Designing Robots that Assist People in Everyday Manual Tasks

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— **Robots that work alongside people in their homes and workplaces could 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. In order to assist people in these ways, robots will need to successfully perform manipulation tasks within human environments. Human environments present special challenges for robot manipulation since they are complex, dynamic, uncontrolled, and difficult to perceive reliably.**

Our approach to robot manipulation overcomes these challenges using a variety of techniques that we group into 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. Within this paper we extend this control system to enable a robot to perform bimanual tasks with everyday handheld objects. In our tests, the robot successfully performs insertion tasks that are akin to common activities such as pouring and stirring using a wide 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.[1]

I. Introduction

Robots that work alongside people in their homes and workplaces could 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. In order to assist people in these ways, robots will need to successfully perform manipulation tasks within human environments. Human environments present special challenges for robot manipulation since they are complex, dynamic, uncontrolled, and difficult to perceive reliably.

Addressing these issues is a focus of several active projects. The ARMAR project is investigating manipulation in human environments and has shown results including the bimanual opening of a jar [29]. 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 tele-operation[23], while work at Stanford has recently investigated learning to grasp unknown, everyday objects [21]. In addition, many groups are pursuing research on autonomous mobile manipulation in human environments [11], [27].

These projects often constrain the robot's environment in order to simplify to issue of perception. In contrast, our approach to robot manipulation addresses the challenges of human environments by using a variety of techniques that can be divided into 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 bimanual tasks with everyday handheld objects.

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 [6]. Its Series Elastic Actuators lower the mechanical impedance of its arms, allowing for safe physical interaction with a person [19], [28]. 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 and a bottle, grasp the objects that have been handed to it, and conduct a visually guided insertion of the bottle into the cup. It then repeats the process with a spoon, and places the cup on a shelf. For an individual with serious physical limitations, this type of help might allow the person to maintain autonomy in everyday activities that would otherwise require help from another person. For a factory worker, this type of help could offload the physically demanding aspects of a task on to the 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

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

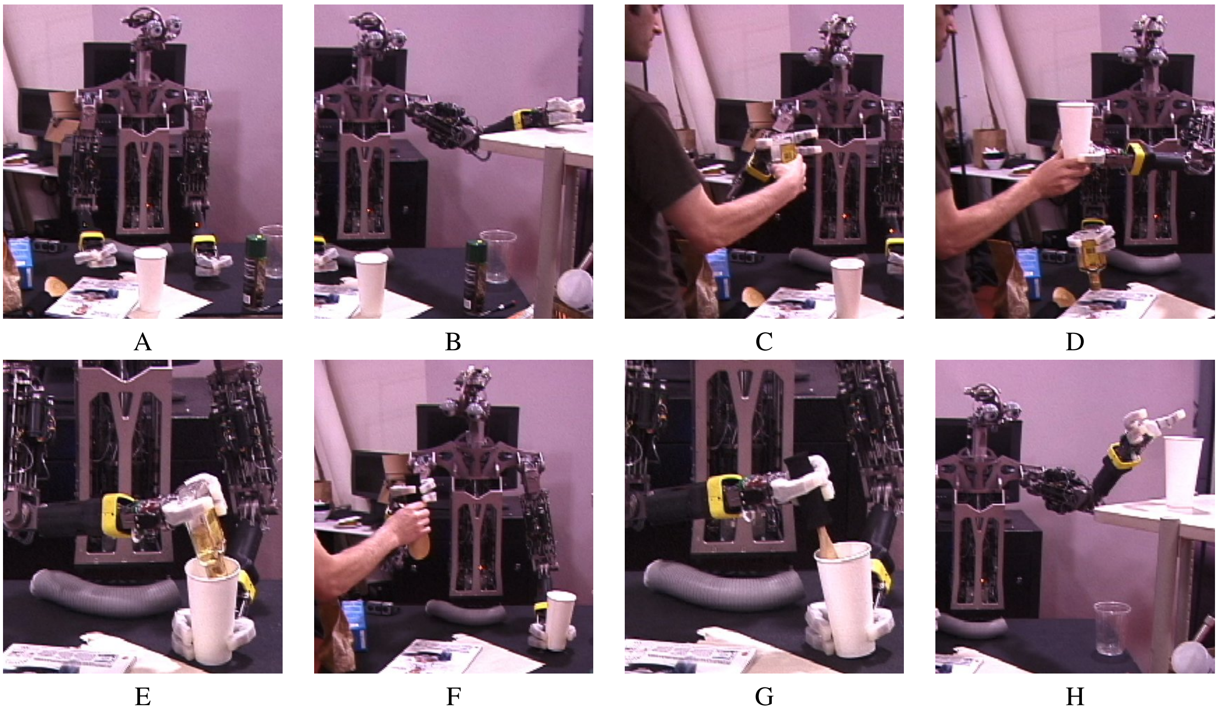

Fig. 1. The humanoid robot Domo assisting a collaborator in a task akin to making a drink. (A-B) Working at a cluttered table, Domo physically verifies the location of a shelf surface. (C-D) Upon request, Domo grasps a bottle and a cup. (E-F) It inserts the bottle into the cup, hands the bottle back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

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, robot autonomy is unnecessary. Careful design can make robots intuitive to use, thereby reducing the required effort.

By treating tasks that involve manipulation as a cooperative process, people and robots can perform tasks that neither one could perform as an individual. Researchers have looked at techniques for cooperative manipulation that physically couple the robot and the person, such as carrying an object together [25], or guiding a person's actions with a Cobot manipulator [14].

B. *Task relevant features*

Donald Norman's book *The Design of Everyday Things* [13], indicates that objects found within human environments are likely to have common structural features that simplify their use (see Figure 2). By developing behaviors that are matched to these structural features, we can simplify robot manipulation tasks. For example, we have previously shown that the manipulation of a large set of human tools can be specified in terms of the tool's tip, such as the tip of a screwdriver [10].

We can define manipulation tasks in terms of behaviors that perform closed-loop control with respect to a feature. Many researchers treat robot manipulation as a planning problem performed with respect to the global state of the world [24], [20]. In contrast, our work is influenced by the work of researchers such as Jagersand, Platt and Grupen, Connell, and Brooks [8], [17], [3], [2], who make use of carefully chosen aspects of the world's state. 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. Other researchers have used task relevant features for manipulation, although typically with fiducial markers or simplified environments [16], [18]. In this paper, we have not altered the world to accommodate perception.

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 are well matched to the human body. Domo's human form allows it to intuitively cue the person with whom it is working. Domo's eye gaze, arm gesture, and open hand are similar in appearance to a human requesting an object. This can help communicate Domo's request and cue the appropriate response more effectively than a wholly alien body.

Fig. 2. Donald Norman's "Coffeepot for Masochists". Many objects in human environments have been designed to match our physical and cognitive abilities. The design of a traditional coffeepot, for example, has evolved such that the pot and coffee can be easily controlled from the handle, the handle is matched to a human-scale power grasp, and the spout is positioned to accommodate perception and control of the spout during pouring. (Personal collection of D. A. Norman. Photograph by Norman. Reproduced with permission).

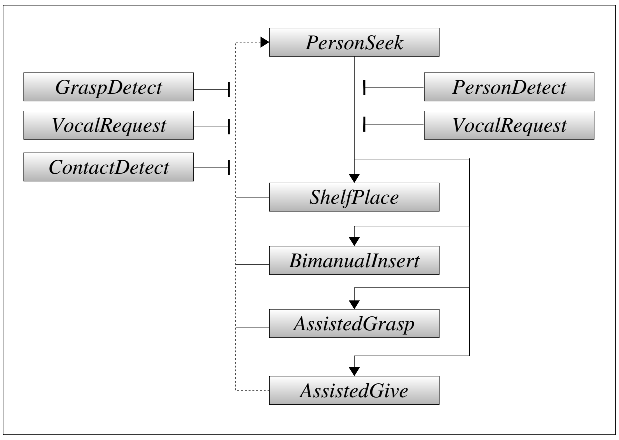

Fig. 3. A collaborator can compose a task using four manipulation behaviors: *ShelfPlace, BimanualInsert, AssistedGrasp*, and *AssistedGive*. Transitions (arrows) occur contingent on perceptual feedback (bars). Exceptions from the expected feedback result in a reset transition (dashed line). The collaborator 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*).

Second, we can mitigate perceptual uncertainty by trading off perceptual computation for physical design. This tradeoff is central to Pfeifer's notion of morphological computation [15]. Morphological computation is characterized as performing a "task distribution" between the robot's controller, body, and environment. This distribution is designed through clever use of sensor placement, material properties, body kinematics, and matching of the robot's body to its environment. This notion has been previously applied to rhythmic manipulation tasks such as hammering [26]. In Domo, the body's passive compliance allows it safely make contact with the world and to take advantage of favorable contact dynamics, such as the tendency of an object to stably align itself with a flat surface.

Third, a physically embodied agent can use its body to to test a perceptual hypothesis, gain a better view on an item of interest, or increase the salience of a sensory signal. For example, a person will tilt their head in order to better hear a speaker. These types of actions can compensate for a robot's physical or perceptual limitations. They can be designed to increase the robot's ability to sense and control important aspects of a task.

III. Designing Tasks

A. Behavior System

Domo accomplishes an assistive task 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 lower-level behavior will attempt to reacquire the object or to smoothly bring the arm to a relaxed posture.

B. Cooperative Design

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 3, 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 them a grasped object (*AssistedGive*). In this way, the robot and the person work as a team. The person intuitively provides task-level planning and guides the robot's action selection. In return, the robot accomplishes manual tasks for the person.

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. Consequently, the robot relies on visual sensing and manipulator compliance to achieve the task.

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 in a hole. Bimanual insertion is much less common. Bimanual insertion requires a more complex body with two arms and two end effectors. One might assume that the task would also be more difficult, since this complex body must be controlled. However, as we demonstrate in this paper, bimanual manipulation can simplify a task in important ways.

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 dramatically simplify perception and action. 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 in order to more easily view its opening and hold it steady while inserting the other object.

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

A. AssistedGrasp

By using *AssistedGrasp*, *BimanualInsert* is able to secure a grasp on a utensil and a container for the task by enlisting a person's help. This is an important form of collaboration. In handing Domo the objects, the person directly specifies the objects that Domo will manipulate. This is both intuitive and effective, and avoids the need for the person to otherwise select objects through speech or gesture. By handing the objects to the robot, the system also avoids the need to autonomously grasp selected objects. Robotic grasping of objects is still a very active field of research and an open problem [22].

AssistedGrasp first locates a person in the scene, and then extends its arm towards the person and opens its hand. By reaching towards the person, the person only needs to move a small amount when handing over the object. In assistive applications for people with motor impairments, this would allow the robot to effectively amplify the person's physical abilities as the robot is able to manipulate the object over its full workspace.

In addition, the robot also cues the person when reaching towards the person. 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 SVM based grasp detector 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.

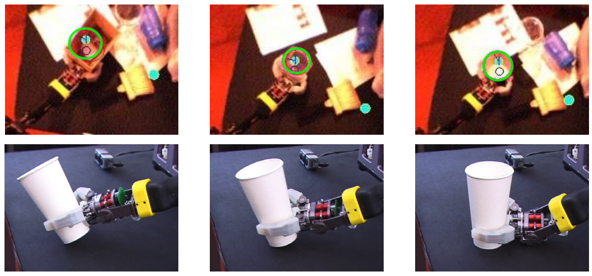

Fig. 4. Execution of the *ContainerPlace* behavior. (Top) The spatio-temporal interest point operator finds the roughly circular opening of a box, jar, and bowl. The detector is robust to cluttered backgrounds. (Bottom) The robot exploits active and passive compliance to align the container to the table.

B. ContainerPlace

The orientation of a grasped object in the hand is typically uncertain. The *ContainerPlace* behavior assists *BimanualInsert* by reducing the orientation uncertainty of a grasped container. Using force control, the behavior lowers the container onto a table while keeping the stiffness of the wrist low. Through this action, the base of the container aligns with the table in a stable configuration. This is shown in Figure 4. Also, the table is used as a stable support during insertion, much like a person resting their cup on a table before pouring a cup of coffee.

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 [10], [9]. 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 treats the tip of the grasped object as a task relevant feature. The behavior brings the 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 4, 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 objects such as drinking glasses, bowls, small boxes, and coffee mugs.

D. TipPose

Once *TipEstimate* has localized the utensil tip within the hand's coordinate system, the *TipPose* behavior controls the feature by essentially 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. These enable 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 [7] where the object is aligned at a 45 degree angle to the table. This advantageous pose prevents 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.

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 zeroed in order to accommodate insertion 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 detect task completion.

V. Results

BimanualInsert can generalize across a variety of insertion objects and containers due to our use of task relevant features. In total, we have executed *BimanualInsert* in nearly one hundred trials with a variety of objects. To demonstrate 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. A trial was successful if the object was fully inserted into the container. The success rates for both experiments are shown in Figure 5. 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 6. While *BimanualInsert* can handle the flexible hose in many cases, a single point representation doesn't provide sufficient information to reorient the hose when it has a large bend. In general, if the 3D orientation of the object tip were sensed using stereo or shape features, the object could be better aligned with the container prior to insertion.

VI. Discussion

We have presented design strategies for building robots that can assist people in everyday tasks. These strategies act to mitigate many of the challenges posed by human environments, which are complex, dynamic, uncontrolled, and difficult to perceive reliably.

In previous work we have shown that these strategies, combined with a behavior-based control system, can enable a robot to assist a person by placing objects on a shelf. In this paper, we have shown that the same system generalizes to bimanual insertion tasks given a variety of everyday handheld objects. This extension provides evidence that our approach is applicable to a broad set of applications, can generalize across objects within a class, can work autonomously without object models, and is robust to the noise and clutter of everyday settings.

References

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

[7] 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.

[8] M. Jagersand and R. Nelson. Visual space task specification, planning and control. In *Proceedings of the IEEE International Symposium on Computer Vision*, pages 521–526, 1995.

[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.

	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 (feedforward)	1/7				

Fig. 5. 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.

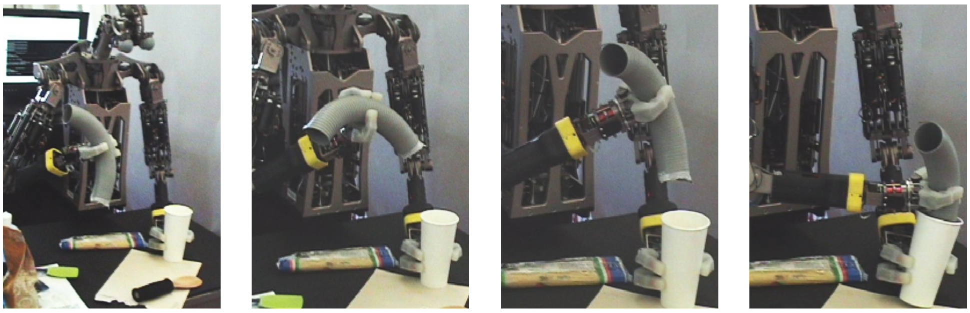

Fig. 6. Execution of *BimanualInsert* using a flexible hose. The unknown bend in the hose requires the active perception of its distal tip and realignment prior to insertion.

[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] Peng Pan, Kevin Lynch, and Michael A. Peshkin. Human interaction with passive assistive robots. In *IEEE 9th International Conference on Rehabilitation Robotics*, June 2005.

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

[16] Justus Piater and Roderic Grupen. Learning appearance features to support robotic manipulation. In *Proceedings of the Cognitive Vision Workshop*, Zurich, Switzerland, 2002.

[17] R. Platt, A. Fagg, and R. Grupen. Manipulation gaits: Sequences of grasp control tasks. In *Proceedings of the 2004 IEEE Conference on Robotics and Automation (ICRA)*, April 2004.

[18] N. Pollard and J.K. Hodgins. Generalizing Demonstrated Manipulation Tasks. In *Proceedings of the Workshop on the Algorithmic Foundations of Robotics (WAFR '02)*, December 2002.

[19] 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.

[20] M. Saha and P. Isto. Motion planning for robotic manipulation of deformable linear objects. In *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2006.

[21] 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.

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

[23] 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.

[24] T. Simeon, J. P. Laumond, J. Cortes, and A. Sahbani. Manipulation planning with probabilistic roadmaps. *IJRR*, 23(7-8):729–746, 2004.

[25] B. Waarsing, M. Nuttin, and H. Van Brussel. Introducing robots into a human-centred environment - the behaviour-based approach. In *Proceedings of the 4th International Conference on CLAWAR*, pages 465–470, May 2001.

[26] Matthew M. Williamson. Exploiting natural dynamics in robot control. In *Fourteenth European Meeting on Cybernetics and Systems Research (EMCSR '98)*, Vienna, Austria, 1998.

[27] 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.

[28] 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.

[29] 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.