objects against a cluttered background, Domo is able to assist a person in a task akin to preparing a drink. As shown in Figure 1, Domo can socially cue a person to hand it a cup

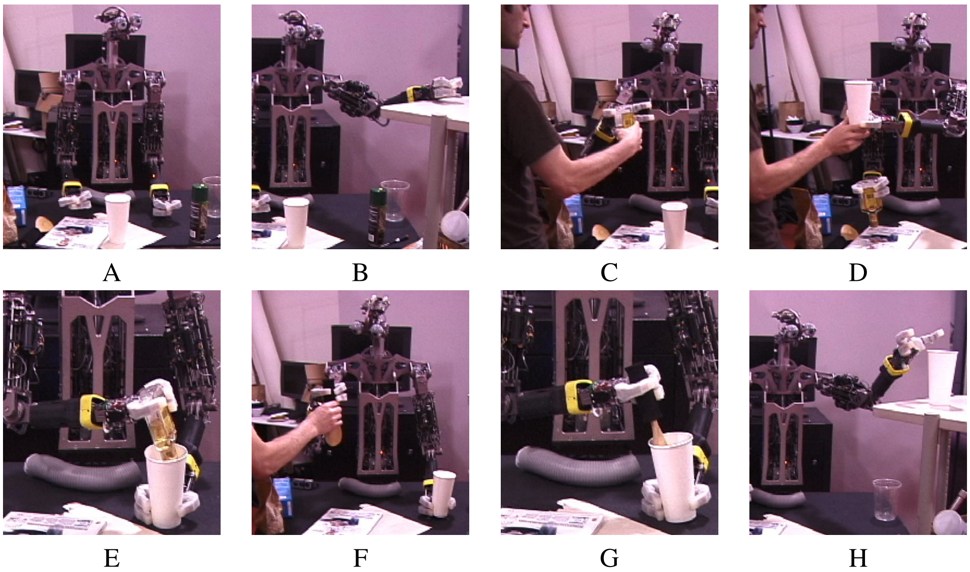

Fig. 1. The humanoid robot Domo assisting a collaborator in a task similar to making a drink. (A-B) Working at a cluttered table, Domo physically verifies the location of a shelf surface. (C-D) Upon request, Domo grasps a bottle and a cup handed to it by the collaborator. (E-F) Domo inserts the bottle into the cup, hands the bottle back to the collaborator, and then acquires a spoon from the collaborator. (G-H) Domo inserts the spoon into the cup, stirs it, and then puts the cup on the shelf.

and a bottle, grasp the objects that have been handed to it, and conduct a visually guided insertion of the bottle into the cup. Domo can then repeat this process using a spoon to stir the interior of the cup, and place the cup on a shelf upon completion. This type of help might enable a person with serious physical limitations to maintain independence in everyday activities that would otherwise require human assistance. For a factory worker, this type of help could potentially offload physically demanding aspects of a task onto a robot.

II. Three Themes for Design

As previously described in [5], three themes characterize our approach to manipulation in human environments. We review these themes here. The first theme, *cooperative manipulation*, refers to the advantages that can be gained by having the robot work with a person to cooperatively perform manipulation tasks. The second theme, *task relevant features*, emphasizes the benefits of carefully selecting the aspects of the world that are to be perceived and acted upon during a manipulation task. The third theme, *let the body do the thinking*, encompasses several ways in which a robot can use its body to simplify manipulation tasks.

A. Cooperative manipulation

For at least the near term, robots in human environments will be dependent on people. Fortunately, people tend to be present within human environments. As long as the robot's usefulness outweighs the efforts required to help it, full autonomy is unnecessary. With careful design robots can be made more intuitive to use, thereby reducing the effort required.

B. Task relevant features

Donald Norman's book *The Design of Everyday Things* [13], emphasizes that objects found within human environments have been designed to match our physical and cognitive abilities. These objects are likely to have common structural features that simplify their use. By developing controllers that are matched to these structural features, we can simplify robot manipulation tasks. Rather than attempting to reconstruct the world in its entirety, we focus the robot's sensory resources on elements of the world that are relevant to the current task.

C. Let the Body Do The Thinking

This theme bundles together design strategies that make use of the robot's body to simplify manipulation in three ways.

First, human environments, interactions, and tasks are well matched to the human body. For example, Domo's eye gaze, arm gesture, and open hand are similar in appearance to a human requesting an object, and are able to intuitively cue uninstructed, non-specialists [6].

Second, we can mitigate the consequences of uncertainty by trading off perception and control for physical design. This tradeoff is central to Pfeifer's notion of morphological computation [14]. For example, Domo uses passive compliance when inserting one object into another.

Third, a physically embodied agent can use its body to test a perceptual hypothesis, gain a better view on an item of interest, or increase the salience of a sensory signal. For example, in this work Domo simultaneously controls two grasped objects in order to better perceive their distal tips.

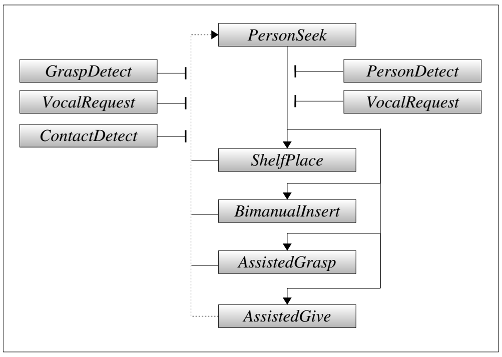

Fig. 2. A collaborator can compose a task using four manipulation behaviors: *ShelfPlace, BimanualInsert, AssistedGrasp*, and *AssistedGive*. Transitions (arrows) occur contingent on perceptual feedback (bars). Exceptions from the expected feedback result in a reset transition (dashed line). The collaborator coordinates the task through voice cues (*VocalRequest)* while the robot tracks the person in the scene (*PersonSeek*, *PersonDetect*). The person can ask the robot to take an object (*AssistedGrasp*), give back an object (*AssistedGive*), insert one object into another (*BimanualInsert*), or place an object on a shelf (*ShelfPlace*). The robot can reattempt a manual skill if failure is signaled (*GraspDetect, VocalRequest*, *ContactDetect*).

III. Behavior-Based Control

A. The Behavior System

Domo performs tasks through the coordination of its perceptual and motor behaviors over time. These behaviors (denoted in italics) are composed hierarchically, and run in a distributed, real-time architecture at $15-100hz$ on a 12 node Linux cluster. We have adopted a layered architecture similar to that of Brooks[2] and Connell[3]. We couple constant perceptual feedback to many simple behaviors in order to increase the task robustness and responsiveness to dynamics in the environment. For example, if a person removes the object from the robot's grasp at anytime during task execution, the active behavior will become inhibited and a secondary behavior will attempt to reacquire the object or to smoothly bring the arm to a relaxed posture.

B. Behaviors

A collaborator coordinates the robot's manual skills to accomplish a task. For example, the task of Figure 1 is accomplished using four manual skills: *ShelfPlace, BimanualInsert, AssistedGrasp*, and *AssistedGive*. As shown in Figure 2, these behaviors run concurrently, allowing a person to vocally request them at any time. If the collaborator notices that Domo is failing at a task, they can provide vocal (*VocalRequest*) or contact (*ContactDetect*) feedback to alert the robot. If Domo accidentally drops an object (*GraspDetect*), the person can pick it up and ask the robot to grasp it again (*AssistedGrasp*). Alternatively, at anytime the person can ask Domo to hand him or her a grasped object (*AssistedGive*). In this way, the robot and the person work as a team. The person provides task-level planning and guides the robot's action selection using intuitive modes of interaction, such as handing objects to the robot and simple verbal commands. In return, the robot performs requested manipulation tasks for the person using the provided objects.

The *AssistedGrasp, AssistedGive,* and *ShelfPlace* behaviors are fully described in [4] and [5]. In the next section we describe the implementation of the *BimanualInsert* behavior in more detail.

IV. The Bimanual Insertion Task

In the *BimanualInsert* behavior, Domo grasps a common object such as a stirring spoon or bottle in one hand and a container such as cup or coffee mug in the other hand. It inserts the object into the container and then optionally stirs the contents. The specific geometric properties and appearance of each object and container are unknown, and their pose in the grasp is uncertain. The robot relies on visual sensing and manipulator compliance to overcome this uncertainty.

This behavior is related to the classic peg-in-hole task often studied in model-based manipulation under uncertainty [12]. For this task a single manipulator controls a peg with the goal of inserting it into a hole. Bimanual insertion is less common.

Through bimanual manipulation a robot can simultaneously control two grasped objects independently. In doing so, the robot can actively control the objects in order to simplify perception and control. For example, Domo wiggles both objects so that it can more easily perceive them through visual motion. Likewise, Domo is able to stabilize the container on a flat surface where it can easily view its opening, hold it steady while inserting the other object, and physically confirm the poses of the objects. Domo is also able to move the objects into its dexterous workspace, where it can more easily perform the physical motions necessary for the task. Finally, by holding both objects at all times, Domo clearly and unambiguously communicates to the person which objects it intends to use for the current task. This is important for cooperative tasks.

The following sections describe the sequential phases of the task in order.

A. AssistedGrasp

AssistedGrasp enlists the person's help in order to secure a grasp on a utensil and a container. By handing Domo the objects, the person directly specifies the objects that Domo will manipulate. In the case of tasks that involve two handheld objects, Domo clearly and unambiguously indicates which objects are in use by holding the objects in its hands. This approach to coordination is both intuitive and effective. It avoids the need for the person to select objects through speech or gesture, and makes it easier for the person to interpret the state or intentions of the robot. By handing the objects to the robot, the system also avoids the challenging robotics problem of locating and autonomously grasping selected

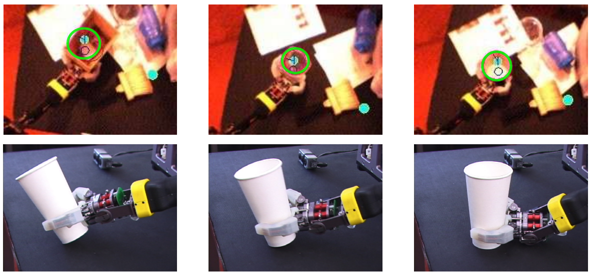

Fig. 3. Execution of the *ContainerPlace* behavior. (Top) The spatio-temporal interest point operator finds the roughly circular opening of a box, jar, and bowl. The detector is robust to cluttered backgrounds. (Bottom) The robot exploits active and passive compliance to align the container to the table.

objects. Robotic grasping of objects is still an active area of research and an open problem [17], [16].

AssistedGrasp locates a person in the scene, extends its arm towards the person, and opens its hand. By reaching towards the person, the robot reduces the need for the person to move when handing over the object. In assistive applications for people with physical limitations, the robot could potentially adapt its reach to the person's capabilities and effectively extend the person's workspace and amplify his or her abilities.

In addition, the robot cues the person through eye contact, directed reaching, and hand opening. This lets him or her know that Domo is ready for an object and prepared to perform the task. The robot monitors contact forces at the hand. If it detects a significant change, it performs a power grasp in an attempt to acquire an object. If the detector *GraspDetect* indicates that an object has been successfully grasped, the robot attempts to acquire another object with its free hand in the same way. Once the robot has an object in each hand, it proceeds to the next phase of the task.

B. ContainerPlace

After *AssistedGrasp*, the orientation of the grasped object in the hand is uncertain. The *ContainerPlace* behavior reduces the orientation uncertainty of a grasped container. Using force control, the behavior lowers the container onto a table while keeping the impedance of the wrist low. This robot behavior is shown in Figure 3.

Since each of the container objects has a flat bottom that is parallel to its opening, this action aligns containers with the table, which results in a stable configuration that is favorable for insertion. This behavior takes advantage of common task relevant features of everyday containers, which have been designed to both accommodate the insertion of objects and stably rest on the flat surfaces that are often found in human environments. For example, people often rest a cup on a table before pouring a cup of coffee.

By using two arms, Domo is able to stably hold the container object against the table throughout the insertion operation. This is important, since compliant contact during the insertion that can generate significant forces and torques on the container. Moreover, throughout the insertion, Domo has the opportunity to physically detect whether or not the object is still being held against the table.

C. TipEstimate

For a wide variety of tools and tasks, control of the tool's endpoint is sufficient for its use. For example, use of a screwdriver requires precise control of the tool blade relative to a screw head but depends little on the details of the tool handle and shaft.

The tip of an object is an important task relevant feature, and we have previously described a method to rapidly localize and control this feature [9], [10]. This method detects fast moving, convex shapes using a form of spatio-temporal interest point operator. As the robot rotates the object, it detects the most rapidly moving convex shape between pairs of consecutive images. Due to the tip's shape and distance from the center of rotation it will tend to produce the most rapidly moving, convex shapes in the image. The robot uses its kinematic model to estimate the 3D point in the hand's coordinate system that best explains these noisy 2D detections.

The *TipEstimate* behavior brings a grasped object into the field of view, rotates its hand, and then localizes the tip. The robot uses the same spatio-temporal interest point operator to detect the opening of the container as it is aligned to the table. As shown in Figure 3, using visual motion and the kinematic model enables the robot to robustly detect this opening on a cluttered table. This method works with a variety of containers such as drinking glasses, bowls, small boxes, and coffee mugs. The opening of the container serves as a form of object tip. Since the tip detector is edge-based, multi-scale, and sensitive to fast moving convex shapes, the edges of the container openings are readily detected.

D. TipPose

Once *TipEstimate* has localized the utensil tip within the hand's coordinate system, the *TipPose* behavior controls the feature by extending the robot's kinematic model by one link. This enables the robot to use traditional Cartesian space control. As the grasped object is moved, the spatio-temporal interest point operator provides visual detections of the tip. This enables the robot to visually servo the tip in the image [4].

Within the insertion task, the *TipPose* behavior visually servoes the object's tip to the container's opening. We adopt an approach similar to [8] where the object is aligned at a 45 degree angle to the table. This advantageous pose avoids visual obstruction of the tip by the hand and expands the range of acceptable misalignment when performing the insertion. During servoing, the tip is kept on the visual ray to the center of the container opening. The depth of the tip is then increased along the ray until the tip is just above the insertion location. This effectively compensates for errors in depth estimation.

Throughout this process, the use of two arms is important. The tip estimation is performed with respect to the hand's coordinate system. By continuing to rigidly grasp an object after estimating the location of its tip, the estimation continues to be relevant and useful. If the robot were to release one of the objects, the uncertainty of the tip's pose relative to the robot's body would be likely to increase, and additional perceptual mechanisms would be required to maintain the estimate, especially in the context of mobile manipulation.

E. CompliantLower

CompliantLower performs the insertion phase of the task by generating a constant downward force at the object's tip. The impedance of the manipulator wrist is also lowered in order to accommodate misalignment. Although the insertion forces are not used for control feedback, the sensed force between the object and the bottom of the container is used to signal task completion.

V. RESULTS

Our three design strategies allow *BimanualInsert* to generalize across a variety of insertion objects and containers. In total, we have executed *BimanualInsert* in nearly one hundred informal trials with a variety of objects. To quantify its performance, we tested *BimanualInsert* in two experiments. In the first experiment, we tested the insertion of a mixing spoon, bottle, paint roller, and paint brush into a paper cup. In the second experiment, we tested the insertion of the mixing spoon into a paper cup, bowl, coffee mug, and jar. On these objects, the size of the container opening varies between 75-100mm and the size of the tool tip varies between 40-60mm. In each experiment, seven trials were conducted on each object pairing.

In a single experiment trial, the object was handed to the robot in an orientation that was deliberately varied between $\pm 20^{\circ}$ along the axis of the hand's power grasp. The grasp location on the object was varied by approximately $\pm$50mm along its length. Each trial took less than 20 seconds to complete and was performed over a visually cluttered table with the collaborating person nearby. A trial was successful if the object was fully inserted into the container. The success rates for both experiments are shown in Figure 4. As the results show, *BimanualInsert* was successful in roughly 90% of the trials. When the visual detection of the tip was disabled, the success rate fell to about 15%.

As a final example, we tested *BimanualInsert* using a flexible hose. The hose has an unknown bend, making it essential that Domo actively sense its distal tip in order to orient the hose prior to insertion. The execution of this test is shown in Figure 5. While *BimanualInsert* can handle the flexible hose in many cases, the single point representation for the tip lacks the orientation information required to reorient the hose and successfully perform the insertion task when the hose has a very large bend. Extending the tip detection system with estimation of the tip's orientation would be useful for these situations.

VI. DISCUSSION

With bimanual manipulation, a robot can simultaneously control two handheld objects in order to better perceive key features, control the objects with respect to one another, and interact with the user. Within this paper, we have presented evidence that these advantages can dramatically simplify manipulation tasks that involve two handheld objects. The control system we have presented relies on both arms, and would not succeed otherwise.

Maintaining rigid grasps on the objects throughout the manipulation task enables the robot to reliably maintain pose estimates for object features, and actively control the objects in order to facilitate new perceptual detections and reestimations. Rigidly grasping the two objects enables the robot to attach the objects to its body and the accompanying coordinate system. Although the world in which the robot is operating is uncontrolled and unmodeled, the robot's body is controlled and well-modeled. Once the robot is holding the two objects, it effectively brings them into a controlled environment.

Within this controlled environment, the robot can efficiently move the objects into favorable configurations for sensing and control. For example, by actively fixturing an object with one arm, the robot can ensure that that the object maintains a favorable configuration in the presence of interaction forces. The ability to handle interaction forces is important to our approach, since it enables the robot to use physical interactions between the objects that help with the task, such as compliance during the insertion. By maintaining contact with the fixtured object, the robot also has the opportunity to physically sense whether or not the fixtured object's state has changed, and provides another channel with which to measure the influence of the interactions between the objects.

With respect to human-robot interaction, the use of two arms enables the robot to directly indicate the objects with which it is working. If the robot is only holding one object, this will be readily apparent to the human. For example, if the task is to pour a drink and the robot is only holding an empty cup, the user can readily infer that the robot should be handed a bottle. Likewise, if the robot is holding a spoon and a mixing bowl, the user can determine an appropriate task for the robot to perform, such as stirring, or decide that the objects are inappropriate.

In the long run, we suspect that these advantages, and others, may outweigh the costs and complexity associated with two armed robots that manipulate in human environments.

REFERENCES

[1] W. Bluethmann, R. Ambrose, A. Fagg, M. Rosenstein, R. Platt, R. Grupen, C. Brezeal, A. Brooks, A. Lockerd, R. Peters, O. Jenk-

	Paper cup	Bowl	Box	Coffee mug	Jar
Mixing spoon	7/7	7/7	7/7	6/7	7/7
Bottle	6/7				
Paint brush	6/7				
Paint roller	5/7				
Spoon (open loop)	1/7				

Fig. 4. Task success for *BimanualInsert*. In a successful trial, Domo inserted the tool (rows, top left) into the container (columns, top right). For comparison, the last row shows results where the visual detection of the tip was disabled. Trials for the blank entries were not attempted.

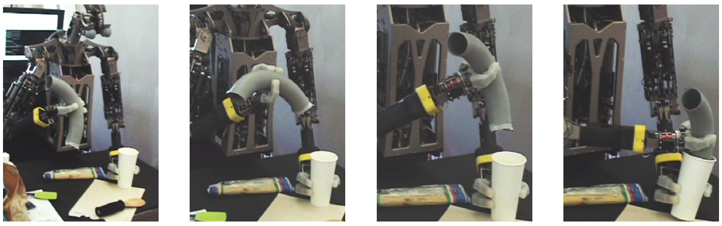

Fig. 5. Execution of *BimanualInsert* using a flexible hose. The unknown bend in the hose requires the active perception of its distal tip and realignment prior to insertion.

ins, M. Mataric, and M. Bugajska. Building an autonomous humanoid tool user. In *Proceedings of the 2004 IEEE International Conference on Humanoid Robots*, Santa Monica, Los Angeles, CA, USA., 2004. IEEE Press.

[2] Rodney A. Brooks. *Cambrian Intelligence*. MIT Press, Cambridge, MA, 1999.

[3] Johnathan Connell. A behavior-based arm controller. *IEEE Transactions on Robotics and Automation*, 5(5):784–791, December 1989.

[4] Aaron Edsinger. *Robot Manipulation in Human Environments*. PhD thesis, Massachusetts Institute of Technology, February 2007.

[5] Aaron Edsinger and Charles C. Kemp. Manipulation in human environments. In *Proceedings of the 2006 IEEE International Conference on Humanoid Robots*, Genoa, Italy, 2006. IEEE Press.

[6] Aaron Edsinger and Charles C. Kemp. Human-robot interaction for cooperative manipulation: Handing objects to one another. In *Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication, ROMAN 2007*, Jeju, Korea, 2007.

[7] Aaron Edsinger-Gonzales and Jeff Weber. Domo: A Force Sensing Humanoid Robot for Manipulation Research. In *Proceedings of the 2004 IEEE International Conference on Humanoid Robots*, Santa Monica, Los Angeles, CA, USA., 2004. IEEE Press.

[8] H. Inoue. Force feedback in precise assembly tasks. In P.H. Winston and R.H. Brown, editors, *Artificial Intelligence: An MIT Perspective*. The MIT Press, 1979.

[9] Charles C. Kemp and Aaron Edsinger. Visual tool tip detection and position estimation for robotic manipulation of unknown human tools. Technical Report AIM-2005-037, MIT Computer Science and Artificial Intelligence Laboratory, 2005.

[10] Charles C. Kemp and Aaron Edsinger. Robot manipulation of human tools: Autonomous detection and control of task relevant features. In *Proceedings of the 5th IEEE International Conference on Development and Learning (ICDL-06)*, Bloomington, Indiana, 2006.

[11] O. Khatib, K. Yokoi, O. Brock, K. Chang, and A. Casal. Robots in human environments: Basic autonomous capabilities. *International Journal of Robotics Research*, 18(684), 1999.

[12] Tomas Lozano-Perez, Matthew Mason, and Russell H. Taylor. Automatic synthesis of fine-motion strategies for robots. *International Journal of Robotics Research*, 3(1), 1984.

[13] D. A. Norman. *The Design of Everyday Things*. Doubleday, New York, New York, 1990.

[14] R. Pfeifer and F. Iida. Morphological computation: Connecting body, brain and environment. *Japanese Scientific Monthly*, 58(2):48–54, 2005.

[15] G. Pratt and M. Williamson. Series Elastic Actuators. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-95)*, volume 1, pages 399–406, Pittsburg, PA, July 1995.

[16] Ashutosh Saxena, Justin Driemeyer, Justin Kearns, Chioma Osondu, and Andrew Y. Ng. Learning to grasp novel objects using vision. In *Proceedings of the International Symposium on Experimental Robotics (ISER)*, 2006.

[17] K. Shimoga. Robot grasp synthesis algorithms: a survey. *International Journal of Robotics Research*, 15(3):230–266, June 1996.

[18] N. Sian, K. Yoki, Y. Kawai, and K. Muruyama. Operating humanoid robots in human environments. In *Proceedings of the Robotics, Science & Systems Workshop on Manipulation for Human Environments*, Philedelphia, Pennsylvania, August 2006.

[19] Y. Yang and O. Brock. Elastic roadmaps: Globally task-consistent motion for autonomous mobile manipulation. In *Proceedings of Robotics: Science and Systems*, Philadelphia, USA, August 2006.

[20] M. Zinn, O. Khatib, B. Roth, and J.K. Salisbury. Playing it safe: human-friendly robots. *IEEE Robotics & Automation Magazine*, 11(2):12–21, June 2004.

[21] R. Zllner, T. Asfour, and R. Dillmann. Programming by demonstration: Dual-arm manipulation tasks for humanoid robots. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004)*, Sendai, Japan, September 28 - October 2, 2004 2004.